Submitted to the Annals of Statistics

SUPPLEMENTARY MATERIAL FOR "EXTREME CONDITIONAL EXPECTILE ESTIMATION IN HEAVY-TAILED HETEROSCEDASTIC REGRESSION MODELS"

BY STÉPHANE GIRARD[1], GILLES STUPFLER[2] AND ANTOINE USSEGLIO-CARLEVE[3]

[1]*Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France, Stephane.Girard@inria.fr*

[2]*Univ Rennes, Ensai, CNRS, CREST - UMR 9194, F-35000 Rennes, France, gilles.stupfler@ensai.fr*

[3]*Toulouse School of Economics, University of Toulouse Capitole, France, antoine.usseglio-carleve@tse-fr.eu*

This supplementary material document contains the proofs of all theoretical results in the main paper, preceded by auxiliary results and their proofs (Sections A and B for the main results, and Sections C and D for the worked-out examples). It also provides further theoretical results related to indirect estimators in Section E, and further details about our finite-sample procedures and studies in Section F.

APPENDIX A: THEORETICAL TOOLBOX: AUXILIARY RESULTS AND THEIR PROOFS

Lemma A.1 below is a result on the mean excess function of a sample of heavy-tailed random variables, used in the proof of Theorem 2.1.

LEMMA A.1. *Assume that ε satisfies condition $\mathcal{C}_1(\gamma)$ with $0<\gamma<1/2$ and $\tau_n \uparrow 1$ is such that $n(1-\tau_n)\to\infty$. Let moreover $t_n\to\infty$ be a nonrandom sequence such that $\overline{F}(t_n)/(1-\tau_n)\to c\in(0,\infty)$. Then*

$$\frac{1}{nt_n(1-\tau_n)}\sum_{i=1}^n \varepsilon_i \mathbb{1}\{\varepsilon_i>t_n\} \xrightarrow{\mathbb{P}} \frac{c}{1-\gamma}.$$

PROOF. Write first

$$\frac{1}{nt_n(1-\tau_n)}\sum_{i=1}^n \varepsilon_i \mathbb{1}\{\varepsilon_i>t_n\} = \frac{c+\mathrm{o}(1)}{nt_n\overline{F}(t_n)}\sum_{i=1}^n \varepsilon_i \mathbb{1}\{\varepsilon_i>t_n\}.$$

The idea is now to split the sum on the right-hand side as follows:

$$\frac{1}{nt_n\overline{F}(t_n)}\sum_{i=1}^n \varepsilon_i \mathbb{1}\{\varepsilon_i>t_n\} = \frac{1}{n\overline{F}(t_n)}\sum_{i=1}^n \mathbb{1}\{\varepsilon_i>t_n\} + \frac{1}{nt_n\overline{F}(t_n)}\sum_{i=1}^n (\varepsilon_i-t_n)\mathbb{1}\{\varepsilon_i>t_n\}.$$

Straightforward expectation and variance calculations yield

$$\mathbb{E}\left(\frac{1}{n\overline{F}(t_n)}\sum_{i=1}^n \mathbb{1}\{\varepsilon_i>t_n\}\right)=1,$$

$$\mathrm{Var}\left(\frac{1}{n\overline{F}(t_n)}\sum_{i=1}^n \mathbb{1}\{\varepsilon_i>t_n\}\right)=\mathrm{O}\left(\frac{1}{n\overline{F}(t_n)}\right)=\mathrm{O}\left(\frac{1}{n(1-\tau_n)}\right)\to 0,$$

MSC 2010 subject classifications: Primary 62G32; secondary 62G08, 62G20, 62G30
Keywords and phrases: Expectiles, extreme value analysis, heavy-tailed distribution, heteroscedasticity, regression models, residual-based estimators, single-index model, tail empirical process of residuals

$$\mathbb{E}\left(\frac{1}{nt_n\overline{F}(t_n)}\sum_{i=1}^n(\varepsilon_i-t_n)\mathbb{1}\{\varepsilon_i>t_n\}\right)=\frac{1}{t_n}\int_{t_n}^{\infty}\frac{\overline{F}(x)}{\overline{F}(t_n)}dx\to\frac{\gamma}{1-\gamma},$$

and $\operatorname{Var}\left(\frac{1}{nt_n\overline{F}(t_n)}\sum_{i=1}^n(\varepsilon_i-t_n)\mathbb{1}\{\varepsilon_i>t_n\}\right)=\mathrm{O}\left(\frac{1}{n\overline{F}(t_n)}\right)=\mathrm{O}\left(\frac{1}{n(1-\tau_n)}\right)\to 0.$

Therefore

$$\frac{1}{nt_n(1-\tau_n)}\sum_{i=1}^n\varepsilon_i\mathbb{1}\{\varepsilon_i>t_n\}\xrightarrow{\mathbb{P}}c\left(1+\frac{\gamma}{1-\gamma}\right)=\frac{c}{1-\gamma}$$

as announced. □

The next auxiliary result is an extension of Theorem 1 in [4]. It drops the assumption of an independent sequence and of an increasing underlying distribution function. We note that the bias term $b(\gamma,\rho)$ of our result below is simpler than the corresponding bias term of Theorem 1 in [4], due to the assumption of a centred noise variable.

PROPOSITION A.1. *Assume that $\mathbb{E}|\varepsilon_-|<\infty$, that condition $\mathcal{C}_2(\gamma,\rho,A)$ holds with $0<\gamma<1$, and that $\mathbb{E}(\varepsilon)=0$. Let $\tau_n\uparrow 1$ be such that $n(1-\tau_n)\to\infty$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1})\to\lambda\in\mathbb{R}$ and $\sqrt{n(1-\tau_n)}/q_{\tau_n}(\varepsilon)=\mathrm{O}(1)$. Then, if*

$$\sqrt{n(1-\tau_n)}\left(\overline{\gamma}-\gamma,\frac{\overline{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)}-1\right)\xrightarrow{d}(\Gamma,\Theta),$$

we have

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}-1\right)\xrightarrow{d}m(\gamma)\Gamma+\Theta-\lambda b(\gamma,\rho)$$

with $m(\gamma)=(1-\gamma)^{-1}-\log(\gamma^{-1}-1)$ and

$$b(\gamma,\rho)=\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}+\frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho}.$$

PROOF. Note that $(\overline{\gamma}^{-1}-1)^{-\overline{\gamma}}\xrightarrow{\mathbb{P}}(\gamma^{-1}-1)^{-\gamma}$ and $\overline{q}_{\tau_n}(\varepsilon)/q_{\tau_n}(\varepsilon)-1\xrightarrow{\mathbb{P}}0$, so that linearising leads to

$$\frac{\widetilde{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}-1=\left(\frac{(\overline{\gamma}^{-1}-1)^{-\overline{\gamma}}}{(\gamma^{-1}-1)^{-\gamma}}-1\right)+\left(\frac{\overline{q}_{\tau_n}(\varepsilon)}{q_{\tau_n}(\varepsilon)}-1\right)(1+\mathrm{o}_{\mathbb{P}}(1))$$

$$+\left(\frac{(\gamma^{-1}-1)^{-\gamma}q_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}-1\right)(1+\mathrm{o}_{\mathbb{P}}(1)).\tag{6}$$

To control the bias term, use Proposition 1 in [5], of which a consequence is, for the centred variable ε,

$$\begin{aligned}\sqrt{n(1-\tau_n)}\left(\frac{(\gamma^{-1}-1)^{-\gamma}q_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}-1\right)&=-\lambda\left[\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}+\frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho}\right]+\mathrm{o}(1)\\&=-\lambda b(\gamma,\rho)+\mathrm{o}(1).\end{aligned}$$

Reporting this in (6) and using the delta-method, we obtain

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}-1\right)\xrightarrow{d}m(\gamma)\Gamma+\Theta-\lambda b(\gamma,\rho).$$

This is precisely the required result. □

The following rearrangement lemma is an extension of Lemma 1 in [10], which we use in the proof of Lemma A.3 below.

LEMMA A.2. *Let $n \geq 2$ and $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$ be two $n-$tuples of real numbers such that for all $i \in \{1, \ldots, n\}$, $a_i \leq b_i$. Then for all $i \in \{1, \ldots, n\}$, $a_{i,n} \leq b_{i,n}$.*

PROOF. See the proof of Lemma 1 in [10], which, although the original result was stated for $n-$tuples featuring no ties, carries over to this more general case with no modification. □

The following lemma is the key to the proof of Theorem 2.2. In our context, its interpretation is that the gap between the tail empirical quantile process of the residuals and the analogue process based on the unobserved errors is bounded above by the gap between errors and their corresponding residuals; this will be used to give an approximation of the tail empirical quantile process of the errors by the tail empirical quantile process of the residuals.

LEMMA A.3. *Let $k = k(n) \to \infty$ be a sequence of integers with $k/n \to 0$. Assume that ε has an infinite right endpoint. Suppose further that the ε_i are independent copies of ε and that the array of random variables $\widehat{\varepsilon}_i^{(n)}$, $1 \leq i \leq n$, satisfies*

$$R_n := \max_{1 \leq i \leq n} \frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{1 + |\varepsilon_i|} \xrightarrow{\mathbb{P}} 0.$$

Then we have both

$$\sup_{0<s\leq 1} \left| \frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks \rfloor,n}}{\varepsilon_{n-\lfloor ks \rfloor,n}} - 1 \right| = \mathrm{O}_{\mathbb{P}}(R_n) \ \text{ and } \ \sup_{0<s\leq 1} \left| \log\left(\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks \rfloor,n}}{\varepsilon_{n-\lfloor ks \rfloor,n}} \right) \right| = \mathrm{O}_{\mathbb{P}}(R_n).$$

PROOF. Clearly:

$$\forall i \in \{1, \ldots, n\}, \ \varepsilon_i - R_n(1 + |\varepsilon_i|) =: \xi_i \leq \widehat{\varepsilon}_i^{(n)} \leq \zeta_i := \varepsilon_i + R_n(1 + |\varepsilon_i|).$$

It then follows from Lemma A.2 that

$$\forall i \in \{1, \ldots, n\}, \ \xi_{i,n} \leq \widehat{\varepsilon}_{i,n}^{(n)} \leq \zeta_{i,n}.$$

Note that for any $r \in (-1, 1)$, the function $x \mapsto x + r(1 + |x|)$ is increasing. Therefore, on the event $\{R_n \leq 1/4\}$, whose probability gets arbitrarily high as n increases, we have:

$$\forall i \in \{1, \ldots, n\}, \ \varepsilon_{i,n} - R_n(1 + |\varepsilon_{i,n}|) = \xi_{i,n} \leq \widehat{\varepsilon}_{i,n}^{(n)} \leq \zeta_{i,n} = \varepsilon_{i,n} + R_n(1 + |\varepsilon_{i,n}|).$$

Now, by Lemma 3.2.1 in [6] together with the equality $\varepsilon \stackrel{d}{=} U(Z)$ where Z has a unit Pareto distribution, we get $\varepsilon_{n-k,n} \xrightarrow{\mathbb{P}} +\infty$. On the event $A_n := \{R_n \leq 1/4\} \cap \{\varepsilon_{n-k,n} \geq 1\}$, which likewise has probability arbitrarily large, we obtain

$$\forall i \geq n-k, \ (1 - R_n)\varepsilon_{i,n} - R_n \leq \widehat{\varepsilon}_{i,n}^{(n)} \leq (1 + R_n)\varepsilon_{i,n} + R_n.$$

In other words, on A_n, and for any $s \in (0, 1]$,

$$-2R_n \leq -R_n \left(1 + \frac{1}{\varepsilon_{n-\lfloor ks \rfloor,n}}\right) \leq \frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks \rfloor,n}}{\varepsilon_{n-\lfloor ks \rfloor,n}} - 1 \leq R_n \left(1 + \frac{1}{\varepsilon_{n-\lfloor ks \rfloor,n}}\right) \leq 2R_n.$$

This shows that

$$\sup_{0<s\leq 1}\left|\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\varepsilon_{n-\lfloor ks\rfloor,n}}-1\right| = \mathrm{O}_{\mathbb{P}}(R_n).$$

Note further that, on A_n,

$$\forall s\in(0,1],\ \log(1-2R_n)\leq \log\left(\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\varepsilon_{n-\lfloor ks\rfloor,n}}\right)\leq \log(1+2R_n).$$

Since $\log(1+x)\leq x$ and $\log(1-x)\geq -2x$ for all $x\in[0,1/2]$, this yields, on A_n,

$$\forall s\in(0,1],\ \left|\log\left(\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\varepsilon_{n-\lfloor ks\rfloor,n}}\right)\right|\leq 4R_n.$$

As a consequence,

$$\sup_{0<s\leq 1}\left|\log\left(\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\varepsilon_{n-\lfloor ks\rfloor,n}}\right)\right| = \mathrm{O}_{\mathbb{P}}(R_n).$$

This concludes the proof. □

The final auxiliary result of this section is used as part of Remark 2. It can be seen as a Breiman-type result, see Proposition 3 in [2] for the original Breiman lemma.

LEMMA A.4. *Suppose that the random variable Y can be written $Y=Z_1+Z_2\,\varepsilon$, where*

- *Z_1 is a bounded random variable,*
- *Z_2 is a (strictly) positive and bounded random variable,*
- *ε satisfies condition $\mathcal{C}_1(\gamma)$,*
- *Z_2 is independent of ε.*

Then Y satisfies condition $\mathcal{C}_1(\gamma)$.

PROOF. We prove that for all $x>0$, $\mathbb{P}(Y>tx)/\mathbb{P}(Y>t)\to x^{-1/\gamma}$ as $t\to\infty$. Note that if a_1,b_1 are such that $Z_1\in[a_1,b_1]$ with probability 1,

$$\frac{\mathbb{P}(Z_2\,\varepsilon>tx-a_1)}{\mathbb{P}(Z_2\,\varepsilon>t-b_1)}\leq\frac{\mathbb{P}(Y>tx)}{\mathbb{P}(Y>t)}\leq\frac{\mathbb{P}(Z_2\,\varepsilon>tx-b_1)}{\mathbb{P}(Z_2\,\varepsilon>t-a_1)}.$$

This entails, for any fixed $\varepsilon\in(0,1)$, that for t large enough,

$$\frac{\mathbb{P}(Z_2\,\varepsilon>t(x+\varepsilon))}{\mathbb{P}(Z_2\,\varepsilon>t(1-\varepsilon))}\leq\frac{\mathbb{P}(Y>tx)}{\mathbb{P}(Y>t)}\leq\frac{\mathbb{P}(Z_2\,\varepsilon>t(x-\varepsilon))}{\mathbb{P}(Z_2\,\varepsilon>t(1+\varepsilon))}.$$

Let $b_2>0$ be such that $Z_2\in(0,b_2]$ with probability 1. Since Z_2 is independent of ε, we have for any $t>0$

$$\frac{\mathbb{P}(Z_2\,\varepsilon>t)}{\mathbb{P}(\varepsilon>t)}=\int_0^{b_2}\frac{\mathbb{P}(\varepsilon>t/z)}{\mathbb{P}(\varepsilon>t)}\mathbb{P}_{Z_2}(dz).$$

Use now Potter bounds (see *e.g.* Proposition B.1.9.5 in [6]) and the dominated convergence theorem to obtain

$$\frac{\mathbb{P}(Z_2\,\varepsilon>t)}{\mathbb{P}(\varepsilon>t)}\to\int_0^{b_2}z^{\gamma}\,\mathbb{P}_{Z_2}(dz)=\mathbb{E}(Z_2^{\gamma})\in(0,\infty).$$

This implies that $Z_2\,\varepsilon$ is, like ε, heavy-tailed with extreme value index γ. In particular

$$\frac{\mathbb{P}(Z_2\,\varepsilon > t(x \mp \varepsilon))}{\mathbb{P}(Z_2\,\varepsilon > t(1 \pm \varepsilon))} = \frac{\mathbb{P}(Z_2\,\varepsilon > t(x \mp \varepsilon))}{\mathbb{P}(Z_2\,\varepsilon > t)} \frac{\mathbb{P}(Z_2\,\varepsilon > t)}{\mathbb{P}(Z_2\,\varepsilon > t(1 \pm \varepsilon))} \to (1 \pm \varepsilon)^{1/\gamma}(x \mp \varepsilon)^{-1/\gamma}$$

as $t \to \infty$. Conclude that

$$(1-\varepsilon)^{1/\gamma}(x+\varepsilon)^{-1/\gamma} \le \liminf_{t\to\infty} \frac{\mathbb{P}(Y > tx)}{\mathbb{P}(Y > t)} \le \limsup_{t\to\infty} \frac{\mathbb{P}(Y > tx)}{\mathbb{P}(Y > t)} \le (1+\varepsilon)^{1/\gamma}(x-\varepsilon)^{-1/\gamma}$$

for any $\varepsilon > 0$, and let $\varepsilon \downarrow 0$ to complete the proof. □

APPENDIX B: THEORETICAL TOOLBOX: PROOFS OF THE MAIN RESULTS

PROOF OF THEOREM 2.1. Note that

$$\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1\right) = \arg\min_{u\in\mathbb{R}} \chi_n(u)$$

with $\chi_n(u) := \dfrac{1}{2\xi^2_{\tau_n}(\varepsilon)} \displaystyle\sum_{i=1}^n \left[\eta_{\tau_n}\left(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - \frac{u\xi_{\tau_n}(\varepsilon)}{\sqrt{n(1-\tau_n)}}\right) - \eta_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon))\right].$

Define

$$\psi_n(u) := \frac{1}{2\xi^2_{\tau_n}(\varepsilon)} \sum_{i=1}^n \left[\eta_{\tau_n}\left(\varepsilon_i - \xi_{\tau_n}(\varepsilon) - \frac{u\xi_{\tau_n}(\varepsilon)}{\sqrt{n(1-\tau_n)}}\right) - \eta_{\tau_n}(\varepsilon_i - \xi_{\tau_n}(\varepsilon))\right].$$

In other words, $\psi_n(u)$ is the counterpart of $\chi_n(u)$ based on the true, unobservable errors ε_i. Note that for any n, $u \mapsto \psi_n(u)$ is a continuously differentiable convex function. We shall prove that, pointwise in u, $\chi_n(u) - \psi_n(u) \xrightarrow{\mathbb{P}} 0$. The result will then be a straightforward consequence of a convexity lemma stated as Theorem 5 in [15] together with the convergence

$$\psi_n(u) \xrightarrow{d} -uZ\sqrt{\frac{2\gamma}{1-2\gamma}} + \frac{u^2}{2\gamma} \ \text{ as } \ n \to \infty$$

(in the sense of finite-dimensional convergence, with Z being standard Gaussian) shown in the proof of Theorem 2 in [4].

We start by recalling that

$$\frac{1}{2}(\eta_\tau(x-y) - \eta_\tau(x)) = -\int_0^y \varphi_\tau(x-t)dt$$

where $\varphi_\tau(y) = |\tau - \mathbb{1}\{y \le 0\}|y$ (see Lemma 2 in [4]). Therefore

$$\begin{aligned}
&\chi_n(u) - \psi_n(u) \\
&= -\frac{1}{\xi^2_{\tau_n}(\varepsilon)} \sum_{i=1}^n \int_0^{u\xi_{\tau_n}(\varepsilon)/\sqrt{n(1-\tau_n)}} [\varphi_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - t) - \varphi_{\tau_n}(\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t)]dt.
\end{aligned}$$

Set $I_n(u) = [0, |u|\xi_{\tau_n}(\varepsilon)/\sqrt{n(1-\tau_n)}]$. Since

$$\begin{aligned}
&|\chi_n(u) - \psi_n(u)| \\
&\le \frac{|u|}{\xi_{\tau_n}(\varepsilon)\sqrt{n(1-\tau_n)}} \sum_{i=1}^n \sup_{|t|\in I_n(u)} |\varphi_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - t) - \varphi_{\tau_n}(\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t)|,
\end{aligned}$$

it is enough to show that

$$T_n(u) := \frac{1}{\xi_{\tau_n}(\varepsilon)\sqrt{n(1-\tau_n)}} \sum_{i=1}^n \sup_{|t|\in I_n(u)} |\varphi_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - t) - \varphi_{\tau_n}(\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t)|$$

$$\xrightarrow{\mathbb{P}} 0. \tag{7}$$

We now apply Lemma 3 in [4], which gives, for any $x, h \in \mathbb{R}$,

$$|\varphi_\tau(x-h) - \varphi_\tau(x)| \le |h|(1 - \tau + 2\mathbb{1}\{x > \min(h,0)\}).$$

This translates into

$$\begin{aligned} &|\varphi_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - t) - \varphi_{\tau_n}(\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t)| \\ &\le |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|(1 - \tau_n + 2\mathbb{1}\{\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t > \min(\varepsilon_i - \widehat{\varepsilon}_i^{(n)}, 0)\}). \end{aligned}$$

Hence the inequality

$$T_n(u) \le T_{1,n} + T_{2,n}(u) \tag{8}$$

with

$$T_{1,n} := \frac{\sqrt{1-\tau_n}}{\xi_{\tau_n}(\varepsilon)\sqrt{n}} \sum_{i=1}^n |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \ \text{ and}$$

$$T_{2,n}(u) := \frac{2}{\xi_{\tau_n}(\varepsilon)\sqrt{n(1-\tau_n)}} \sum_{i=1}^n \sup_{|t|\in I_n(u)} |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \mathbb{1}\{\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t > \min(\varepsilon_i - \widehat{\varepsilon}_i^{(n)}, 0)\}.$$

We first focus on $T_{1,n}$. Define $R_{n,i} := |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|/(1+|\varepsilon_i|)$ and $R_n = \max_{1\le i\le n} R_{n,i}$. We have

$$T_{1,n} \le \left[\frac{\sqrt{n(1-\tau_n)}}{\xi_{\tau_n}(\varepsilon)} R_n\right] \times \frac{1}{n}\sum_{i=1}^n (1+|\varepsilon_i|) = \mathrm{O}_{\mathbb{P}}\left(\frac{\sqrt{n(1-\tau_n)}}{\xi_{\tau_n}(\varepsilon)} R_n\right)$$

by the law of large numbers. Note now that $\xi_{\tau_n}(\varepsilon) \to \infty$ and thus

$$T_{1,n} = \mathrm{O}_{\mathbb{P}}\left(\frac{\sqrt{n(1-\tau_n)}}{q_{\tau_n}(\varepsilon)} R_n\right) = \mathrm{o}_{\mathbb{P}}\left(\sqrt{n(1-\tau_n)} R_n\right) \xrightarrow{\mathbb{P}} 0 \tag{9}$$

by assumption. We now turn to the control of $T_{2,n}(u)$, for which we write, for any t,

$$\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t > \min(\varepsilon_i - \widehat{\varepsilon}_i^{(n)}, 0) \Rightarrow \varepsilon_i - \xi_{\tau_n}(\varepsilon) - t > 0 \ \text{ or } \ \widehat{\varepsilon}_i^{(n)} - \xi_{\tau_n}(\varepsilon) - t > 0.$$

It follows that, for n large enough, we have, for any t such that $|t| \in I_n(u)$,

$$\varepsilon_i - \xi_{\tau_n}(\varepsilon) - t > \min(\varepsilon_i - \widehat{\varepsilon}_i^{(n)}, 0) \Rightarrow \varepsilon_i > \frac{\xi_{\tau_n}(\varepsilon)}{2} \ \text{ or } \ \widehat{\varepsilon}_i^{(n)} > \frac{\xi_{\tau_n}(\varepsilon)}{2}. \tag{10}$$

Now, for n large enough and with arbitrarily large probability as $n \to \infty$, $|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \le (1 + |\varepsilon_i|)/2$ for any $i \in \{1, \ldots, n\}$, so that after some algebra,

$$\widehat{\varepsilon}_i^{(n)} > \frac{\xi_{\tau_n}(\varepsilon)}{2} \Rightarrow \varepsilon_i + \frac{1}{2}|\varepsilon_i| > \frac{1}{2}(\xi_{\tau_n}(\varepsilon) - 1) \Rightarrow \varepsilon_i + \frac{1}{2}|\varepsilon_i| > \frac{1}{4}\xi_{\tau_n}(\varepsilon)$$

because $\xi_{\tau_n}(\varepsilon) \to \infty$. Since the quantity $x + |x|/2$ can only be positive if $x > 0$, it follows that, with arbitrarily large probability,

$$\widehat{\varepsilon}_i^{(n)} > \frac{\xi_{\tau_n}(\varepsilon)}{2} \Rightarrow \varepsilon_i > \frac{1}{6}\xi_{\tau_n}(\varepsilon). \tag{11}$$

Combining (10) and (11) results in the following bound, valid with arbitrarily large probability as $n \to \infty$:

$$T_{2,n}(u) \le \frac{2}{\xi_{\tau_n}(\varepsilon)\sqrt{n(1-\tau_n)}} \sum_{i=1}^{n} |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i| \mathbb{1}\left\{\varepsilon_i > \frac{1}{6}\xi_{\tau_n}(\varepsilon)\right\}.$$

By assumption on $|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|$, this leads to

$$T_{2,n}(u) \le 4\left[\frac{\sqrt{n(1-\tau_n)}}{\xi_{\tau_n}(\varepsilon)} R_n\right] \times \frac{1}{n(1-\tau_n)} \sum_{i=1}^{n} \varepsilon_i \mathbb{1}\left\{\varepsilon_i > \frac{1}{6}\xi_{\tau_n}(\varepsilon)\right\}.$$

Finally, the regular variation property of $\overline{F}$ and the asymptotic proportionality relationship between $\xi_{\tau_n}(\varepsilon)$ and $q_{\tau_n}(\varepsilon)$ ensure that

$$\lim_{n\to\infty} \frac{\overline{F}\,(\xi_{\tau_n}(\varepsilon)/6)}{1-\tau_n} \text{ exists, is positive and finite.}$$

Lemma A.1 then entails

$$T_{2,n}(u) = \mathrm{O}_{\mathbb{P}}\left(\sqrt{n(1-\tau_n)} R_n\right) \xrightarrow{\mathbb{P}} 0 \tag{12}$$

by assumption. Combining (7), (8), (9) and (12) completes the proof. □

PROOF OF THEOREM 2.2. To prove the first expansion, write

$$\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{q_{1-k/n}(\varepsilon)} - s^{-\gamma} = \frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\varepsilon_{n-\lfloor ks\rfloor,n}}\left(\frac{\varepsilon_{n-\lfloor ks\rfloor,n}}{q_{1-k/n}(\varepsilon)} - s^{-\gamma}\right) + s^{-\gamma}\left(\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\varepsilon_{n-\lfloor ks\rfloor,n}} - 1\right).$$

Use Lemma A.3 and Theorem 2.4.8 in [6] to get

$$\begin{aligned} &\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\varepsilon_{n-\lfloor ks\rfloor,n}}\left(\frac{\varepsilon_{n-\lfloor ks\rfloor,n}}{q_{1-k/n}(\varepsilon)} - s^{-\gamma}\right) \\ &= \frac{1}{\sqrt{k}}\left[\gamma s^{-\gamma-1} W_n(s) + \sqrt{k} A(n/k) s^{-\gamma}\frac{s^{-\rho}-1}{\rho} + s^{-\gamma-1/2-\delta}\,\mathrm{o}_{\mathbb{P}}(1)\right] \end{aligned} \tag{13}$$

uniformly in $s \in (0,1]$. Applying Lemma A.3 again gives

$$s^{-\gamma}\left|\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\varepsilon_{n-\lfloor ks\rfloor,n}} - 1\right| \le s^{-\gamma-1/2-\delta}\left|\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\varepsilon_{n-\lfloor ks\rfloor,n}} - 1\right| = \frac{s^{-\gamma-1/2-\delta}}{\sqrt{k}}\,\mathrm{o}_{\mathbb{P}}(1) \tag{14}$$

uniformly in $s \in (0,1]$. Combine (13) and (14) to complete the proof of the first expansion. The proof of the second expansion is based on the equality

$$\log\left(\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{q_{1-k/n}(\varepsilon)}\right) = \log\left(\frac{\varepsilon_{n-\lfloor ks\rfloor,n}}{q_{1-k/n}(\varepsilon)}\right) + \log\left(\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\varepsilon_{n-\lfloor ks\rfloor,n}}\right)$$

and follows exactly the same ideas. □

PROOF OF COROLLARY 2.1. Notice that, by Theorem 2.2, there is a sequence W_n of standard Brownian motions such that, for any $\delta > 0$ sufficiently small:

$$\widehat{\gamma}_k = \int_0^1 \log\left(\frac{\widehat{\varepsilon}^{(n)}_{n-\lfloor ks\rfloor,n}}{\widehat{\varepsilon}^{(n)}_{n-k,n}}\right) ds$$

$$= \int_0^1 \left\{\gamma \log\frac{1}{s} + \frac{\gamma}{\sqrt{k}}\left[s^{-1}W_n(s) - W_n(1)\right] + A\left(\frac{n}{k}\right)\left[\frac{s^{-\rho}-1}{\rho} + s^{-1/2-\delta}\,\mathrm{o}_{\mathbb{P}}(1)\right]\right\} ds.$$

We then obtain that $\widehat{\gamma}_k$ can be written

$$\sqrt{k}(\widehat{\gamma}_k - \gamma) = \frac{\lambda}{1-\rho} + \gamma\int_0^1 \left[s^{-1}W_n(s) - W_n(1)\right] ds + \mathrm{o}_{\mathbb{P}}(1).$$

Similarly,

$$\sqrt{k}\left(\frac{\widehat{\varepsilon}^{(n)}_{n-k,n}}{q_{1-k/n}(\varepsilon)} - 1\right) = \gamma W_n(1) + \mathrm{o}_{\mathbb{P}}(1).$$

Noting that the Gaussian terms in these two asymptotic expansions are independent completes the proof. □

PROOF OF THEOREM 2.3. The key is to note that

$$\frac{\overline{\xi}^{\star}_{\tau_n'}(Y|\boldsymbol{x})}{\xi_{\tau_n'}(Y|\boldsymbol{x})} - 1 = \left(1 + \frac{g(\boldsymbol{x})}{\sigma(\boldsymbol{x})\xi_{\tau_n'}(\varepsilon)}\right)^{-1}\left(\frac{\overline{\xi}^{\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)} - 1\right)$$
$$+ \frac{\overline{g}(\boldsymbol{x}) - g(\boldsymbol{x})}{g(\boldsymbol{x}) + \sigma(\boldsymbol{x})\xi_{\tau_n'}(\varepsilon)} + \left(1 + \frac{g(\boldsymbol{x})}{\sigma(\boldsymbol{x})\xi_{\tau_n'}(\varepsilon)}\right)^{-1}\frac{\overline{\sigma}(\boldsymbol{x}) - \sigma(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\frac{\overline{\xi}^{\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}.$$

Using the convergence $\xi_\tau(\varepsilon)/q_\tau(\varepsilon) \to (\gamma^{-1}-1)^{-\gamma}$ as $\tau \uparrow 1$ and the heavy-tailed condition, we find $1/\xi_{\tau_n'}(\varepsilon) = \mathrm{o}(1/\xi_{\tau_n}(\varepsilon)) = \mathrm{o}(1/q_{\tau_n}(\varepsilon))$. Our assumptions show that this is a $\mathrm{o}(1/\sqrt{n(1-\tau_n)})$ and therefore

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\overline{\xi}^{\star}_{\tau_n'}(Y|\boldsymbol{x})}{\xi_{\tau_n'}(Y|\boldsymbol{x})} - 1\right)$$
$$= \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\overline{\xi}^{\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)} - 1\right)(1 + \mathrm{o}_{\mathbb{P}}(1)) + \mathrm{o}_{\mathbb{P}}(1).$$

Our result is then shown by adapting the proof of Theorem 5 of [5], with the condition $\rho < 0$ being used exclusively to control the bias term appearing naturally because of the extrapolation procedure applied to the heavy-tailed random variable ε. We omit the details. □

APPENDIX C: WORKED-OUT EXAMPLES: AUXILIARY RESULTS AND THEIR PROOFS

Lemma C.1 gives the rate of convergence of the weighted least squares estimators in model (M_1). Here and throughout all $\mathrm{O}_{\mathbb{P}}(1)$ statements are meant componentwise.

LEMMA C.1. *Assume that* $(\boldsymbol{X}_i, Y_i)_{i\geq 1}$ *are independent random pairs generated from model* (M_1). *Suppose further that* $\mathbb{E}(\varepsilon^2) < \infty$. *Then we have*

$$\sqrt{n}(\widehat{\alpha} - \alpha) = \mathrm{O}_{\mathbb{P}}(1), \ \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathrm{O}_{\mathbb{P}}(1) \ \textit{and} \ \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathrm{O}_{\mathbb{P}}(1).$$

PROOF. We introduce the notation

$$\mathfrak{X} = \begin{pmatrix} 1 \ \boldsymbol{X}_1^\top \\ \vdots \ \ \vdots \\ 1 \ \boldsymbol{X}_n^\top \end{pmatrix}, \ \boldsymbol{Y} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} \text{ and } \boldsymbol{\Omega} = \mathrm{diag}([1+\boldsymbol{\theta}^\top \boldsymbol{X}_1]^2, \ldots, [1+\boldsymbol{\theta}^\top \boldsymbol{X}_n]^2).$$

A preliminary step is to remark that for any $\boldsymbol{a} = (a_0, a_1, \ldots, a_d)^\top \in \mathbb{R}^{d+1}$,

$$\boldsymbol{a}^\top \mathfrak{X}^\top \mathfrak{X} \boldsymbol{a} = \sum_{i=1}^n [a_0 + (a_1, \ldots, a_d)\boldsymbol{X}_i]^2 > 0$$
$$\text{and } \boldsymbol{a}^\top \mathfrak{X}^\top \boldsymbol{\Omega}^{-1} \mathfrak{X} \boldsymbol{a} = \sum_{i=1}^n \left[1+\boldsymbol{\theta}^\top \boldsymbol{X}_i\right]^{-2} [a_0 + (a_1, \ldots, a_d)\boldsymbol{X}_i]^2 > 0$$

with probability 1, because $\boldsymbol{X}$ has a continuous distribution (and as such, does not put mass on affine hyperplanes of $\mathbb{R}^d$). The symmetric matrices $\mathfrak{X}^\top \mathfrak{X}$ and $\mathfrak{X}^\top \boldsymbol{\Omega}^{-1} \mathfrak{X}$ therefore have full rank with probability 1. Since, by the law of large numbers,

$$\frac{1}{n}\left[\mathfrak{X}^\top \mathfrak{X}\right]_{i+1,j+1} \xrightarrow{\mathbb{P}} \mathbb{E}\left(X_i X_j\right) \text{ and } \frac{1}{n}\left[\mathfrak{X}^\top \boldsymbol{\Omega}^{-1} \mathfrak{X}\right]_{i+1,j+1} \xrightarrow{\mathbb{P}} \mathbb{E}\left(\left[1+\boldsymbol{\theta}^\top \boldsymbol{X}\right]^{-2} X_i X_j\right)$$

(where $X_0 = 1$ for notational convenience), the same argument shows that $\mathfrak{X}^\top \mathfrak{X}/n$ and $\mathfrak{X}^\top \boldsymbol{\Omega}^{-1} \mathfrak{X}/n$ converge in probability to symmetric positive definite matrices, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ say.

Our first step is to show that the preliminary estimators $\widetilde{\alpha}$, $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$ are $\sqrt{n}$−consistent. Rewrite model (M_1) for the available data as

$$\boldsymbol{Y} = \mathfrak{X}\begin{pmatrix} \alpha \\ \boldsymbol{\beta} \end{pmatrix} + \left[\mathfrak{X}\begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix}\right] \circ \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}^\top = (\varepsilon_1, \ldots, \varepsilon_n)$ and $\circ$ denotes the Hadamard (entrywise) product of matrices. By standard least squares theory,

$$\begin{pmatrix} \widetilde{\alpha} \\ \widetilde{\boldsymbol{\beta}} \end{pmatrix} = \left(\mathfrak{X}^\top \mathfrak{X}\right)^{-1} \mathfrak{X}^\top \boldsymbol{Y}.$$

A direct calculation then yields

$$\begin{pmatrix} \sqrt{n}(\widetilde{\alpha} - \alpha) \\ \sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \end{pmatrix} = n\left(\mathfrak{X}^\top \mathfrak{X}\right)^{-1} \times \frac{1}{\sqrt{n}} \mathfrak{X}^\top \left\{\left[\mathfrak{X}\begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix}\right] \circ \boldsymbol{\varepsilon}\right\}$$
$$= n\left(\mathfrak{X}^\top \mathfrak{X}\right)^{-1} \times \begin{pmatrix} n^{-1/2} \sum_{i=1}^n \left[1+\boldsymbol{\theta}^\top \boldsymbol{X}_i\right] \varepsilon_i \\ n^{-1/2} \sum_{i=1}^n \left[1+\boldsymbol{\theta}^\top \boldsymbol{X}_i\right] X_{i1} \varepsilon_i \\ \vdots \\ n^{-1/2} \sum_{i=1}^n \left[1+\boldsymbol{\theta}^\top \boldsymbol{X}_i\right] X_{id} \varepsilon_i \end{pmatrix}.$$

Set for notational convenience $X_{i0} = 1$. Since, for any $m \in \{0, 1 \ldots, d\}$, the random variables $\left[1+\boldsymbol{\theta}^\top \boldsymbol{X}_i\right] X_{im} \varepsilon_i$, $1 \leq i \leq n$, are independent, centred and square-integrable, the standard multivariate central limit theorem combined with the convergence $n\left(\mathfrak{X}^\top\mathfrak{X}\right)^{-1} \xrightarrow{\mathbb{P}} \boldsymbol{\Sigma}_1^{-1}$ yields

$$\sqrt{n}(\widetilde{\alpha}-\alpha) = \mathrm{O}_{\mathbb{P}}(1) \text{ and } \sqrt{n}(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \mathrm{O}_{\mathbb{P}}(1). \tag{15}$$

We then prove that $\sqrt{n}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}) = \mathrm{O}_{\mathbb{P}}(1)$. Recalling that

$$\widetilde{\boldsymbol{\theta}} = \frac{\widetilde{\boldsymbol{\nu}}}{\widetilde{\mu}} \text{ and } \boldsymbol{\theta} = \frac{\boldsymbol{\nu}}{\mu}$$

where $\mu = \mathbb{E}|\varepsilon| > 0$ and $\boldsymbol{\nu} = \mu\boldsymbol{\theta}$, it is enough to show that $\sqrt{n}(\widetilde{\mu}-\mu) = \mathrm{O}_{\mathbb{P}}(1)$ and $\sqrt{n}(\widetilde{\boldsymbol{\nu}}-\boldsymbol{\nu}) = \mathrm{O}_{\mathbb{P}}(1)$. Defining

$$\boldsymbol{Z} = \begin{pmatrix} |Y_1 - (\alpha + \boldsymbol{\beta}^\top \boldsymbol{X}_1)| \\ \vdots \\ |Y_n - (\alpha + \boldsymbol{\beta}^\top \boldsymbol{X}_n)| \end{pmatrix} = \mathfrak{X}\begin{pmatrix} \mu \\ \boldsymbol{\nu} \end{pmatrix} + \left[\mathfrak{X}\begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix}\right] \circ \boldsymbol{e},$$

where $\boldsymbol{e}^\top = (|\varepsilon_1| - \mathbb{E}|\varepsilon|, \ldots, |\varepsilon_n| - \mathbb{E}|\varepsilon|)$, and defining then $\widetilde{\boldsymbol{Z}}$ in the obvious way, we have

$$\begin{pmatrix} \widetilde{\mu} \\ \widetilde{\boldsymbol{\nu}} \end{pmatrix} = \left(\mathfrak{X}^\top\mathfrak{X}\right)^{-1}\mathfrak{X}^\top\widetilde{\boldsymbol{Z}}.$$

We therefore obtain

$$\begin{aligned} \begin{pmatrix} \sqrt{n}(\widetilde{\mu}-\mu) \\ \sqrt{n}(\widetilde{\boldsymbol{\nu}}-\boldsymbol{\nu}) \end{pmatrix} &= n\left(\mathfrak{X}^\top\mathfrak{X}\right)^{-1} \times \frac{1}{\sqrt{n}}\mathfrak{X}^\top\left\{\left[\mathfrak{X}\begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix}\right] \circ \boldsymbol{e}\right\} \\ &+ n\left(\mathfrak{X}^\top\mathfrak{X}\right)^{-1} \times \mathfrak{X}^\top\left(\frac{1}{\sqrt{n}}\left[\widetilde{\boldsymbol{Z}} - \boldsymbol{Z}\right]\right). \end{aligned} \tag{16}$$

Since $e = |\varepsilon| - \mathbb{E}|\varepsilon|$ is independent of $\boldsymbol{X}$ and has a finite variance, repeating the proof of (15) gives

$$n\left(\mathfrak{X}^\top\mathfrak{X}\right)^{-1} \times \frac{1}{\sqrt{n}}\mathfrak{X}^\top\left\{\left[\mathfrak{X}\begin{pmatrix} 1 \\ \boldsymbol{\theta} \end{pmatrix}\right] \circ \boldsymbol{e}\right\} = \mathrm{O}_{\mathbb{P}}(1). \tag{17}$$

Furthermore,

$$\mathfrak{X}^\top\left(\frac{1}{\sqrt{n}}\left[\widetilde{\boldsymbol{Z}} - \boldsymbol{Z}\right]\right) = \begin{pmatrix} n^{-1/2}\sum_{i=1}^n \left[\widetilde{Z}_i - Z_i\right] \\ n^{-1/2}\sum_{i=1}^n X_{i1}\left[\widetilde{Z}_i - Z_i\right] \\ \vdots \\ n^{-1/2}\sum_{i=1}^n X_{id}\left[\widetilde{Z}_i - Z_i\right] \end{pmatrix}.$$

Recalling that $\boldsymbol{X}$ lies in a compact set, we find that for any $m \in \{0, 1, \ldots, d\}$,

$$\left| n^{-1/2}\sum_{i=1}^n X_{im}\left[\widetilde{Z}_i - Z_i\right] \right| = \mathrm{O}_{\mathbb{P}}\left(\sqrt{n}\max_{1\le i\le n}|(\widetilde{\alpha}-\alpha) + (\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_i|\right) = \mathrm{O}_{\mathbb{P}}(1)$$

by (15). Combining this with (16), (17) and the convergence $n\left(\mathfrak{X}^\top\mathfrak{X}\right)^{-1} \xrightarrow{\mathbb{P}} \boldsymbol{\Sigma}_1^{-1}$, we get indeed $\sqrt{n}(\widetilde{\mu}-\mu) = \mathrm{O}_{\mathbb{P}}(1)$ and $\sqrt{n}(\widetilde{\boldsymbol{\nu}}-\boldsymbol{\nu}) = \mathrm{O}_{\mathbb{P}}(1)$ and thus

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}) = \mathrm{O}_{\mathbb{P}}(1). \tag{18}$$

We are now ready to prove the convergence of the weighted estimators $\widehat{\alpha}$, $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$. By standard weighted least squares theory,

$$\begin{pmatrix}\widehat{\alpha}\\ \widehat{\boldsymbol{\beta}}\end{pmatrix} = \left(\mathfrak{X}^\top \widetilde{\boldsymbol{\Omega}}^{-1}\mathfrak{X}\right)^{-1} \mathfrak{X}^\top \widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{Y}.$$

It follows that

$$\begin{pmatrix}\sqrt{n}(\widehat{\alpha}-\alpha)\\ \sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})\end{pmatrix} = n\left(\mathfrak{X}^\top \widetilde{\boldsymbol{\Omega}}^{-1}\mathfrak{X}\right)^{-1} \times \frac{1}{\sqrt{n}}\mathfrak{X}^\top \widetilde{\boldsymbol{\Omega}}^{-1}\left\{\left[\mathfrak{X}\begin{pmatrix}1\\ \boldsymbol{\theta}\end{pmatrix}\right]\circ\boldsymbol{\varepsilon}\right\} \tag{19}$$

where $\widetilde{\boldsymbol{\Omega}}$ is obtained from $\boldsymbol{\Omega}$ in the obvious manner. Note that for any $i,j\in\{0,\ldots,d\}$,

$$\frac{1}{n}\left[\mathfrak{X}^\top \boldsymbol{\Omega}^{-1}\mathfrak{X}\right]_{i+1,j+1} = \frac{1}{n}\sum_{k=1}^n \frac{X_{ki}X_{kj}}{\left[1+\boldsymbol{\theta}^\top \boldsymbol{X}_k\right]^2} \text{ and}$$

$$\frac{1}{n}\left[\mathfrak{X}^\top \widetilde{\boldsymbol{\Omega}}^{-1}\mathfrak{X}\right]_{i+1,j+1} = \frac{1}{n}\sum_{k=1}^n \frac{X_{ki}X_{kj}}{\left[1+\widetilde{\boldsymbol{\theta}}^\top \boldsymbol{X}_k\right]^2}.$$

Recalling once again that $\boldsymbol{X}$ lies in a compact set, that $1+\boldsymbol{\theta}^\top\boldsymbol{X}$ is bounded from below by a positive constant, and (18), we find, by the law of large numbers,

$$\frac{1}{n}\left[\mathfrak{X}^\top \widetilde{\boldsymbol{\Omega}}^{-1}\mathfrak{X}\right] - \frac{1}{n}\left[\mathfrak{X}^\top \boldsymbol{\Omega}^{-1}\mathfrak{X}\right] \xrightarrow{\mathbb{P}} 0 \text{ and thus } n\left(\mathfrak{X}^\top \widetilde{\boldsymbol{\Omega}}^{-1}\mathfrak{X}\right)^{-1} \xrightarrow{\mathbb{P}} \boldsymbol{\Sigma}_2^{-1}. \tag{20}$$

Besides, for any $m\in\{0,1,\ldots,d\}$,

$$\begin{aligned}&\left[\frac{1}{\sqrt{n}}\mathfrak{X}^\top \widetilde{\boldsymbol{\Omega}}^{-1}\left\{\left[\mathfrak{X}\begin{pmatrix}1\\ \boldsymbol{\theta}\end{pmatrix}\right]\circ\boldsymbol{\varepsilon}\right\} - \frac{1}{\sqrt{n}}\mathfrak{X}^\top \boldsymbol{\Omega}^{-1}\left\{\left[\mathfrak{X}\begin{pmatrix}1\\ \boldsymbol{\theta}\end{pmatrix}\right]\circ\boldsymbol{\varepsilon}\right\}\right]_{m+1}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[1+\boldsymbol{\theta}^\top\boldsymbol{X}_i\right] X_{im}\varepsilon_i \left[\frac{1}{\left[1+\widetilde{\boldsymbol{\theta}}^\top\boldsymbol{X}_i\right]^2} - \frac{1}{\left[1+\boldsymbol{\theta}^\top\boldsymbol{X}_i\right]^2}\right]\\
&= -\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^\top \left\{\frac{1}{n}\sum_{i=1}^n X_{im}\varepsilon_i \frac{2+\boldsymbol{\theta}^\top\boldsymbol{X}_i+\widetilde{\boldsymbol{\theta}}^\top\boldsymbol{X}_i}{\left[1+\widetilde{\boldsymbol{\theta}}^\top\boldsymbol{X}_i\right]^2\left[1+\boldsymbol{\theta}^\top\boldsymbol{X}_i\right]}\boldsymbol{X}_i\right\}.\end{aligned}$$

Using again the properties of $\boldsymbol{X}$ and (18), some straightforward algebra yields that

$$R_n := \sqrt{n}\max_{1\le i\le n}\left|\frac{2+\boldsymbol{\theta}^\top\boldsymbol{X}_i+\widetilde{\boldsymbol{\theta}}^\top\boldsymbol{X}_i}{\left[1+\widetilde{\boldsymbol{\theta}}^\top\boldsymbol{X}_i\right]^2} - \frac{2}{1+\boldsymbol{\theta}^\top\boldsymbol{X}_i}\right| = \mathrm{O}_{\mathbb{P}}(1).$$

Conclude that

$$\begin{aligned}&\left[\frac{1}{\sqrt{n}}\mathfrak{X}^\top \widetilde{\boldsymbol{\Omega}}^{-1}\left\{\left[\mathfrak{X}\begin{pmatrix}1\\ \boldsymbol{\theta}\end{pmatrix}\right]\circ\boldsymbol{\varepsilon}\right\} - \frac{1}{\sqrt{n}}\mathfrak{X}^\top \boldsymbol{\Omega}^{-1}\left\{\left[\mathfrak{X}\begin{pmatrix}1\\ \boldsymbol{\theta}\end{pmatrix}\right]\circ\boldsymbol{\varepsilon}\right\}\right]_{m+1}\\
&= -2\sqrt{n}\left(\widetilde{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^\top \left\{\frac{1}{n}\sum_{i=1}^n X_{im}\varepsilon_i \left[\frac{2}{[1+\boldsymbol{\theta}^\top\boldsymbol{X}_i]^2} + \mathrm{O}_{\mathbb{P}}\left(\frac{R_n}{\sqrt{n}}\right)\right]\boldsymbol{X}_i\right\}.\end{aligned}$$

Since ε is centred and independent of $\boldsymbol{X}$, we may combine the properties of $\boldsymbol{X}$ and (18) with the law of large numbers to get

$$\frac{1}{\sqrt{n}}\mathfrak{X}^\top \widetilde{\boldsymbol{\Omega}}^{-1}\left\{\left[\mathfrak{X}\begin{pmatrix}1\\ \boldsymbol{\theta}\end{pmatrix}\right]\circ\boldsymbol{\varepsilon}\right\} - \frac{1}{\sqrt{n}}\mathfrak{X}^\top \boldsymbol{\Omega}^{-1}\left\{\left[\mathfrak{X}\begin{pmatrix}1\\ \boldsymbol{\theta}\end{pmatrix}\right]\circ\boldsymbol{\varepsilon}\right\} = \mathrm{o}_{\mathbb{P}}(1). \tag{21}$$

Now clearly

$$\left[\frac{1}{\sqrt{n}}\mathfrak{X}^\top\boldsymbol{\Omega}^{-1}\left\{\left[\mathfrak{X}\begin{pmatrix}1\\ \boldsymbol{\theta}\end{pmatrix}\right]\circ\boldsymbol{\varepsilon}\right\}\right]_{m+1}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{X_{im}}{1+\boldsymbol{\theta}^\top\boldsymbol{X}_i}\varepsilon_i$$

so that, by the standard multivariate central limit theorem,

$$\frac{1}{\sqrt{n}}\mathfrak{X}^\top\boldsymbol{\Omega}^{-1}\left\{\left[\mathfrak{X}\begin{pmatrix}1\\ \boldsymbol{\theta}\end{pmatrix}\right]\circ\boldsymbol{\varepsilon}\right\}=\mathrm{O}_{\mathbb{P}}(1). \tag{22}$$

Combining (19), (20), (21) and (22) results in

$$\sqrt{n}(\widehat{\alpha}-\alpha)=\mathrm{O}_{\mathbb{P}}(1)\ \text{ and }\ \sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})=\mathrm{O}_{\mathbb{P}}(1).$$

We complete the proof by showing that $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta})=\mathrm{O}_{\mathbb{P}}(1)$. It is again enough to show that $\sqrt{n}(\widehat{\mu}-\mu)=\mathrm{O}_{\mathbb{P}}(1)$ and $\sqrt{n}(\widehat{\boldsymbol{\nu}}-\boldsymbol{\nu})=\mathrm{O}_{\mathbb{P}}(1)$. Write

$$\begin{aligned}\begin{pmatrix}\sqrt{n}(\widehat{\mu}-\mu)\\ \sqrt{n}(\widehat{\boldsymbol{\nu}}-\boldsymbol{\nu})\end{pmatrix}&=n\left(\mathfrak{X}^\top\widetilde{\boldsymbol{\Omega}}^{-1}\mathfrak{X}\right)^{-1}\times\frac{1}{\sqrt{n}}\mathfrak{X}^\top\widetilde{\boldsymbol{\Omega}}^{-1}\left\{\left[\mathfrak{X}\begin{pmatrix}1\\ \boldsymbol{\theta}\end{pmatrix}\right]\circ\boldsymbol{e}\right\}\\ &\quad+n\left(\mathfrak{X}^\top\widetilde{\boldsymbol{\Omega}}^{-1}\mathfrak{X}\right)^{-1}\times\mathfrak{X}^\top\widetilde{\boldsymbol{\Omega}}^{-1}\left(\frac{1}{\sqrt{n}}\left[\widehat{\boldsymbol{Z}}-\boldsymbol{Z}\right]\right).\end{aligned}$$

Furthermore,

$$\mathfrak{X}^\top\widetilde{\boldsymbol{\Omega}}^{-1}\left(\frac{1}{\sqrt{n}}\left[\widehat{\boldsymbol{Z}}-\boldsymbol{Z}\right]\right)=\begin{pmatrix}n^{-1/2}\sum_{i=1}^{n}\left[1+\widetilde{\boldsymbol{\theta}}^\top\boldsymbol{X}_i\right]^{-2}\left[\widehat{Z}_i-Z_i\right]\\ n^{-1/2}\sum_{i=1}^{n}X_{i1}\left[1+\widetilde{\boldsymbol{\theta}}^\top\boldsymbol{X}_i\right]^{-2}\left[\widehat{Z}_i-Z_i\right]\\ \vdots\\ n^{-1/2}\sum_{i=1}^{n}X_{id}\left[1+\widetilde{\boldsymbol{\theta}}^\top\boldsymbol{X}_i\right]^{-2}\left[\widehat{Z}_i-Z_i\right]\end{pmatrix}.$$

Recalling the properties of $\boldsymbol{X}$ and the $\sqrt{n}-$convergence of $\widehat{\alpha}$, $\widehat{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\theta}}$, we find that for any $m\in\{0,1,\ldots,d\}$,

$$\begin{aligned}\left|n^{-1/2}\sum_{i=1}^{n}X_{im}\left[1+\widetilde{\boldsymbol{\theta}}^\top\boldsymbol{X}_i\right]^{-2}\left[\widehat{Z}_i-Z_i\right]\right|&=\mathrm{O}_{\mathbb{P}}\left(\sqrt{n}\max_{1\le i\le n}|(\widehat{\alpha}-\alpha)+(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top\boldsymbol{X}_i|\right)\\ &=\mathrm{O}_{\mathbb{P}}(1).\end{aligned}$$

Combining this with (20) and straightforward adaptations of (21) and (22) with $\boldsymbol{e}$ in place of $\boldsymbol{\varepsilon}$, we find $\sqrt{n}(\widehat{\mu}-\mu)=\mathrm{O}_{\mathbb{P}}(1)$ and $\sqrt{n}(\widehat{\boldsymbol{\nu}}-\boldsymbol{\nu})=\mathrm{O}_{\mathbb{P}}(1)$ as required. □

Lemma C.2 is a general uniform consistency result which is useful for the analysis of the single-index model (M_2).

LEMMA C.2. *Assume that $(\mathcal{X}_i,\mathcal{Y}_i)_{i\ge1}$ are independent copies of a bivariate random pair $(\mathcal{X},\mathcal{Y})$ such that:*

- *$\mathcal{X}$ has support $[a,b]$, with $a<b$, and a density function $f_{\mathcal{X}}$ which is uniformly bounded on compact sub-intervals of (a,b).*
- *There exists $\delta>0$ such that $\mathbb{E}|\mathcal{Y}|^{2+\delta}<\infty$ and the conditional moment function $z\mapsto\mathbb{E}\left[|\mathcal{Y}|^{2+\delta}|\mathcal{X}=z\right]$ is uniformly bounded on compact sub-intervals of (a,b).*

Let further:

- $(\mathcal{V}_i)$ *be a sequence of independent copies of a bounded random variable* $\mathcal{V}$.
- L *be a Lipschitz continuous function with support contained in* $[-1,1]$.

Assume finally that $nh_n^5 \to c \in (0,\infty)$*, and* $t_n = n^t$ *with* $2/(5+\delta) < t < 2/5$*. Then for any* $a_1, b_1 \in [a,b]$ *with* $a < a_1 < b_1 < b$*,*

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \le z \le b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n \mathcal{Y}_i \mathbb{1}\{|\mathcal{Y}_i| \le t_n\} \mathcal{V}_i L\left(\frac{z-\mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E}\left[\mathcal{Y}\mathcal{V} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|$$
$$= \mathrm{O}_{\mathbb{P}}(1).$$

We note that, as a consequence, we have a similar uniform consistency result for the non-truncated version of the smoothed empirical moment, that is

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \le z \le b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n \mathcal{Y}_i \mathcal{V}_i L\left(\frac{z-\mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E}\left[\mathcal{Y}\mathcal{V} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right| = \mathrm{O}_{\mathbb{P}}(1)$$

under the further assumption $\mathbb{E}|\mathcal{Y}|^{5/2+\delta} < \infty$. This follows from noting that

$$\mathbb{P}\left(\bigcup_{i=1}^n \{|\mathcal{Y}_i| > t_n\}\right) \le n\mathbb{P}\left(|\mathcal{Y}| > t_n\right) = \mathrm{O}\left(\frac{n}{t_n^{5/2+\delta}}\right) = \mathrm{O}\left(n^{1-(5+2\delta)/(5+\delta)}\right) = \mathrm{o}(1)$$

by Markov's inequality. The stronger moment assumption $\mathbb{E}|\mathcal{Y}|^{5/2+\delta} < \infty$ already appears in [17] in the context of local polynomial estimation.

PROOF. The basic idea is to control the oscillation of the random function

$$z \mapsto \frac{n^{2/5}}{\sqrt{\log n}} \left| \frac{1}{nh_n} \sum_{i=1}^n \mathcal{Y}_i \mathbb{1}\{|\mathcal{Y}_i| \le t_n\} \mathcal{V}_i L\left(\frac{z-\mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E}\left[\mathcal{Y}\mathcal{V} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|$$

and then use this control to prove that it is sufficient to show uniform consistency over a fine grid instead, which can be done by using Bernstein's exponential inequality. Our proof adapts the method of [12] (proof of Theorem 2).

Define $\mathrm{Y}_i^{(n)} := \mathcal{Y}_i \mathcal{V}_i \mathbb{1}\{|\mathcal{Y}_i| \le t_n\}$ and $\mathrm{Y}^{(n)} := \mathcal{Y}\mathcal{V}\mathbb{1}\{|\mathcal{Y}| \le t_n\}$. Then

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \le z \le b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n \mathcal{Y}_i \mathbb{1}\{|\mathcal{Y}_i| \le t_n\} \mathcal{V}_i L\left(\frac{z-\mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E}\left[\mathcal{Y}\mathcal{V} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|$$
$$\le \frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \le z \le b_1} \left| \frac{1}{nh_n} \sum_{i=1}^n \mathrm{Y}_i^{(n)} L\left(\frac{z-\mathcal{X}_i}{h_n}\right) - \frac{1}{h_n} \mathbb{E}\left[\mathrm{Y}^{(n)} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|$$
$$+ \frac{n^{2/5}}{\sqrt{\log n}} \sup_{a_1 \le z \le b_1} \frac{1}{h_n} \mathbb{E}\left[|\mathcal{Y}|\,|\mathcal{V}|\mathbb{1}\{|\mathcal{Y}| > t_n\} \left|L\left(\frac{z-\mathcal{X}}{h_n}\right)\right|\right]. \tag{23}$$

The second term on the right-hand side of (23) is controlled by noting that, thanks to a change of variables,

$$\frac{1}{h_n} \mathbb{E}\left[|\mathcal{Y}|\,|\mathcal{V}|\mathbb{1}\{|\mathcal{Y}| > t_n\} \left|L\left(\frac{z-\mathcal{X}}{h_n}\right)\right|\right]$$
$$= \mathrm{O}\left(\int_{-1}^1 \mathbb{E}\left[|\mathcal{Y}|\mathbb{1}\{|\mathcal{Y}| > t_n\}|\mathcal{X} = z - h_n u\right] |L(u)|\, f_{\mathcal{X}}(z-h_n u)\, du\right)$$

$$= \mathrm{O}\left(t_n^{-1-\delta}\int_{-1}^{1}\mathbb{E}\left[|\mathcal{Y}|^{2+\delta}|\mathcal{X}=z-h_nu\right]|L(u)|\,f_{\mathcal{X}}(z-h_nu)\,du\right)=\mathrm{O}(t_n^{-1-\delta})$$

uniformly in $z\in[a_1,b_1]$. Here the boundedness of $\mathcal{V}$, the integrability of $|L|$ and the assumption that the $(2+\delta)-$conditional moment of $\mathcal{Y}$ and the density function $f_{\mathcal{X}}$ are uniformly bounded on compact sub-intervals of (a,b) were all used. Finally

$$t_n^{-1-\delta}=n^{-(1+\delta)t}=\mathrm{o}(n^{-2/5})=\mathrm{o}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right)$$

so that

$$\frac{n^{2/5}}{\sqrt{\log n}}\sup_{a_1\le z\le b_1}\frac{1}{h_n}\mathbb{E}\left[|\mathcal{Y}|\,|\mathcal{V}|\mathbb{1}\{|\mathcal{Y}|>t_n\}\left|L\left(\frac{z-\mathcal{X}}{h_n}\right)\right|\right]=\mathrm{o}(1). \tag{24}$$

Combining (23) and (24), we find that it is sufficient to show that

$$\frac{n^{2/5}}{\sqrt{\log n}}\sup_{a_1\le z\le b_1}\left|\frac{1}{nh_n}\sum_{i=1}^{n}\mathrm{Y}_i^{(n)}\,L\left(\frac{z-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[\mathrm{Y}^{(n)}\,L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|=\mathrm{O}_{\mathbb{P}}(1). \tag{25}$$

We now replace the supremum in (25) by a supremum over a grid by focusing on the oscillation of the left-hand side. For a given $z\in\mathbb{R}$, let

$$A_n(z):=\left\{z'\in[a_1,b_1]\;\middle|\;|z'-z|\le h_n\frac{\sqrt{\log n}}{n^{2/5}}\right\}.$$

Then $[a_1,b_1]$ is covered by the $A_n(z_{n,j})$, with

$$z_{n,j}=a_1+j\,h_n\frac{\sqrt{\log n}}{n^{2/5}},\ j=1,\ldots,\left\lfloor\frac{b_1-a_1}{h_n\frac{\sqrt{\log n}}{n^{2/5}}}\right\rfloor=:N_n,$$

where $\lfloor\cdot\rfloor$ denotes the floor function. Besides, writing $|L(z')-L(z)|\le C_L|z'-z|$ by Lipschitz continuity of L, we also find

$$|z'-z|\le 1\Rightarrow|L(z')-L(z)|\le|z'-z|\mathcal{L}(z)\ \text{ with } \mathcal{L}(z):=C_L\mathbb{1}\{|z|\le 2\}.$$

Let $z_{n,j}$ be a grid point and $z\in A_n(z_{n,j})$. By construction $|z-z_{n,j}|/h_n\le\sqrt{\log(n)}/n^{2/5}$ which converges to 0, so that, for n large enough,

$$\forall i\in\{1,\ldots,n\},\ \left|L\left(\frac{z-\mathcal{X}_i}{h_n}\right)-L\left(\frac{z_{n,j}-\mathcal{X}_i}{h_n}\right)\right|\le\frac{\sqrt{\log n}}{n^{2/5}}\mathcal{L}\left(\frac{z_{n,j}-\mathcal{X}_i}{h_n}\right).$$

Then

$$\begin{aligned}
&\frac{n^{2/5}}{\sqrt{\log n}}\sup_{z\in A_n(z_{n,j})}\left|\frac{1}{nh_n}\sum_{i=1}^{n}\mathrm{Y}_i^{(n)}\,L\left(\frac{z-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[\mathrm{Y}^{(n)}\,L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|\\
&\le\frac{n^{2/5}}{\sqrt{\log n}}\left|\frac{1}{nh_n}\sum_{i=1}^{n}\mathrm{Y}_i^{(n)}\,L\left(\frac{z_{n,j}-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[\mathrm{Y}^{(n)}\,L\left(\frac{z_{n,j}-\mathcal{X}}{h_n}\right)\right]\right|\\
&+\frac{1}{nh_n}\sum_{i=1}^{n}|\mathrm{Y}_i^{(n)}|\,\mathcal{L}\left(\frac{z_{n,j}-\mathcal{X}_i}{h_n}\right)+\frac{1}{h_n}\mathbb{E}\left[|\mathrm{Y}^{(n)}|\,\mathcal{L}\left(\frac{z_{n,j}-\mathcal{X}}{h_n}\right)\right]\\
&\le\frac{n^{2/5}}{\sqrt{\log n}}\left|\frac{1}{nh_n}\sum_{i=1}^{n}\mathrm{Y}_i^{(n)}\,L\left(\frac{z_{n,j}-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[\mathrm{Y}^{(n)}\,L\left(\frac{z_{n,j}-\mathcal{X}}{h_n}\right)\right]\right|
\end{aligned}$$

$$+\left|\frac{1}{nh_n}\sum_{i=1}^n |\mathrm{Y}_i^{(n)}|\,\mathcal{L}\left(\frac{z_{n,j}-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[|\mathrm{Y}^{(n)}|\,\mathcal{L}\left(\frac{z_{n,j}-\mathcal{X}}{h_n}\right)\right]\right|$$
$$+\,2\times\frac{1}{h_n}\mathbb{E}\left[|\mathrm{Y}^{(n)}|\,\mathcal{L}\left(\frac{z_{n,j}-\mathcal{X}}{h_n}\right)\right].$$

By the boundedness of $\mathcal{V}$, of $f_{\mathcal{X}}$ and of $z\mapsto\mathbb{E}\left[|\mathcal{Y}|\,|\mathcal{X}=z\right]$ over compact sub-intervals of (a,b), we find, for n large enough,

$$\sup_{a_1\le z\le b_1}\frac{1}{h_n}\mathbb{E}\left[|\mathrm{Y}^{(n)}|\,\mathcal{L}\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\le C_0$$

where C_0 is a finite constant. Consequently, for any constant $C>2C_0$,

$$\frac{n^{2/5}}{\sqrt{\log n}}\sup_{z\in A_n(z_{n,j})}\left|\frac{1}{nh_n}\sum_{i=1}^n \mathrm{Y}_i^{(n)}\,L\left(\frac{z-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[\mathrm{Y}^{(n)}\,L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|$$
$$\le\frac{n^{2/5}}{\sqrt{\log n}}\left|\frac{1}{nh_n}\sum_{i=1}^n \mathrm{Y}_i^{(n)}\,L\left(\frac{z_{n,j}-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[\mathrm{Y}^{(n)}\,L\left(\frac{z_{n,j}-\mathcal{X}}{h_n}\right)\right]\right|$$
$$+\frac{n^{2/5}}{\sqrt{\log n}}\left|\frac{1}{nh_n}\sum_{i=1}^n |\mathrm{Y}_i^{(n)}|\,\mathcal{L}\left(\frac{z_{n,j}-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[|\mathrm{Y}^{(n)}|\,\mathcal{L}\left(\frac{z_{n,j}-\mathcal{X}}{h_n}\right)\right]\right|+C$$

where the (crude) inequality $n^{2/5}/\sqrt{\log n}\ge 1$, for n large enough, was used. Conclude, by writing $[a_1,b_1]\subset\cup_{1\le j\le N_n}A_n(z_{n,j})$, that

$$\mathbb{P}\left(\frac{n^{2/5}}{\sqrt{\log n}}\sup_{a_1\le z\le b_1}\left|\frac{1}{nh_n}\sum_{i=1}^n \mathrm{Y}_i^{(n)}\,L\left(\frac{z-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[\mathrm{Y}^{(n)}\,L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|>3C\right)$$
$$\le N_n\max_{1\le j\le N_n}\mathbb{P}\left(\frac{n^{2/5}}{\sqrt{\log n}}\left|\frac{1}{nh_n}\sum_{i=1}^n \mathrm{Y}_i^{(n)}\,L\left(\frac{z_{n,j}-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[\mathrm{Y}^{(n)}\,L\left(\frac{z_{n,j}-\mathcal{X}}{h_n}\right)\right]\right|>C\right)$$
$$+N_n\max_{1\le j\le N_n}\mathbb{P}\left(\frac{n^{2/5}}{\sqrt{\log n}}\left|\frac{1}{nh_n}\sum_{i=1}^n |\mathrm{Y}_i^{(n)}|\,\mathcal{L}\left(\frac{z_{n,j}-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[|\mathrm{Y}^{(n)}|\,\mathcal{L}\left(\frac{z_{n,j}-\mathcal{X}}{h_n}\right)\right]\right|>C\right).$$

We finish the proof by showing

(26)
$$\mathbb{P}\left(\frac{n^{2/5}}{\sqrt{\log n}}\left|\frac{1}{nh_n}\sum_{i=1}^n \mathrm{Y}_i^{(n)}\,L\left(\frac{z-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[\mathrm{Y}^{(n)}\,L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|>C\right)=\mathrm{O}\left(\frac{1}{n}\right)$$

and

(27)
$$\mathbb{P}\left(\frac{n^{2/5}}{\sqrt{\log n}}\left|\frac{1}{nh_n}\sum_{i=1}^n |\mathrm{Y}_i^{(n)}|\,\mathcal{L}\left(\frac{z-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[|\mathrm{Y}^{(n)}|\,\mathcal{L}\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|>C\right)=\mathrm{O}\left(\frac{1}{n}\right)$$

for C large enough, uniformly in $z\in[a_1,b_1]$. Since N_n is of order $n^{2/5}/(h_n\sqrt{\log(n)})\approx n^{3/5}/\sqrt{\log(n)}=\mathrm{o}(n)$, this will entail

$$\mathbb{P}\left(\frac{n^{2/5}}{\sqrt{\log n}}\sup_{a_1\le z\le b_1}\left|\frac{1}{nh_n}\sum_{i=1}^n \mathrm{Y}_i^{(n)}\,L\left(\frac{z-\mathcal{X}_i}{h_n}\right)-\frac{1}{h_n}\mathbb{E}\left[\mathrm{Y}^{(n)}\,L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right|>3C\right)=\mathrm{o}(1)$$

for C large enough, which is sufficient for our purposes. We only show (26) uniformly in $z \in [a_1, b_1]$; the proof of (27) is identical. Rewrite the left-hand side of (26) as

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\left\{\mathrm{Y}_i^{(n)} L\left(\frac{z-\mathcal{X}_i}{h_n}\right) - \mathbb{E}\left[\mathrm{Y}^{(n)} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right\}\right| > Cu_n\right),$$

with $u_n := n^{3/5} h_n \sqrt{\log n}$. Let v be a constant such that $|\mathcal{V}| \leq v$ with probability 1. Note that for any i we have the crude bound

$$\left|\mathrm{Y}_i^{(n)} L\left(\frac{z-\mathcal{X}_i}{h_n}\right) - \mathbb{E}\left[\mathrm{Y}^{(n)} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right| \leq 2vt_n \max_{-1\leq u\leq 1} |L(u)|.$$

Remark also that, for n large enough,

$$\mathrm{Var}\left(\mathrm{Y}^{(n)} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right) \leq v^2 \mathbb{E}\left[\mathcal{Y}^2 L^2\left(\frac{z-\mathcal{X}}{h_n}\right)\right] \leq Dh_n$$

for some finite constant D, by uniform boundedness of $f_{\mathcal{X}}$ and $z \mapsto \mathbb{E}\left[\mathcal{Y}^2 \,|\mathcal{X} = z\right]$ over compact sub-intervals of (a,b). By the Bernstein exponential inequality we get

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\left\{\mathrm{Y}_i^{(n)} L\left(\frac{z-\mathcal{X}_i}{h_n}\right) - \mathbb{E}\left[\mathrm{Y}^{(n)} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right\}\right| > Cu_n\right)$$
$$\leq 2\exp\left(-\frac{C^2 u_n^2/2}{Dnh_n + 2Cvt_n u_n \max_{[-1,1]} |L|/3}\right).$$

Recalling that $t_n = n^t$ with $2/(5+\delta) < t < 2/5$, $u_n = n^{3/5} h_n \sqrt{\log n}$ and $nh_n^5 \to c \in (0,\infty)$, one finds

$$\frac{1}{\log n} \times \frac{C^2 u_n^2/2}{Dnh_n + 2Cvt_n u_n \max_{[-1,1]} |L|/3} \to \frac{c^{1/5} C^2}{2D} \quad \text{as} \quad n \to \infty$$

and therefore there is a constant $C' > 0$, independent of C, such that for n large enough

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\left\{\mathrm{Y}_i^{(n)} L\left(\frac{z-\mathcal{X}_i}{h_n}\right) - \mathbb{E}\left[\mathrm{Y}^{(n)} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right\}\right| > Cu_n\right) \leq 2\exp\left(-C' C^2 \log n\right)$$

uniformly in $z \in [a_1, b_1]$. For C large enough, this yields

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}\left\{\mathrm{Y}_i^{(n)} L\left(\frac{z-\mathcal{X}_i}{h_n}\right) - \mathbb{E}\left[\mathrm{Y}^{(n)} L\left(\frac{z-\mathcal{X}}{h_n}\right)\right]\right\}\right| > Cu_n\right) = \mathrm{O}\left(\frac{1}{n}\right)$$

which is equivalent to (26). This completes the proof. □

Lemma C.3 provides a uniform control, tailored to the assumptions of Proposition C.1, of the gap between smoothed moments and their asymptotic equivalents.

LEMMA C.3. *Assume that the bivariate random pair* $(\mathcal{X}, \mathcal{Y})$ *is such that:*

- $\mathcal{X}$ *has support* $[a,b]$*, with* $a < b$*, and a density function* $f_{\mathcal{X}}$ *which has a continuous derivative on* (a,b)*.*
- *The conditional moment function* $m_{\mathcal{Y}|\mathcal{X}} : z \mapsto \mathbb{E}(\mathcal{Y}|\mathcal{X} = z)$ *is well-defined and has a continuous derivative on* (a,b)*.*
- L *is a bounded measurable function with support contained in* $[-1,1]$*.*

Then, as $h \to 0$*:*

(i) *For any* $a_1, b_1 \in [a,b]$ *with* $a < a_1 < b_1 < b$, *we have, uniformly in* $z \in [a_1, b_1]$,

$$\frac{1}{h}\mathbb{E}\left[\mathcal{Y}\, L\left(\frac{z-\mathcal{X}}{h}\right)\right] = m_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z)\int_{-1}^{1} L(u)du$$
$$- h\{m'_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z) + m_{\mathcal{Y}|\mathcal{X}}(z) f'_{\mathcal{X}}(z)\}\int_{-1}^{1} uL(u)du + \mathrm{o}(h).$$

(ii) *If moreover* $f_{\mathcal{X}}$ *and* $m_{\mathcal{Y}|\mathcal{X}}$ *are twice continuously differentiable on* (a,b) *then, uniformly in* $z \in [a_1, b_1]$,

$$\frac{1}{h}\mathbb{E}\left[\mathcal{Y}\, L\left(\frac{z-\mathcal{X}}{h}\right)\right]$$
$$= m_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z)\int_{-1}^{1} L(u)du - h\{m'_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z) + m_{\mathcal{Y}|\mathcal{X}}(z) f'_{\mathcal{X}}(z)\}\int_{-1}^{1} uL(u)du$$
$$+ \frac{h^2}{2}\{m''_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z) + 2m'_{\mathcal{Y}|\mathcal{X}}(z) f'_{\mathcal{X}}(z) + m_{\mathcal{Y}|\mathcal{X}}(z) f''_{\mathcal{X}}(z)\}\int_{-1}^{1} u^2 L(u)du + \mathrm{o}(h^2).$$

PROOF. Note that

$$\frac{1}{h}\mathbb{E}\left[\mathcal{Y}\, L\left(\frac{z-\mathcal{X}}{h}\right)\right] = \int_{-1}^{1} m_{\mathcal{Y}|\mathcal{X}}(z-hu) f_{\mathcal{X}}(z-hu) L(u)\, du.$$

Parts (i) and (ii) are obtained by using the following Taylor formulae with integral remainder:

$$\varphi(z+\delta) = \varphi(z) + \delta\varphi'(z) + \int_z^{z+\delta} [\varphi'(t) - \varphi'(z)]\, dt$$

and

$$\varphi(z+\delta) = \varphi(z) + \delta\varphi'(z) + \frac{\delta^2}{2}\varphi''(z) + \int_z^{z+\delta} (z+\delta-t)[\varphi''(t) - \varphi''(z)]\, dt$$

applied to the function $\varphi : z \mapsto m_{\mathcal{Y}|\mathcal{X}}(z) f_{\mathcal{X}}(z)$. To get a uniform control of the remainders, use the fact that this function has uniformly continuous derivatives on any compact subinterval of $[a,b]$, by Heine's theorem. □

Our next auxiliary result is the uniform consistency (with rate) of the estimators of g and σ in the heteroscedastic single-index model of Section 3.2.

PROPOSITION C.1. *Assume that* $(\boldsymbol{X}_i, Y_i)_{i\geq 1}$ *are independent random pairs generated from the single-index model* (M_2). *Assume further that:*

- *The functions* g *and* $\sigma > 0$ *are continuous on* $K_{\boldsymbol{\beta}}$ *and twice continuously differentiable on the interior* $K^{\circ}_{\boldsymbol{\beta}}$ *of* $K_{\boldsymbol{\beta}}$.
- *The projection* $\boldsymbol{\beta}^\top \boldsymbol{X}$ *has a density function* $f_{\boldsymbol{\beta}^\top \boldsymbol{X}}$ *which is twice continuously differentiable and positive on* $K^{\circ}_{\boldsymbol{\beta}}$.
- *Each of the conditional moment functions* $z \mapsto \mathbb{E}(X_j | \boldsymbol{\beta}^\top \boldsymbol{X} = z)$, $j \in \{1, \ldots, d\}$ *is continuously differentiable on* $K^{\circ}_{\boldsymbol{\beta}}$.
- *There is* $\delta > 0$ *such that* $\mathbb{E}|\varepsilon|^{2+\delta} < \infty$.
- L *is a twice continuously differentiable and symmetric probability density function with support contained in* $[-1,1]$.

Assume also that $nh_n^5 \to c \in (0,\infty)$, *and* $t_n = n^t$ *with* $2/(5+\delta) < t < 2/5$. *Then, for any compact subset* K_0 *of* K^{o} *and any estimator* $\widehat{\boldsymbol{\beta}}$ *such that* $\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \mathrm{O}_{\mathbb{P}}(1)$, *we have*

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{\boldsymbol{x}\in K_0} \left|\widehat{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - g\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| = \mathrm{O}_{\mathbb{P}}(1)$$

$$\text{and} \quad \frac{n^{2/5}}{\sqrt{\log n}} \sup_{\boldsymbol{x}\in K_0} \left|\widehat{\sigma}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - \sigma\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| = \mathrm{O}_{\mathbb{P}}(1).$$

Before proving this result, note that when K is convex, its projection $K_{\boldsymbol{\beta}} := \{\boldsymbol{\beta}^\top \boldsymbol{x},\, \boldsymbol{x} \in K\}$, which is also the support of $\boldsymbol{\beta}^\top \boldsymbol{X}$, is a compact interval containing at least two points (because K has a nonempty interior). Note also that Proposition C.1 is tailored to our framework in the sense that the assumption $\mathbb{E}|\varepsilon|^{2+\delta} < \infty$, which puts a constraint on the tail heaviness of the noise variable, is intuitively close to minimal for the estimation of g and σ by estimators of Nadaraya-Watson type. An inspection of the proof reveals that a similar theorem holds if $\widehat{g}_{h_n,t_n}$ and $\widehat{\sigma}_{h_n,t_n}$ are replaced by non-truncated versions, under the stronger moment assumption $\mathbb{E}|\varepsilon|^{5/2+\delta} < \infty$; see the comment below the statement of Lemma C.2. The regularity assumption on $z \mapsto \mathbb{E}(X_j|\boldsymbol{\beta}^\top \boldsymbol{X} = z)$ is a technical requirement, which is for instance satisfied if the density function $f_{\boldsymbol{X}}$ is continuously differentiable and positive on K°.

PROOF. We start by proving the assertion on $\widehat{g}_{h_n,t_n}$. Define a truncated pseudo-Nadaraya-Watson estimator by

$$\widetilde{g}_{h_n,t_n}(z) = \sum_{i=1}^n Y_i \mathbb{1}\{|Y_i| \le t_n\}\, L\left(\frac{z - \boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n}\right) \Big/ \sum_{i=1}^n L\left(\frac{z - \boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n}\right).$$

The idea is to write

$$\begin{aligned}\left|\widehat{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - g\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| &\le \left|g\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - g\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| \\ &\quad + \left|\widetilde{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - g\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right)\right| \\ &\quad + \left|\widehat{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - \widetilde{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right)\right| \end{aligned} \tag{28}$$

and control each term on the right-hand side of (28) separately. To control the first term, we first apply the mean value theorem:

$$\left|g\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - g\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| \le \left|\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^\top \boldsymbol{x}\right| \times \sup_{\lambda\in[0,1]} \left|g'\left(\boldsymbol{\beta}^\top \boldsymbol{x} + \lambda\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^\top \boldsymbol{x}\right)\right|.$$

Since $K_0 \subset K^{\mathrm{o}}$, the distance between the compact set K_0 and the (compact) topological boundary of K is positive, *i.e.* $\rho := \inf\{\|x - y\|,\, x \in K_0,\, y \in K \setminus K^{\mathrm{o}}\} > 0$. It is then straightforward to show that, letting $K_{\boldsymbol{\beta}} = [u,v]$, we have $\boldsymbol{\beta}^\top \boldsymbol{x} \in [u + \rho/2, v - \rho/2]$ for any $\boldsymbol{x} \in K_0$. Since $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, we obtain that, with arbitrarily large probability as $n \to \infty$,

$$\forall \lambda \in [0,1],\ \forall \boldsymbol{x} \in K_0,\ \boldsymbol{\beta}^\top \boldsymbol{x} + \lambda\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right)^\top \boldsymbol{x} \in [u + \rho/4, v - \rho/4]. \tag{29}$$

Because g' is continuous and therefore bounded on compact intervals contained in (u,v), this gives

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{\boldsymbol{x}\in K_0} \left|g\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - g\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| = \mathrm{O}_{\mathbb{P}}\left(\frac{n^{2/5}}{\sqrt{\log n}} \times \frac{1}{\sqrt{n}}\right) = \mathrm{o}_{\mathbb{P}}(1). \tag{30}$$

To control the second term, we show the uniform consistency of the regression pseudo-estimator $\widetilde{g}_{h_n,t_n}$. The assumptions of Lemma C.2 are fulfilled for $(\mathcal{X},\mathcal{Y},\mathcal{V})=(\boldsymbol{\beta}^\top\boldsymbol{X},Y,1)=(\boldsymbol{\beta}^\top\boldsymbol{X},g\left(\boldsymbol{\beta}^\top\boldsymbol{X}\right)+\sigma\left(\boldsymbol{\beta}^\top\boldsymbol{X}\right)\varepsilon,1)$ and $(\mathcal{X},\mathcal{Y},\mathcal{V})=(\boldsymbol{\beta}^\top\boldsymbol{X},1,1)$. Recalling that ε is independent of $\boldsymbol{X}$ and centred, Lemma C.2 then provides

$$\widetilde{g}_{h_n,t_n}(z)=\frac{\dfrac{1}{h_n}\mathbb{E}\left[Y\,L\left(\dfrac{z-\boldsymbol{\beta}^\top\boldsymbol{X}}{h_n}\right)\right]+\mathrm{O}_{\mathbb{P}}\left(\dfrac{\sqrt{\log n}}{n^{2/5}}\right)}{\dfrac{1}{h_n}\mathbb{E}\left[L\left(\dfrac{z-\boldsymbol{\beta}^\top\boldsymbol{X}}{h_n}\right)\right]+\mathrm{O}_{\mathbb{P}}\left(\dfrac{\sqrt{\log n}}{n^{2/5}}\right)}$$

uniformly on any (fixed) compact subset of $K_{\boldsymbol{\beta}}^{\circ}=(u,v)$. Noting that $h_n\sim(c/n)^{1/5}$ and $\int_{-1}^{1}uL(u)du=0$ (because L is symmetric), Lemma C.3(ii) therefore entails

$$\widetilde{g}_{h_n,t_n}(z)=\frac{f_{\boldsymbol{\beta}^\top\boldsymbol{X}}(z)g(z)+\mathrm{O}_{\mathbb{P}}\left(\dfrac{\sqrt{\log n}}{n^{2/5}}\right)}{f_{\boldsymbol{\beta}^\top\boldsymbol{X}}(z)+\mathrm{O}_{\mathbb{P}}\left(\dfrac{\sqrt{\log n}}{n^{2/5}}\right)}=g(z)+\mathrm{O}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right)$$

uniformly on any compact subset of (u,v), the last equality being correct because $f_{\boldsymbol{\beta}^\top\boldsymbol{X}}$ is bounded from below by a positive constant on such sets. Together with (29) for $\lambda=1$, this yields

$$\frac{n^{2/5}}{\sqrt{\log n}}\sup_{\boldsymbol{x}\in K_0}\left|\widetilde{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top\boldsymbol{x}\right)-g\left(\widehat{\boldsymbol{\beta}}^\top\boldsymbol{x}\right)\right|=\mathrm{O}_{\mathbb{P}}(1). \tag{31}$$

We conclude by controlling the third term in the right-hand side of (28). The idea is to define $\mathfrak{Y}_i^{(n)}:=Y_i\mathbb{1}\{|Y_i|\le t_n\}$ and, for any z and $p=0,1$,

$$\widehat{m}_n^{(p)}(z):=\frac{1}{nh_n}\sum_{i=1}^{n}\left[\mathfrak{Y}_i^{(n)}\right]^p\,L\left(\frac{z-\widehat{\boldsymbol{\beta}}^\top\boldsymbol{X}_i}{h_n}\right)$$
$$\text{and}\ \ \widetilde{m}_n^{(p)}(z):=\frac{1}{nh_n}\sum_{i=1}^{n}\left[\mathfrak{Y}_i^{(n)}\right]^p\,L\left(\frac{z-\boldsymbol{\beta}^\top\boldsymbol{X}_i}{h_n}\right).$$

With this notation,

$$\begin{aligned}\widehat{g}_{h_n,t_n}(z)-\widetilde{g}_{h_n,t_n}(z)&=\frac{\widehat{m}_n^{(1)}(z)}{\widehat{m}_n^{(0)}(z)}-\frac{\widetilde{m}_n^{(1)}(z)}{\widetilde{m}_n^{(0)}(z)}\\&=\frac{[\widehat{m}_n^{(1)}(z)-\widetilde{m}_n^{(1)}(z)]\widetilde{m}_n^{(0)}(z)-[\widehat{m}_n^{(0)}(z)-\widetilde{m}_n^{(0)}(z)]\widetilde{m}_n^{(1)}(z)}{(\widetilde{m}_n^{(0)}(z)+[\widehat{m}_n^{(0)}(z)-\widetilde{m}_n^{(0)}(z)])\widetilde{m}_n^{(0)}(z)}.\end{aligned}\tag{32}$$

Since

$$\left|\widetilde{m}_n^{(0)}(z)-f_{\mathcal{X}}(z)\right|=\mathrm{o}_{\mathbb{P}}(1)\ \text{and}\ \left|\widetilde{m}_n^{(1)}(z)-f_{\mathcal{X}}(z)g(z)\right|=\mathrm{o}_{\mathbb{P}}(1) \tag{33}$$

uniformly on any compact subset of (u,v) by Lemmas C.2 and C.3(ii), we concentrate on differences of the form

$$\widehat{m}_n^{(p)}(z)-\widetilde{m}_n^{(p)}(z)=\frac{1}{nh_n}\sum_{i=1}^{n}\left[\mathfrak{Y}_i^{(n)}\right]^p\left\{L\left(\frac{z-\widehat{\boldsymbol{\beta}}^\top\boldsymbol{X}_i}{h_n}\right)-L\left(\frac{z-\boldsymbol{\beta}^\top\boldsymbol{X}_i}{h_n}\right)\right\}.$$

By Taylor's theorem with integral remainder applied to the function L, we find

$$\widehat{m}_n^{(p)}(z) - \widetilde{m}_n^{(p)}(z)$$

$$= -\frac{1}{nh_n}\sum_{i=1}^n \left[\mathfrak{Y}_i^{(n)}\right]^p \times \frac{(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top \boldsymbol{X}_i}{h_n} L'\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n}\right)$$

$$+ \frac{1}{nh_n}\sum_{i=1}^n \left[\mathfrak{Y}_i^{(n)}\right]^p \times \frac{1}{2}\left\{\frac{(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top \boldsymbol{X}_i}{h_n}\right\}^2 L''\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n}\right)$$

$$+ \frac{1}{nh_n}\sum_{i=1}^n \left[\mathfrak{Y}_i^{(n)}\right]^p \times \int_{(z-\boldsymbol{\beta}^\top \boldsymbol{X}_i)/h_n}^{(z-\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i)/h_n} \left(\frac{z-\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i}{h_n} - s\right)\left\{L''(s) - L''\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n}\right)\right\} ds$$

(34)
$$=: T_{1,n}(z) + T_{2,n}(z) + T_{3,n}(z).$$

We handle these three terms separately.

Control of $T_{1,n}(z)$: Note that

$$T_{1,n}(z) = -\frac{1}{h_n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top \left[\frac{1}{nh_n}\sum_{i=1}^n \left[\mathfrak{Y}_i^{(n)}\right]^p L'\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n}\right)\boldsymbol{X}_i\right].$$

Recall that $\boldsymbol{X}$ has compact support; Lemma C.2 (choosing $\mathcal{V} = X_j$, $1 \le j \le d$) then yields

$$T_{1,n}(z) = -\frac{1}{h_n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^\top \left[\frac{1}{h_n}\mathbb{E}\left(Y^p L'\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}}{h_n}\right)\boldsymbol{X}\right) + \mathrm{O}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right)\right]$$

uniformly on any compact subset of (u, v). Because for any $j \in \{1, \ldots, d\}$,

$$\mathbb{E}(Y^p X_j | \boldsymbol{\beta}^\top \boldsymbol{X} = z) = [\mathbb{1}\{p=0\} + g(z)\mathbb{1}\{p=1\}]\,\mathbb{E}(X_j | \boldsymbol{\beta}^\top \boldsymbol{X} = z),$$

the conditional moment function $z \mapsto \mathbb{E}(Y^p X_j | \boldsymbol{\beta}^\top \boldsymbol{X} = z)$ satisfies the regularity requirements of Lemma C.3(i). By Lemma C.3(i) and the symmetry of L,

$$\frac{1}{h_n}\mathbb{E}\left(Y^p L'\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}}{h_n}\right)\boldsymbol{X}\right) = \mathrm{O}(h_n)$$

uniformly on any compact subset of (u, v). Since $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathrm{O}_{\mathbb{P}}(1/\sqrt{n})$, this yields

(35) $$T_{1,n}(z) = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) = \mathrm{o}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right) \quad \text{uniformly on any compact subset of } (u, v).$$

Control of $T_{2,n}(z)$: Recall that $\boldsymbol{X}$ has compact support, $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathrm{O}_{\mathbb{P}}(1/\sqrt{n})$, and L'' is bounded to obtain, using the law of large numbers,

(36)
$$\sup_{z\in\mathbb{R}} |T_{2,n}(z)| = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{nh_n^3} \times \frac{1}{n}\sum_{i=1}^n |Y_i|^p\right) = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{nh_n^3}\right) = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{n^{2/5}}\right) = \mathrm{o}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right).$$

Control of $T_{3,n}(z)$: Use a change of variables to rewrite the integral term in $T_{3,n}(z)$ as

$$\int_{(z-\boldsymbol{\beta}^\top \boldsymbol{X}_i)/h_n}^{(z-\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i)/h_n} \left(\frac{z-\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i}{h_n} - s\right)\left\{L''(s) - L''\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n}\right)\right\} ds$$

$$= \int_0^{(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^\top \boldsymbol{X}_i/h_n} \left(\frac{(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^\top \boldsymbol{X}_i}{h_n} - u\right)\left\{L''\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n} + u\right) - L''\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n}\right)\right\} du.$$

Since $\boldsymbol{X}$ has compact support and $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathrm{O}_{\mathbb{P}}(1/\sqrt{n})$ we have

$$\max_{1\le i\le n} \left| \frac{(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^\top \boldsymbol{X}_i}{h_n} \right| = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{h_n\sqrt{n}} \right) = \mathrm{o}_{\mathbb{P}}(1).$$

By uniform continuity of the continuous and compactly supported function L'', it follows that

$$\max_{1\le i\le n} \sup_{z\in\mathbb{R}} \sup_{|u|\le |(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^\top \boldsymbol{X}_i|/h_n} \left| L''\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n} + u \right) - L''\left(\frac{z-\boldsymbol{\beta}^\top \boldsymbol{X}_i}{h_n} \right) \right| = \mathrm{o}_{\mathbb{P}}(1).$$

We then get

$$\begin{aligned}
\sup_{z\in\mathbb{R}} |T_{3,n}(z)| &= \mathrm{o}_{\mathbb{P}}\left(\frac{1}{nh_n} \sum_{i=1}^n |Y_i|^p \left| \int_0^{(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^\top \boldsymbol{X}_i/h_n} \left| \frac{(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^\top \boldsymbol{X}_i}{h_n} - u \right| du \right| \right) \\
&= \mathrm{o}_{\mathbb{P}}\left(\frac{1}{nh_n} \sum_{i=1}^n |Y_i|^p \left[\frac{(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}})^\top \boldsymbol{X}_i}{h_n} \right]^2 \right) \\
&= \mathrm{o}_{\mathbb{P}}\left(\frac{1}{nh_n^3} \times \frac{1}{n} \sum_{i=1}^n |Y_i|^p \right) = \mathrm{O}_{\mathbb{P}}\left(\frac{1}{nh_n^3} \right) = \mathrm{o}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}} \right). \qquad (37)
\end{aligned}$$

Combine (32), (33), (34), (35), (36) and (37) to obtain

$$\widehat{g}_{h_n,t_n}(z) - \widetilde{g}_{h_n,t_n}(z) = \frac{\mathrm{o}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}} \right)}{f_{\boldsymbol{\beta}^\top \boldsymbol{X}}(z) + \mathrm{o}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}} \right)} = \mathrm{o}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}} \right)$$

uniformly on any compact subset of (u,v). Using (29) again with $\lambda = 1$, we get

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{\boldsymbol{x}\in K_0} \left| \widehat{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} \right) - \widetilde{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} \right) \right| = \mathrm{o}_{\mathbb{P}}(1). \qquad (38)$$

Combining (28), (30), (31) and (38) concludes the proof of the assertion on $\widehat{g}_{h_n,t_n}$.

We turn to the control of $\widehat{\sigma}_{h_n,t_n}$, where the added difficulty is that the computation of the estimator is based on the absolute residuals $\widehat{Z}_{i,h_n,t_n} = \left| Y_i - \widehat{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i \right) \right|$ rather than on the "true values" $Z_i := \left| Y_i - g\left(\boldsymbol{\beta}^\top \boldsymbol{X}_i \right) \right|$. We thus introduce its pseudo-estimator analogue based on the Z_i,

$$\overline{\sigma}_{h_n,t_n}(z) := \sum_{i=1}^n Z_i \mathbb{1}\left\{ Z_i \le t_n \right\} L\left(\frac{z-\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i}{h_n} \right) \Bigg/ \sum_{i=1}^n L\left(\frac{z-\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i}{h_n} \right)$$

and we seek to control $|\widehat{\sigma}_{h_n,t_n}(z) - \overline{\sigma}_{h_n,t_n}(z)|$, for $z = \widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}$, uniformly in $\boldsymbol{x} \in K_0$. Write

$$\begin{aligned}
&\widehat{\sigma}_{h_n,t_n}(z) - \overline{\sigma}_{h_n,t_n}(z) \\
&= \sum_{i=1}^n \left[\widehat{Z}_{i,h_n,t_n} \mathbb{1}\left\{ \widehat{Z}_{i,h_n,t_n} \le t_n \right\} - Z_i \mathbb{1}\left\{ Z_i \le t_n \right\} \right] L\left(\frac{z-\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i}{h_n} \right) \Bigg/ \sum_{i=1}^n L\left(\frac{z-\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i}{h_n} \right).
\end{aligned}$$

Note that the only pairs $(\boldsymbol{X}_i, Y_i)$ making a nonzero contribution to this difference are those for which $|z - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i| \le h_n$. For $\boldsymbol{x} \in K_0$, we thus focus on controlling

$$\sup_{\boldsymbol{x}\in K_0} \left| \widehat{Z}_{i,h_n,t_n} \mathbb{1}\left\{ \widehat{Z}_{i,h_n,t_n} \le t_n \right\} - Z_i \mathbb{1}\left\{ Z_i \le t_n \right\} \right| \mathbb{1}\left\{ \left| \widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i \right| \le h_n \right\}.$$

Since $\left|\widehat{Z}_{i,h_n,t_n} - Z_i\right| \leq \left|\widehat{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i\right) - g\left(\boldsymbol{\beta}^\top \boldsymbol{X}_i\right)\right|$, the triangle inequality yields

$$\sup_{\boldsymbol{x}\in K_0} \left|\widehat{Z}_{i,h_n,t_n}\mathbb{1}\left\{\widehat{Z}_{i,h_n,t_n} \leq t_n\right\} - Z_i\mathbb{1}\left\{Z_i \leq t_n\right\}\right| \mathbb{1}\left\{\left|\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i\right| \leq h_n\right\}$$

$$\leq \sup_{\boldsymbol{x}\in K_0} \max_{i:\,|\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i| \leq h_n} \left|\widehat{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i\right) - g\left(\boldsymbol{\beta}^\top \boldsymbol{X}_i\right)\right| \tag{39}$$

$$+ \sup_{\boldsymbol{x}\in K_0} Z_i \left|\mathbb{1}\left\{\widehat{Z}_{i,h_n,t_n} \leq t_n\right\} - \mathbb{1}\left\{Z_i \leq t_n\right\}\right| \mathbb{1}\left\{\left|\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i\right| \leq h_n\right\}. \tag{40}$$

We focus on (39) first, where the idea is to use our uniform convergence result on $\widehat{g}_{h_n,t_n}$. Write

$$|\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i| \leq h_n \Rightarrow |\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \boldsymbol{\beta}^\top \boldsymbol{X}_i| \leq h_n + |(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^\top \boldsymbol{X}_i| = h_n + \mathrm{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$$

irrespective of the index i and $\boldsymbol{x} \in K_0$, so that, with arbitrarily large probability as $n \to \infty$,

$$\forall i \in \{1, \ldots, n\},\ \forall \boldsymbol{x} \in K_0,\ |\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i| \leq h_n \Rightarrow |\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \boldsymbol{\beta}^\top \boldsymbol{X}_i| \leq 2h_n.$$

Recall that, by (29), $\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} \in [u + \rho/4, v - \rho/4]$ with arbitrarily large probability as $n \to \infty$, irrespective of $\boldsymbol{x} \in K_0$. Since $h_n \to 0$, this yields, with arbitrarily large probability as $n \to \infty$,

$$\forall i \in \{1, \ldots, n\},\ \forall \boldsymbol{x} \in K_0,\ |\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i| \leq h_n \Rightarrow \boldsymbol{\beta}^\top \boldsymbol{X}_i \in [u + \rho/8, v - \rho/8].$$

In other words, for such indices i, $\boldsymbol{X}_i$ belongs to the intersection of K and the inverse image of the closed interval $[u + \rho/8, v - \rho/8]$ by the (continuous) projection mapping $\boldsymbol{x} \mapsto \boldsymbol{\beta}^\top \boldsymbol{x}$. This intersection is itself a compact set K_1, say, and therefore, with arbitrarily large probability as $n \to \infty$,

$$\forall i \in \{1, \ldots, n\},\ \forall \boldsymbol{x} \in K_0,\ |\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i| \leq h_n \Rightarrow \boldsymbol{X}_i \in K_1.$$

Note also that $K_1 \subset K^\circ$ since K_1 is contained in the (open) inverse image of the open interval $(u + \rho/16, v - \rho/16)$ by the same projection mapping. It then follows from our uniform convergence result on $\widehat{g}_{h_n,t_n}$ that

$$\sup_{\boldsymbol{x}\in K_0} \max_{i:\,|\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i| \leq h_n} \left|\widehat{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i\right) - g\left(\boldsymbol{\beta}^\top \boldsymbol{X}_i\right)\right| = \mathrm{O}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right). \tag{41}$$

We can now control (40). Clearly

$$\begin{aligned}
&\left|\mathbb{1}\left\{\widehat{Z}_{i,h_n,t_n} \leq t_n\right\} - \mathbb{1}\left\{Z_i \leq t_n\right\}\right| \\
&= \mathbb{1}\left\{\widehat{Z}_{i,h_n,t_n} \leq t_n,\, Z_i > t_n\right\} + \mathbb{1}\left\{\widehat{Z}_{i,h_n,t_n} > t_n,\, Z_i \leq t_n\right\}.
\end{aligned}$$

Recall that $\left|\widehat{Z}_{i,h_n,t_n} - Z_i\right| \leq \left|\widehat{g}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i\right) - g\left(\boldsymbol{\beta}^\top \boldsymbol{X}_i\right)\right|$ and use (41) together with the assumption $t_n \to \infty$ to find that, with arbitrarily large probability as $n \to \infty$,

$$\forall i \in \{1, \ldots, n\},\ \sup_{\boldsymbol{x}\in K_0} Z_i \left|\mathbb{1}\left\{\widehat{Z}_{i,h_n,t_n} \leq t_n\right\} - \mathbb{1}\left\{Z_i \leq t_n\right\}\right| \mathbb{1}\left\{\left|\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} - \widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i\right| \leq h_n\right\}$$

$$\leq Z_i \mathbb{1}\left\{Z_i \leq 2t_n,\, Z_i > t_n\right\} + Z_i \mathbb{1}\left\{Z_i > t_n/2,\, Z_i \leq t_n\right\}$$

$$\leq Z_i \mathbb{1}\left\{t_n/2 < Z_i \leq 2t_n\right\}. \tag{42}$$

Combine (41) and (42) to obtain, with arbitrarily large probability as $n \to \infty$,

$$\sup_{\boldsymbol{x}\in K_0}\left|\widehat{\sigma}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - \overline{\sigma}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right)\right|$$

$$\leq \sup_{\boldsymbol{x}\in K_0}\left[\overline{\sigma}_{h_n,2t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - \overline{\sigma}_{h_n,t_n/2}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right)\right] + \mathrm{O}_{\mathbb{P}}\left(\frac{\sqrt{\log n}}{n^{2/5}}\right). \tag{43}$$

To conclude, note that since $\mathbb{E}|\varepsilon| = 1$,

$$Z := \left|Y - g\left(\boldsymbol{\beta}^\top \boldsymbol{X}\right)\right| = \sigma\left(\boldsymbol{\beta}^\top \boldsymbol{X}\right) + \sigma\left(\boldsymbol{\beta}^\top \boldsymbol{X}\right)(|\varepsilon| - \mathbb{E}|\varepsilon|).$$

This single-index model linking Z to $\boldsymbol{X}$ has the same structure as model (M_2) and satisfies our assumptions, with g replaced by σ and ε replaced by $|\varepsilon| - \mathbb{E}|\varepsilon|$. Since for this model $\overline{\sigma}_{h_n,t_n}$ plays the role of $\widehat{g}_{h_n,t_n}$, we can use the first part of the Proposition to get

$$\frac{n^{2/5}}{\sqrt{\log n}} \sup_{\boldsymbol{x}\in K_0}\left|\overline{\sigma}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - \sigma\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| = \mathrm{O}_{\mathbb{P}}(1). \tag{44}$$

The result then follows by using (43) to write

$$\begin{aligned}\frac{n^{2/5}}{\sqrt{\log n}} \sup_{\boldsymbol{x}\in K_0}\left|\widehat{\sigma}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - \sigma\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| &\leq \frac{n^{2/5}}{\sqrt{\log n}} \sup_{\boldsymbol{x}\in K_0}\left|\overline{\sigma}_{h_n,2t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - \sigma\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| \\ &+ \frac{n^{2/5}}{\sqrt{\log n}} \sup_{\boldsymbol{x}\in K_0}\left|\overline{\sigma}_{h_n,t_n/2}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - \sigma\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| \\ &+ \frac{n^{2/5}}{\sqrt{\log n}} \sup_{\boldsymbol{x}\in K_0}\left|\overline{\sigma}_{h_n,t_n}\left(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}\right) - \sigma\left(\boldsymbol{\beta}^\top \boldsymbol{x}\right)\right| \\ &+ \mathrm{O}_{\mathbb{P}}(1)\end{aligned}$$

and then by using (44) as well as its analogues with t_n replaced by $t_n/2$ and $2t_n$. □

The following de-conditioning lemma is a stronger version of Lemma 8 in [18].

LEMMA C.4. *Let $N = N(n) \xrightarrow{\mathbb{P}} \infty$ be a random sequence of integers that, for each n, takes its values in $\{0, 1, \ldots, n\}$. Suppose that (G_n) and (H_m) are sequences of random elements taking values in a metric space S endowed with its Borel $\sigma-$field. Assume that*

$$\forall n \geq 1,\ \forall m \in \{1, \ldots, n\},\ G_n \,|\, \{N(n) = m\} \stackrel{d}{=} H_m.$$

Then:

(i) *If $H_m \xrightarrow{d} H$ as $m \to \infty$, we have $G_n \xrightarrow{d} H$ as $n \to \infty$.*

If moreover S is a linear space endowed with a norm $\|\cdot\|$, then:

(ii) *If $\|H_m\| = \mathrm{O}_{\mathbb{P}}(1)$, we have $\|G_n\| = \mathrm{O}_{\mathbb{P}}(1)$.*

Finally, in the case $S = \mathbb{R}$:

(iii) *If $H_m \xrightarrow{\mathbb{P}} +\infty$ as $m \to \infty$, we have $G_n \xrightarrow{\mathbb{P}} +\infty$ as $n \to \infty$.*

PROOF. Use the law of total probability to write, for any positive integer m_0 and any Borel subset A of S,

$$\mathbb{P}(G_n \in A) = \mathbb{P}(G_n \in A, N(n) \le m_0) + \sum_{m=m_0+1}^{n} \mathbb{P}(G_n \in A \,|\, N(n) = m)\mathbb{P}(N(n) = m)$$

$$= \mathbb{P}(G_n \in A, N(n) \le m_0) + \sum_{m=m_0+1}^{n} \mathbb{P}(H_m \in A)\mathbb{P}(N(n) = m). \tag{45}$$

To show (i), let A be a continuity set of H (in the sense that $\mathbb{P}(H \in \partial A) = 0$, where ∂A is the topological boundary of A). By the Portmanteau theorem, there is an integer m_0 such that for $m > m_0$, $|\mathbb{P}(H_m \in A) - \mathbb{P}(H \in A)| \le \varepsilon/3$. With this choice of m_0 we have, for n large enough,

$$|\mathbb{P}(G_n \in A) - \mathbb{P}(H \in A)|$$

$$\le \mathbb{P}(G_n \in A, N(n) \le m_0) + \mathbb{P}(H \in A)\mathbb{P}(N(n) \le m_0) + \frac{\varepsilon}{3} \sum_{m=m_0+1}^{n} \mathbb{P}(N(n) = m)$$

$$\le \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

This proves (i). To show statements (ii) and (iii), deduce from (45) that for any m_0,

$$\mathbb{P}(G_n \in A) \le \sup_{m > m_0} \mathbb{P}(H_m \in A) + \mathrm{o}(1) \text{ as } n \to \infty.$$

Fix $\varepsilon > 0$. To prove (ii), let $C > 0$ and m_0 be such that $\mathbb{P}(\|H_m\| > C) \le \varepsilon/2$ for any $m > m_0$, and apply the above inequality with A being the complement of the closed ball with centre the origin and radius C along with this choice of m_0 to get $\mathbb{P}(\|G_n\| > C) \le \varepsilon$ for n large enough, which is the desired result. Finally, to prove (iii), pick an arbitrary t and set $A = A_t = (-\infty, t]$. There is an integer m_0 such that $\mathbb{P}(H_m \in A_t) \le \varepsilon/2$ for $m > m_0$; applying the above inequality with this choice of m_0 yields $\mathbb{P}(G_n \in A_t) \le \varepsilon$ for n large enough, which is (iii). □

Our next result is a technical extension of Theorem 2.1 to the case when the sample size n is random. This will be key to the proof of our main theorems in Sections 3.2 and 3.3, where one has to work with a selected subset of observations whose size N is indeed random.

LEMMA C.5. *Assume that there is $\delta > 0$ such that $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$, that ε satisfies condition $\mathcal{C}_1(\gamma)$ with $0 < \gamma < 1/2$ and $\tau_n \uparrow 1$ is such that $n(1-\tau_n) \to \infty$. Let $N = N(n) \xrightarrow{\mathbb{P}} \infty$ be a random sequence of integers that, for each n, takes its values in $\{0, 1, \ldots, n\}$. Suppose that, for any n and on the event $\{N > 0\}$, $\widehat{\varepsilon}_i^{(n)}$ and $\varepsilon_i^{(n)}$, $1 \le i \le N$ are given such that*

- *For any $n \ge 1$ and any $m \in \{1, \ldots, n\}$, the distribution of $(\varepsilon_1^{(n)}, \ldots, \varepsilon_N^{(n)})$ given $N = m$ is the distribution of m independent copies of ε,*
- *We have*

$$\sqrt{N(1-\tau_N)} \max_{1 \le i \le N} \frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i^{(n)}|}{1 + |\varepsilon_i^{(n)}|} \xrightarrow{\mathbb{P}} 0.$$

Let finally $\widehat{\xi}_{\tau_N}(\varepsilon) = \arg\min_{u \in \mathbb{R}} \sum_{i=1}^{N} \eta_{\tau_N}(\widehat{\varepsilon}_i^{(n)} - u)$ on $\{N > 0\}$ and 0 otherwise, as well as

$$\psi_N(u) = \frac{1}{2\xi_{\tau_N}^2(\varepsilon)} \sum_{i=1}^{N} \left[\eta_{\tau_N}\left(\varepsilon_i^{(n)} - \xi_{\tau_N}(\varepsilon) - \frac{u\xi_{\tau_N}(\varepsilon)}{\sqrt{N(1-\tau_N)}} \right) - \eta_{\tau_N}(\varepsilon_i^{(n)} - \xi_{\tau_N}(\varepsilon)) \right]$$

and $\chi_N(u) = \dfrac{1}{2\xi_{\tau_N}^2(\varepsilon)} \displaystyle\sum_{i=1}^{N} \left[\eta_{\tau_N} \left(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_N}(\varepsilon) - \frac{u\xi_{\tau_N}(\varepsilon)}{\sqrt{N(1-\tau_N)}} \right) - \eta_{\tau_N}(\widehat{\varepsilon}_i^{(n)} - \xi_{\tau_N}(\varepsilon)) \right]$

on $\{N > 0\}$*, and 0 otherwise. Then we have* $\chi_N(u) - \psi_N(u) \xrightarrow{\mathbb{P}} 0$ *as* $n \to \infty$ *and*

$$\sqrt{N(1-\tau_N)} \left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\gamma^3}{1-2\gamma}\right).$$

PROOF. To show that $\chi_N(u) - \psi_N(u) \xrightarrow{\mathbb{P}} 0$, following the ideas of the proof of Theorem 2.1, it is enough to prove that

$$T_{1,N} = \frac{\sqrt{1-\tau_N}}{\xi_{\tau_N}(\varepsilon)\sqrt{N}} \sum_{i=1}^{N} |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i^{(n)}| \xrightarrow{\mathbb{P}} 0 \tag{46}$$

and that, if $I_N(u) = [0, |u|\xi_{\tau_N}(\varepsilon)/\sqrt{N(1-\tau_N)}]$,

$$\begin{aligned} T_{2,N}(u) = {} & \frac{2}{\xi_{\tau_N}(\varepsilon)\sqrt{N(1-\tau_N)}} \\ & \times \sum_{i=1}^{N} \sup_{|t| \in I_N(u)} |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i^{(n)}| \mathbb{1}\{\varepsilon_i^{(n)} - \xi_{\tau_N}(\varepsilon) - t > \min(\varepsilon_i^{(n)} - \widehat{\varepsilon}_i^{(n)}, 0)\} \\ & \xrightarrow{\mathbb{P}} 0. \end{aligned} \tag{47}$$

Clearly, since $N = N(n) \xrightarrow{\mathbb{P}} \infty$ and in particular $N > 0$ with arbitrarily large probability,

$$T_{1,N} = \mathrm{o}_{\mathbb{P}}\left(\frac{1}{N} \sum_{i=1}^{N} (1 + |\varepsilon_i^{(n)}|) \right) = \mathrm{o}_{\mathbb{P}}(1)$$

where the law of large numbers is combined with the de-conditioning Lemma C.4(i), to show that $N^{-1}\sum_{i=1}^{N}(1 + |\varepsilon_i^{(n)}|) \xrightarrow{\mathbb{P}} 1 + \mathbb{E}|\varepsilon| < \infty$. This proves (46). We now turn to the control of $T_{2,N}(u)$. Use that $N = N(n) \xrightarrow{\mathbb{P}} \infty$ and follow the ideas leading to (11) in the proof of Theorem 2.1 to find, for n large enough,

$$\varepsilon_i^{(n)} - \xi_{\tau_N}(\varepsilon) - t > \min(\varepsilon_i^{(n)} - \widehat{\varepsilon}_i^{(n)}, 0) \Rightarrow \varepsilon_i^{(n)} > \frac{1}{6}\xi_{\tau_N}(\varepsilon)$$

with arbitrarily large probability, irrespective of $i \in \{1, \ldots, N\}$ and t such that $|t| \in I_N(u)$. Therefore, with arbitrarily large probability as $n \to \infty$:

$$\begin{aligned} T_{2,N}(u) &\le \frac{2}{\xi_{\tau_N}(\varepsilon)\sqrt{N(1-\tau_N)}} \sum_{i=1}^{N} |\widehat{\varepsilon}_i^{(n)} - \varepsilon_i^{(n)}| \mathbb{1}\left\{\varepsilon_i^{(n)} > \frac{1}{6}\xi_{\tau_N}(\varepsilon)\right\} \\ &= \mathrm{o}_{\mathbb{P}}\left(\frac{1}{N\xi_{\tau_N}(\varepsilon)(1-\tau_N)} \sum_{i=1}^{N} \varepsilon_i^{(n)} \mathbb{1}\left\{\varepsilon_i^{(n)} > \frac{1}{6}\xi_{\tau_N}(\varepsilon)\right\} \right). \end{aligned}$$

Combine Lemma A.1 with the de-conditioning Lemma C.4(i) to get

$$T_{2,N}(u) = \mathrm{o}_{\mathbb{P}}(1).$$

This is (47). Combine (46) and (47) to get $\chi_N(u) - \psi_N(u) \xrightarrow{\mathbb{P}} 0$. Now a combination of the conclusion of the proof of Theorem 2 in [4] and the de-conditioning Lemma C.4(i) yields

$$\chi_N(u) = \psi_N(u) + \mathrm{o}_{\mathbb{P}}(1) \xrightarrow{d} -uZ\sqrt{\frac{2\gamma}{1-2\gamma}} + \frac{u^2}{2\gamma} \text{ as } n \to \infty$$

in the sense of finite-dimensional convergence, with Z being standard Gaussian. Since $\chi_N(u)$ is convex in u, the conclusion follows using the convexity lemma stated as Theorem 5 in [15]. □

Lemma C.6(i) below is a technical extension of Lemma A.3 to the case of a random sample size. It is essential in, among others, proving that the Hill estimator based on a random number of residuals is asymptotically Gaussian, which is stated below as Lemma C.6(ii); this will be used extensively in Sections 3.2 and 3.3.

LEMMA C.6. *Let $k = k(n) \to \infty$ be a sequence of integers with $k/n \to 0$. Assume that ε has an infinite right endpoint. Let $N = N(n) \xrightarrow{\mathbb{P}} \infty$ be a random sequence of integers that, for each n, takes its values in $\{0, 1, \ldots, n\}$. Suppose that, for any n and on the event $\{N > 0\}$, $\widehat{\varepsilon}_i^{(n)}$ and $\varepsilon_i^{(n)}$, $1 \le i \le N$ are given such that*

- *For any $n \ge 1$ and any $m \in \{1, \ldots, n\}$, the distribution of $(\varepsilon_1^{(n)}, \ldots, \varepsilon_N^{(n)})$ given $N = m$ is the distribution of m independent copies of ε,*
- *We have*

$$R_N := \max_{1\le i\le N} \frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i^{(n)}|}{1 + |\varepsilon_i^{(n)}|} \xrightarrow{\mathbb{P}} 0.$$

(i) Then we have both

$$\sup_{0<s\le 1} \left| \frac{\widehat{\varepsilon}^{(n)}_{N-\lfloor k(N)s\rfloor,N}}{\varepsilon^{(n)}_{N-\lfloor k(N)s\rfloor,N}} - 1 \right| = \mathrm{O}_{\mathbb{P}}(R_N) \ \ \textit{and} \ \ \sup_{0<s\le 1} \left| \log\left(\frac{\widehat{\varepsilon}^{(n)}_{N-\lfloor k(N)s\rfloor,N}}{\varepsilon^{(n)}_{N-\lfloor k(N)s\rfloor,N}} \right) \right| = \mathrm{O}_{\mathbb{P}}(R_N).$$

Here by convention $\widehat{\varepsilon}^{(n)}_{N-\lfloor k(N)s\rfloor,N}$ and $\varepsilon^{(n)}_{N-\lfloor k(N)s\rfloor,N}$ are equal to 1 on the event $\{N = 0\}$.

(ii) If moreover ε satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$ and $\tau_n \uparrow 1$ is such that $n(1-\tau_n) \to \infty$, $\sqrt{n(1-\tau_n)}A((1-\tau_n)^{-1}) \to \lambda \in \mathbb{R}$ and $\sqrt{N(1-\tau_N)}R_N \xrightarrow{\mathbb{P}} 0$, then the Hill estimator

$$\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor} = \frac{1}{\lfloor N(1-\tau_N)\rfloor} \sum_{i=1}^{\lfloor N(1-\tau_N)\rfloor} \log \frac{\widehat{\varepsilon}^{(n)}_{N-i+1,N}}{\widehat{\varepsilon}^{(n)}_{N-\lfloor N(1-\tau_N)\rfloor,N}}$$

is such that $\sqrt{N(1-\tau_N)}(\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor} - \gamma) \xrightarrow{d} \mathcal{N}(\lambda/(1-\rho), \gamma^2)$.

PROOF. We follow the proof of Lemma A.3. On the event $\{N > 0\} \cap \{R_N \le 1/4\}$, having arbitrarily high probability, we may write

$$\forall i \in \{1, \ldots, N\},\ \varepsilon^{(n)}_{i,N} - R_N(1 + |\varepsilon^{(n)}_{i,N}|) \le \widehat{\varepsilon}^{(n)}_{i,N} \le \varepsilon^{(n)}_{i,N} + R_N(1 + |\varepsilon^{(n)}_{i,N}|).$$

Given $N = m$, the random variable $\varepsilon^{(n)}_{N-k(N),N}$ has the same distribution as $\varepsilon_{m-k(m),m}$, the $(m - k(m))$th order statistic of a sample of m independent copies of ε. Since

$\varepsilon_{m-k(m),m} \xrightarrow{\mathbb{P}} +\infty$ as $m \to \infty$, we obtain likewise $\varepsilon^{(n)}_{N-k(N),N} \xrightarrow{\mathbb{P}} +\infty$ by the deconditioning Lemma C.4(iii). On the event $A_n := \{N > 0\} \cap \{R_N \le 1/4\} \cap \{\varepsilon_{N-k(N),N} \ge 1\}$, whose probability tends to 1, we have

$$\forall i \ge N-k(N),\ (1-R_N)\varepsilon^{(n)}_{i,N} - R_N \le \widehat{\varepsilon}^{(n)}_{i,N} \le (1+R_N)\varepsilon^{(n)}_{i,N} + R_N.$$

Therefore, on A_N,

$$\forall s \in (0,1],\ -2R_N \le \frac{\widehat{\varepsilon}^{(n)}_{N-\lfloor k(N)s\rfloor,N}}{\varepsilon^{(n)}_{N-\lfloor k(N)s\rfloor,N}} - 1 \le 2R_N.$$

Mimic then the final stages of the proof of Lemma A.3 to conclude the proof of (i).

(ii) Define

$$\widetilde{\gamma}_{\lfloor N(1-\tau_N)\rfloor} = \frac{1}{\lfloor N(1-\tau_N)\rfloor} \sum_{i=1}^{\lfloor N(1-\tau_N)\rfloor} \log \frac{\varepsilon^{(n)}_{N-i+1,N}}{\varepsilon^{(n)}_{N-\lfloor N(1-\tau_N)\rfloor,N}}.$$

By (i) and the assumption $\sqrt{N(1-\tau_N)}R_N \xrightarrow{\mathbb{P}} 0$,

$$\sqrt{N(1-\tau_N)}(\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor} - \gamma) = \sqrt{N(1-\tau_N)}(\widetilde{\gamma}_{\lfloor N(1-\tau_N)\rfloor} - \gamma) + \mathrm{o}_{\mathbb{P}}(1).$$

Combine Lemma C.4(i) and Theorem 3.2.5 in [6] to conclude the proof of (ii). □

Lemma C.7 contains the crucial arguments behind our construction in Section 3.3.

LEMMA C.7. *Work in model (M_3). Assume that ε satisfies condition $\mathcal{C}_1(\gamma)$ and that K_0 is a measurable subset of the support of $\boldsymbol{X}$ such that $\mathbb{P}(\boldsymbol{X} \in K_0) > 0$.*

(i) *There exists $\tau_c \in (0,1)$ such that $q_\tau(Y|\boldsymbol{x}) = g(\boldsymbol{x}) + \sigma(\boldsymbol{x})q_\tau(\varepsilon)$ for any $\tau \in [\tau_c, 1]$ and any $\boldsymbol{x}$ in the support of $\boldsymbol{X}$.*

(ii) *If $\mathbb{E}|\varepsilon_-| < \infty$ and $0 < \gamma < 1$, one has*

$$\xi_\tau(Y|\boldsymbol{x}) = g(\boldsymbol{x}) + \sigma(\boldsymbol{x})\xi_\tau(\max(\varepsilon, (y_0 - g(\boldsymbol{x}))/\sigma(\boldsymbol{x}))).$$

In particular the expectile $\xi_\tau(Y|\boldsymbol{X} = \boldsymbol{x})$ is asymptotically equivalent to $\xi_\tau(g(\boldsymbol{X}) + \sigma(\boldsymbol{X})\varepsilon|\boldsymbol{X} = \boldsymbol{x})$ as $\tau \uparrow 1$.

(iii) *The probability $\mathbb{P}(\varepsilon > (y_0 - g(\boldsymbol{X}))/\sigma(\boldsymbol{X}), \boldsymbol{X} \in K_0)$ is not zero. Let e have the same distribution as $(Y - g(\boldsymbol{X}))/\sigma(\boldsymbol{X})$ given that $g(\boldsymbol{X}) + \sigma(\boldsymbol{X})\varepsilon > y_0$ and $\boldsymbol{X} \in K_0$. Then for t so large that $(y_0 - g(\boldsymbol{X}))/\sigma(\boldsymbol{X}) \le t$ with probability 1,*

$$\mathbb{P}(e > t) = \frac{\mathbb{P}(\varepsilon > t)}{\mathbb{P}(\varepsilon > (y_0 - g(\boldsymbol{X}))/\sigma(\boldsymbol{X}) \,|\, \boldsymbol{X} \in K_0)}.$$

In particular, e satisfies condition $\mathcal{C}_1(\gamma)$.

(iv) *Let $p = \mathbb{P}(\varepsilon > (y_0 - g(\boldsymbol{X}))/\sigma(\boldsymbol{X}) \,|\, \boldsymbol{X} \in K_0)$. Then $q_\tau(\varepsilon)/q_\tau(e) \to p^\gamma$ as $\tau \uparrow 1$. If moreover $\mathbb{E}|\varepsilon_-| < \infty$ and $0 < \gamma < 1$, then $\xi_\tau(\varepsilon)/\xi_\tau(e) \to p^\gamma$ as $\tau \uparrow 1$.*

(v) *If, in addition to $\mathbb{E}|\varepsilon_-| < \infty$ and $0 < \gamma < 1$, the random variable ε satisfies condition $\mathcal{C}_2(\gamma, \rho, A)$, then e satisfies condition $\mathcal{C}_2(\gamma, \rho, p^{-\rho}A)$ and, as $\tau \uparrow 1$,*

$$p^\gamma \frac{\xi_\tau(e)}{\xi_\tau(\varepsilon)} = 1 + p^\gamma \frac{\gamma(\gamma^{-1}-1)^\gamma}{q_\tau(\varepsilon)} \left(\mathbb{E}\left[\varepsilon \,\middle|\, \varepsilon > \frac{y_0 - g(\boldsymbol{X})}{\sigma(\boldsymbol{X})}, \boldsymbol{X} \in K_0 \right] + \mathrm{o}(1) \right)$$
$$+ \frac{p^{-\rho} - 1}{\rho} \left(1 + \rho \left[\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho} + \frac{(\gamma^{-1}-1)^{-\rho} - 1}{\rho} \right] + \mathrm{o}(1) \right) A((1-\tau)^{-1}).$$

(vi) *Under the assumptions of (v), as* $\tau \uparrow 1$,

$$\frac{\xi_\tau(Y|\boldsymbol{x})}{g(\boldsymbol{x}) + \sigma(\boldsymbol{x})\xi_\tau(\varepsilon)}$$
$$= 1 + \frac{\gamma(\gamma^{-1}-1)^\gamma}{q_\tau(\varepsilon)}\left(\mathbb{E}\left[\max\left(\varepsilon, \frac{y_0 - g(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\right)\right] + \mathrm{o}(1)\right) + \mathrm{o}(|A((1-\tau)^{-1})|).$$

PROOF. The key point is to remark that $Y = \max(g(\boldsymbol{X}) + \sigma(\boldsymbol{X})\varepsilon, y_0)$. By independence between $\boldsymbol{X}$ and ε, the conditional distribution of Y given $\boldsymbol{X} = \boldsymbol{x}$ is then the distribution of $\max(g(\boldsymbol{x}) + \sigma(\boldsymbol{x})\varepsilon, y_0) = g(\boldsymbol{x}) + \sigma(\boldsymbol{x})\max(\varepsilon, (y_0 - g(\boldsymbol{x}))/\sigma(\boldsymbol{x}))$.

(i) The τth conditional quantile of Y given $\boldsymbol{X} = \boldsymbol{x}$ is

$$q_\tau(Y|\boldsymbol{x}) = g(\boldsymbol{x}) + \sigma(\boldsymbol{x})\max(q_\tau(\varepsilon), (y_0 - g(\boldsymbol{x}))/\sigma(\boldsymbol{x})).$$

Since g and $1/\sigma$ are bounded on the support of $\boldsymbol{X}$ and $q_\tau(\varepsilon) \to \infty$ as $\tau \uparrow 1$, one has $q_\tau(\varepsilon) > (y_0 - g(\boldsymbol{x}))/\sigma(\boldsymbol{x})$ for τ large enough, irrespective of $\boldsymbol{x}$. Conclude that there is $\tau_c \in (0,1)$ with $q_\tau(Y|\boldsymbol{x}) = g(\boldsymbol{x}) + \sigma(\boldsymbol{x})q_\tau(\varepsilon)$ for any $\tau \in [\tau_c, 1]$ and any $\boldsymbol{x}$ in the support of $\boldsymbol{X}$, as required.

(ii) By location equivariance and positive homogeneity of expectiles, the τth conditional expectile of Y given $\boldsymbol{X} = \boldsymbol{x}$ is

$$\xi_\tau(Y|\boldsymbol{x}) = g(\boldsymbol{x}) + \sigma(\boldsymbol{x})\xi_\tau(\max(\varepsilon, (y_0 - g(\boldsymbol{x}))/\sigma(\boldsymbol{x}))).$$

To conclude, it is sufficient to show that for any t_0, the extreme expectiles of ε and $\max(\varepsilon, t_0)$ are asymptotically equivalent. To do so we note that the definition of the τth unconditional expectile $\xi_\tau(\varepsilon)$ of ε as

$$\xi_\tau(\varepsilon) = \mathop{\arg\min}_{\theta \in \mathbb{R}} \mathbb{E}(\eta_\tau(\varepsilon - \theta) - \eta_\tau(\varepsilon))$$

can equivalently be obtained as the τth quantile associated to the distribution function E defined as

$$1 - E(y) = \frac{\mathbb{E}\left[(\varepsilon - y)\mathbb{1}_{\{\varepsilon > y\}}\right]}{2\,\mathbb{E}\left[(\varepsilon - y)\mathbb{1}_{\{\varepsilon > y\}}\right] + y - \mathbb{E}[\varepsilon]}.$$

See *e.g.* the final paragraph of p.373 in [1]. Similarly the τth expectile $\xi_\tau(\max(\varepsilon, t_0))$ of $\max(\varepsilon, t_0)$ is obtained as the τth quantile associated to the distribution function E_0 defined as

$$1 - E_0(y) = \frac{\mathbb{E}\left[(\max(\varepsilon, t_0) - y)\mathbb{1}_{\{\max(\varepsilon, t_0) > y\}}\right]}{2\,\mathbb{E}\left[(\max(\varepsilon, t_0) - y)\mathbb{1}_{\{\max(\varepsilon, t_0) > y\}}\right] + y - \mathbb{E}[\max(\varepsilon, t_0)]}.$$

It is straightforward to check that for $y > t_0$

$$1 - E_0(y) = \frac{\mathbb{E}\left[(\varepsilon - y)\mathbb{1}_{\{\varepsilon > y\}}\right]}{2\,\mathbb{E}\left[(\varepsilon - y)\mathbb{1}_{\{\varepsilon > y\}}\right] + y - \mathbb{E}[\max(\varepsilon, t_0)]}.$$

Lemma 3(i) in [18] (with f therein chosen as the identity function and $a = 1$) entails that $y \mapsto 1/(1 - E(y))$ and $y \mapsto 1/(1 - E_0(y))$ are asymptotically equivalent as $y \to \infty$ and regularly varying with positive index. Let U and U_0 denote the pertaining tail quantile functions, *i.e.* the left-continuous inverses of $1/(1 - E)$ and $1/(1 - E_0)$; these are also regularly varying, and we will conclude by proving that U and U_0 are asymptotically equivalent. A combination of Equations (1.2.26) and (1.2.28) in [6] and the regular variation property of U entails

$$\lim_{t\to\infty} \frac{t^{-1}}{(1-E)(U(t))} = \lim_{t\to\infty} \frac{t^{-1}}{(1-E_0)(U(t))} = \lim_{t\to\infty} \frac{t^{-1}}{(1-E_0)(U_0(t))} = 1,$$

$$\lim_{t\to\infty} t^{-1}U(1/(1-E)(t)) = \lim_{t\to\infty} t^{-1}U(1/(1-E_0)(t)) = \lim_{t\to\infty} t^{-1}U_0(1/(1-E_0)(t)) = 1.$$

Apply Proposition B.1.9.10 in [6] to obtain that U and U_0 are indeed asymptotically equivalent, thus completing the proof of (ii).

(iii) First of all, if $\mathbb{P}_{\boldsymbol{X}}$ denotes the distribution of $\boldsymbol{X}$,

$$\mathbb{P}(\varepsilon > (y_0 - g(\boldsymbol{X}))/\sigma(\boldsymbol{X}), \boldsymbol{X} \in K_0) = \int_{K_0} \mathbb{P}(\varepsilon > (y_0 - g(\boldsymbol{x}))/\sigma(\boldsymbol{x}))\, \mathbb{P}_{\boldsymbol{X}}(d\boldsymbol{x}) > 0$$

because $\mathbb{P}(\varepsilon > (y_0 - g(\boldsymbol{x}))/\sigma(\boldsymbol{x})) > 0$ for any $\boldsymbol{x}$ (since ε is heavy-tailed) and $\mathbb{P}(\boldsymbol{X} \in K_0) > 0$. Write then

$$\begin{aligned}\mathbb{P}(e > t) &= \mathbb{P}(\varepsilon > t \,|\, g(\boldsymbol{X}) + \sigma(\boldsymbol{X})\varepsilon > y_0, \boldsymbol{X} \in K_0) \\ &= \frac{\mathbb{P}(\varepsilon > t, \varepsilon > (y_0 - g(\boldsymbol{X}))/\sigma(\boldsymbol{X}), \boldsymbol{X} \in K_0)}{\mathbb{P}(\varepsilon > (y_0 - g(\boldsymbol{X}))/\sigma(\boldsymbol{X}), \boldsymbol{X} \in K_0)}.\end{aligned}$$

It is indeed possible to take t so large that $(y_0 - g(\boldsymbol{X}))/\sigma(\boldsymbol{X}) \le t$ with probability 1 since g and $1/\sigma$ are bounded on the support of $\boldsymbol{X}$. For such t,

$$\mathbb{P}(e > t) = \frac{\mathbb{P}(\varepsilon > t, \boldsymbol{X} \in K_0)}{\mathbb{P}(\varepsilon > (y_0 - g(\boldsymbol{X}))/\sigma(\boldsymbol{X}), \boldsymbol{X} \in K_0)} = \frac{\mathbb{P}(\varepsilon > t)}{\mathbb{P}(\varepsilon > (y_0 - g(\boldsymbol{X}))/\sigma(\boldsymbol{X}) \,|\, \boldsymbol{X} \in K_0)}$$

by independence between $\boldsymbol{X}$ and ε, which is the required result.

(iv) That $q_\tau(\varepsilon)/q_\tau(e) \to p^\gamma$ as $\tau \uparrow 1$ directly follows from the identity $\mathbb{P}(e > t) = p^{-1}\mathbb{P}(\varepsilon > t)$ for t large enough, and therefore $q_\tau(e) = q_{1-p(1-\tau)}(\varepsilon)$ for τ close enough to 1, combined with the regular variation property of $t \mapsto U(t) = q_{1-t^{-1}}(\varepsilon)$. The convergence $\xi_\tau(\varepsilon)/\xi_\tau(e) \to p^\gamma$ as $\tau \uparrow 1$ follows from the asymptotic proportionality relationship between extreme quantiles and expectiles applied to both e and ε (which have the same extreme value index).

(v) Recall from the proof of (iv) that for τ close enough to 1, $q_\tau(e) = q_{1-p(1-\tau)}(\varepsilon)$. Set $V(t) = q_{1-t^{-1}}(e)$ and pick $x > 0$. For t large enough, we find

$$\begin{aligned}\frac{V(tx)}{V(t)} = \frac{U(p^{-1}tx)}{U(p^{-1}t)} &= x^\gamma + A(p^{-1}t)\left(x^\gamma \frac{x^\rho - 1}{\rho} + \mathrm{o}(1)\right) \\ &= x^\gamma + p^{-\rho}A(t)\left(x^\gamma \frac{x^\rho - 1}{\rho} + \mathrm{o}(1)\right)\end{aligned}$$

by assumption $\mathcal{C}_2(\gamma, \rho, A)$ on ε and regular variation of $|A|$ with index ρ (see Section 2.3 in [6]). This exactly means that e satisfies condition $\mathcal{C}_2(\gamma, \rho, p^{-\rho}A)$. Write then

$$p^\gamma \frac{\xi_\tau(e)}{\xi_\tau(\varepsilon)} = p^\gamma \frac{q_\tau(e)}{q_\tau(\varepsilon)} \times (\gamma^{-1} - 1)^\gamma \frac{\xi_\tau(e)}{q_\tau(e)} \times (\gamma^{-1} - 1)^{-\gamma} \frac{q_\tau(\varepsilon)}{\xi_\tau(\varepsilon)}. \tag{48}$$

Use again the identity $q_\tau(e) = q_{1-p(1-\tau)}(\varepsilon)$ for τ close enough to 1 to get

$$p^\gamma \frac{q_\tau(e)}{q_\tau(\varepsilon)} = p^\gamma \frac{U(p^{-1}(1-\tau)^{-1})}{U((1-\tau)^{-1})} = 1 + \left(\frac{p^{-\rho} - 1}{\rho} + \mathrm{o}(1)\right) A((1-\tau)^{-1}). \tag{49}$$

Proposition 1(i) in [5] applied to the random variable ε (having expectation 0) entails

$$(\gamma^{-1}-1)^{-\gamma}\frac{q_\tau(\varepsilon)}{\xi_\tau(\varepsilon)}$$

$$(50)\qquad = 1-\left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}+\frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho}+\mathrm{o}(1)\right)A((1-\tau)^{-1})+\mathrm{o}\left(\frac{1}{q_\tau(\varepsilon)}\right).$$

This same result applied to the random variable e, which satisfies condition $\mathcal{C}_2(\gamma,\rho,p^{-\rho}A)$, gives

$$\begin{aligned}(\gamma^{-1}-1)^{\gamma}\frac{\xi_\tau(e)}{q_\tau(e)} &= 1+\frac{\gamma(\gamma^{-1}-1)^{\gamma}}{q_\tau(e)}(\mathbb{E}(e)+\mathrm{o}(1)) \\ &\quad +\left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}+\frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho}+\mathrm{o}(1)\right)p^{-\rho}A((1-\tau)^{-1}) \\ &= 1+p^{\gamma}\frac{\gamma(\gamma^{-1}-1)^{\gamma}}{q_\tau(\varepsilon)}\left(\mathbb{E}\left[\varepsilon\,\middle|\,\varepsilon>\frac{y_0-g(\boldsymbol{X})}{\sigma(\boldsymbol{X})},\,\boldsymbol{X}\in K_0\right]+\mathrm{o}(1)\right) \\ (51)\qquad &\quad +p^{-\rho}\left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}+\frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho}+\mathrm{o}(1)\right)A((1-\tau)^{-1}).\end{aligned}$$

Combine (48), (49), (50) and (51) to get (v).

(vi) From (ii),

$$\begin{aligned}\frac{\xi_\tau(Y|\boldsymbol{x})}{g(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_\tau(\varepsilon)}-1 &= \frac{\sigma(\boldsymbol{x})[\xi_\tau(\max(\varepsilon,(y_0-g(\boldsymbol{x}))/\sigma(\boldsymbol{x})))-\xi_\tau(\varepsilon)]}{g(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_\tau(\varepsilon)} \\ &= \left(\frac{\xi_\tau(\max(\varepsilon,(y_0-g(\boldsymbol{x}))/\sigma(\boldsymbol{x})))}{\xi_\tau(\varepsilon)}-1\right)(1+\mathrm{o}(1))\end{aligned}$$

because $\xi_\tau(\varepsilon)\to\infty$ as $\tau\uparrow 1$. To complete the proof we show that for any t_0,

$$\frac{\xi_\tau(\max(\varepsilon,t_0))}{\xi_\tau(\varepsilon)}=1+\frac{\gamma(\gamma^{-1}-1)^{\gamma}}{q_\tau(\varepsilon)}\left(\mathbb{E}[\max(\varepsilon,t_0)]+\mathrm{o}(1)\right)+\mathrm{o}(|A((1-\tau)^{-1})|)$$

as $\tau\uparrow 1$. This is done by, first, writing

$$\frac{\xi_\tau(\max(\varepsilon,t_0))}{\xi_\tau(\varepsilon)}=\frac{\xi_\tau(\max(\varepsilon,t_0))}{q_\tau(\max(\varepsilon,t_0))}\times\frac{q_\tau(\max(\varepsilon,t_0))}{q_\tau(\varepsilon)}\times\frac{q_\tau(\varepsilon)}{\xi_\tau(\varepsilon)}=\frac{\xi_\tau(\max(\varepsilon,t_0))}{q_\tau(\max(\varepsilon,t_0))}\times\frac{q_\tau(\varepsilon)}{\xi_\tau(\varepsilon)}$$

for τ close enough to 1. Then, using the fact that $\max(\varepsilon,t_0)$ and ε have the same quantile function for τ large enough, we obtain, by Proposition 1(i) in [5],

$$\begin{aligned}(\gamma^{-1}-1)^{\gamma}\frac{\xi_\tau(\max(\varepsilon,t_0))}{q_\tau(\max(\varepsilon,t_0))} &= 1+\frac{\gamma(\gamma^{-1}-1)^{\gamma}}{q_\tau(\varepsilon)}(\mathbb{E}[\max(\varepsilon,t_0)]+\mathrm{o}(1)) \\ &\quad +\left(\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho}+\frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho}+\mathrm{o}(1)\right)A((1-\tau)^{-1}).\end{aligned}$$

Combining this with (50) completes the proof. □

Our final auxiliary result is a direct extension of Theorem 2.1 to the case when the residuals $\widehat{\varepsilon}_i^{(n)}$ approximate an array $\varepsilon_i^{(n)}$, with $1\le i\le s_n\to\infty$. This will be useful to deal with the case of ARMA and GARCH models.

LEMMA C.8. *Let (s_n) be a positive sequence of integers tending to infinity. Assume that, for any n, the $\varepsilon_i^{(n)}$, $1 \le i \le s_n$, are independent copies of a random variable ε such that there is $\delta > 0$ with $\mathbb{E}|\varepsilon_-|^{2+\delta} < \infty$ and ε satisfies condition $\mathcal{C}_1(\gamma)$ with $0 < \gamma < 1/2$. Let $\tau_n \uparrow 1$ be such that $s_n(1-\tau_n) \to \infty$. Suppose moreover that the array of random variables $\widehat{\varepsilon}_i^{(n)}$, $1 \le i \le s_n$, satisfies*

$$\sqrt{s_n(1-\tau_n)} \max_{1 \le i \le s_n} \frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i^{(n)}|}{1 + |\varepsilon_i^{(n)}|} \xrightarrow{\mathbb{P}} 0.$$

Define

$$\widehat{\xi}_{\tau_n}(\varepsilon) = \arg\min_{u \in \mathbb{R}} \sum_{i=1}^{s_n} \eta_{\tau_n}(\widehat{\varepsilon}_i^{(n)} - u).$$

Then we have $\sqrt{s_n(1-\tau_n)} \left(\dfrac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \dfrac{2\gamma^3}{1-2\gamma}\right)$.

APPENDIX D: WORKED-OUT EXAMPLES: PROOFS OF THE MAIN RESULTS

PROOF OF COROLLARY 3.1. (i) The key is to write

$$\begin{aligned}
&\sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(Y|\boldsymbol{x})}{\xi_{\tau_n}(Y|\boldsymbol{x})} - 1 \right) \\
&= \frac{(1+\boldsymbol{\theta}^\top \boldsymbol{x})\xi_{\tau_n}(\varepsilon)}{\alpha + \boldsymbol{\beta}^\top \boldsymbol{x} + (1+\boldsymbol{\theta}^\top \boldsymbol{x})\xi_{\tau_n}(\varepsilon)} \times \sqrt{n(1-\tau_n)} \left(\frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1 \right) \\
&+ \frac{\sqrt{1-\tau_n}}{\alpha + \boldsymbol{\beta}^\top \boldsymbol{x} + (1+\boldsymbol{\theta}^\top \boldsymbol{x})\xi_{\tau_n}(\varepsilon)} \times \sqrt{n} \left[\widehat{\alpha} - \alpha + \left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \right)^\top \boldsymbol{x} \right] \\
&+ \frac{\sqrt{1-\tau_n}\, \widehat{\xi}_{\tau_n}(\varepsilon)}{\alpha + \boldsymbol{\beta}^\top \boldsymbol{x} + (1+\boldsymbol{\theta}^\top \boldsymbol{x})\xi_{\tau_n}(\varepsilon)} \times \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})^\top \boldsymbol{x}.
\end{aligned}$$

Now

$$\widehat{\varepsilon}_i^{(n)} - \varepsilon_i = \frac{\alpha - \widehat{\alpha} + (\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})^\top \boldsymbol{X}_i}{1 + \widehat{\boldsymbol{\theta}}^\top \boldsymbol{X}_i} + \frac{(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^\top \boldsymbol{X}_i}{1 + \widehat{\boldsymbol{\theta}}^\top \boldsymbol{X}_i} \varepsilon_i.$$

Then clearly, by Lemma C.1 and since $\boldsymbol{X}$ has a compact support,

$$\sqrt{n} \max_{1 \le i \le n} \frac{|\widehat{\varepsilon}_i^{(n)} - \varepsilon_i|}{1 + |\varepsilon_i|} = \mathrm{O}_{\mathbb{P}}(1), \tag{52}$$

which proves the high-level condition (2). We conclude by combining Lemma C.1, Theorem 2.1 and the convergence $\xi_{\tau_n}(\varepsilon) \to \infty$.

(ii) Combine (i) with the second convergence in Theorem 2.3. □

PROOF OF THEOREM 3.1. (i) We first show

$$\sqrt{N(1-\tau_N)} \left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)} - 1 \right) \xrightarrow{d} \mathcal{N}\left(0, \frac{2\gamma^3}{1-2\gamma}\right). \tag{53}$$

Let $\varepsilon_{1,K_0},\ldots,\varepsilon_{N,K_0}$ be those noise variables whose corresponding covariates $\boldsymbol{X}_i\in K_0$, and note that given $N=m>0$, $(\varepsilon_{1,K_0},\ldots,\varepsilon_{N,K_0})\stackrel{d}{=}(\varepsilon_1,\ldots,\varepsilon_m)$. Besides, $N=N(K_0,n)$ is a binomial random variable with parameters n and $\mathbb{P}(\boldsymbol{X}\in K_0)$, so that $N/n\stackrel{\mathbb{P}}{\longrightarrow}\mathbb{P}(\boldsymbol{X}\in K_0)>0$. Since $\tau_n=1-n^{-a}$ with $a\in(1/5,1)$,

$$\sqrt{N(1-\tau_N)}=N^{(1-a)/2}=\mathrm{O}_{\mathbb{P}}(n^{(1-a)/2})=\mathrm{o}_{\mathbb{P}}(n^{2/5}/\sqrt{\log n})$$

so that

$$\sqrt{N(1-\tau_N)}\max_{1\le i\le N}\frac{|\widehat{\varepsilon}_{i,K_0}^{(n)}-\varepsilon_{i,K_0}|}{1+|\varepsilon_{i,K_0}|}=\mathrm{o}_{\mathbb{P}}\left(\frac{n^{2/5}}{\sqrt{\log n}}\max_{1\le i\le n}\frac{|\widehat{\varepsilon}_i^{(n)}-\varepsilon_i|}{1+|\varepsilon_i|}\mathbb{1}\{\boldsymbol{X}_i\in K_0\}\right)=\mathrm{o}_{\mathbb{P}}(1).$$

Apply then Lemma C.5 to get (53). Statement (i) then follows in a straightforward way from Proposition C.1 and the representation

$$\begin{aligned}\frac{\widehat{\xi}_{\tau_N}(Y|\boldsymbol{x})}{\xi_{\tau_N}(Y|\boldsymbol{x})}-1&=\frac{\widehat{g}_{h_n,t_n}(\widehat{\boldsymbol{\beta}}^\top\boldsymbol{x})-g(\boldsymbol{\beta}^\top\boldsymbol{x})}{g(\boldsymbol{\beta}^\top\boldsymbol{x})+\sigma(\boldsymbol{\beta}^\top\boldsymbol{x})\xi_{\tau_N}(\varepsilon)}+\frac{\widehat{\sigma}_{h_n,t_n}(\widehat{\boldsymbol{\beta}}^\top\boldsymbol{x})-\sigma(\boldsymbol{\beta}^\top\boldsymbol{x})}{g(\boldsymbol{\beta}^\top\boldsymbol{x})+\sigma(\boldsymbol{\beta}^\top\boldsymbol{x})\xi_{\tau_N}(\varepsilon)}\widehat{\xi}_{\tau_N}(\varepsilon)\\&\quad+\frac{\sigma(\boldsymbol{\beta}^\top\boldsymbol{x})}{\sigma(\boldsymbol{\beta}^\top\boldsymbol{x})+g(\boldsymbol{\beta}^\top\boldsymbol{x})/\xi_{\tau_N}(\varepsilon)}\left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)}-1\right).\end{aligned}$$

(ii) Set $\widehat{\xi}^{\star}_{\tau'_N}(\varepsilon)=\left(\dfrac{1-\tau'_N}{1-\tau_N}\right)^{-\overline{\gamma}}\widehat{\xi}_{\tau_N}(\varepsilon)$. Use the ideas of the proof of Theorem 2.3 to find that

$$\frac{\sqrt{N(1-\tau_N)}}{\log[(1-\tau_N)/(1-\tau'_N)]}\left(\frac{\widehat{\xi}^{\star}_{\tau'_N}(Y|\boldsymbol{x})}{\xi_{\tau'_N}(Y|\boldsymbol{x})}-1\right)\ \text{ and }\ \frac{\sqrt{N(1-\tau_N)}}{\log[(1-\tau_N)/(1-\tau'_N)]}\left(\frac{\widehat{\xi}^{\star}_{\tau'_N}(\varepsilon)}{\xi_{\tau'_N}(\varepsilon)}-1\right)$$

have the same asymptotic distribution. Our result is then shown by using the assumption $\sqrt{N(1-\tau_N)}(\overline{\gamma}-\gamma)\stackrel{d}{\longrightarrow}\Gamma$, as well as convergence (53) and by adapting directly the proof of Theorem 5 of [5] to obtain

$$\frac{\sqrt{N(1-\tau_N)}}{\log[(1-\tau_N)/(1-\tau'_N)]}\left(\frac{\widehat{\xi}^{\star}_{\tau'_N}(\varepsilon)}{\xi_{\tau'_N}(\varepsilon)}-1\right)\stackrel{d}{\longrightarrow}\Gamma.$$

We omit the details. □

PROOF OF THEOREM 3.2. First of all, define

$$\widehat{\xi}_{\tau_N}(\varepsilon):=\left(\frac{N}{N_0}\right)^{\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor}}\widehat{\xi}_{\tau_N}(e)$$

so that $\widehat{\xi}_{\tau_N}(Y|\boldsymbol{x})=\widehat{g}(\boldsymbol{x})+\widehat{\sigma}(\boldsymbol{x})\widehat{\xi}_{\tau_N}(\varepsilon)$. Then

$$\begin{aligned}&\frac{\widehat{\xi}_{\tau_N}(Y|\boldsymbol{x})}{\xi_{\tau_N}(Y|\boldsymbol{x})}-1\\&=\left(\frac{\widehat{g}(\boldsymbol{x})+\widehat{\sigma}(\boldsymbol{x})\widehat{\xi}_{\tau_N}(\varepsilon)}{g(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau_N}(\varepsilon)}-1\right)\frac{g(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(Y|\boldsymbol{x})}\\&+\left(\frac{g(\boldsymbol{x})+\sigma(\boldsymbol{x})\xi_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(Y|\boldsymbol{x})}-1\right)\\&=\left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)}-1\right)(1+\mathrm{o}_{\mathbb{P}}(1))+\mathrm{o}_{\mathbb{P}}(|\widehat{g}(\boldsymbol{x})-g(\boldsymbol{x})|)+\mathrm{O}_{\mathbb{P}}\left(\left|\frac{\widehat{\sigma}(\boldsymbol{x})}{\sigma(\boldsymbol{x})}-1\right|\right)\end{aligned}$$

$$- \frac{\gamma(\gamma^{-1}-1)^{\gamma}}{q_{\tau_N}(\varepsilon)} \left(\mathbb{E}\left[\max\left(\varepsilon, \frac{y_0 - g(\boldsymbol{x})}{\sigma(\boldsymbol{x})} \right) \right] + \mathrm{o}_{\mathbb{P}}(1) \right) + \mathrm{o}_{\mathbb{P}}(|A((1-\tau_N)^{-1})|)$$

by Lemma C.7(vi), the consistency assumption on $\widehat{g}$ and $\widehat{\sigma}$, and $N = N(n) \xrightarrow{\mathbb{P}} \infty$. Now $1/v_n = \mathrm{o}_{\mathbb{P}}(1/\sqrt{N(1-\tau_N)})$, because $n^{1-a}/v_n^2 \to 0$ and $N(1-\tau_N) = N^{1-a} \le n^{1-a}$. The v_n-consistency of $\widehat{g}$ and $\widehat{\sigma}$ then entails

$$\sqrt{N(1-\tau_N)} \left(\frac{\widehat{\xi}_{\tau_N}(Y|\boldsymbol{x})}{\xi_{\tau_N}(Y|\boldsymbol{x})} - 1 \right) = \sqrt{N(1-\tau_N)} \left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)} - 1 \right) (1 + \mathrm{o}_{\mathbb{P}}(1))$$
$$- \gamma(\gamma^{-1}-1)^{\gamma} \mathbb{E}\left[\max\left(\varepsilon, \frac{y_0 - g(\boldsymbol{x})}{\sigma(\boldsymbol{x})} \right) \right] \mu + \mathrm{o}_{\mathbb{P}}(1). \tag{54}$$

It is therefore sufficient to consider the convergence of $\widehat{\xi}_{\tau_N}(\varepsilon)$. Write

$$\log\left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)} \right) = (\widehat{\gamma}_{\lfloor N(1-\tau_N) \rfloor} - \gamma) \log\left(\frac{N}{N_0} \right) + \gamma \left[\log\left(\frac{N}{N_0} \right) - \log p \right]$$
$$+ \log\left(\frac{\widehat{\xi}_{\tau_N}(e)}{\xi_{\tau_N}(e)} \right) + \log\left(p^{\gamma} \frac{\xi_{\tau_N}(e)}{\xi_{\tau_N}(\varepsilon)} \right).$$

The quantity N/N_0 is a $\sqrt{n}-$consistent estimator of $p > 0$, thus making the second term a $\mathrm{O}_{\mathbb{P}}(1/\sqrt{n}) = \mathrm{o}_{\mathbb{P}}(1/\sqrt{N(1-\tau_N)})$, and the fourth term is controlled with Lemma C.7(v) and a Taylor expansion. Therefore

$$\log\left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)} \right) = [\log p + \mathrm{o}_{\mathbb{P}}(1)](\widehat{\gamma}_{\lfloor N(1-\tau_N) \rfloor} - \gamma) + \log\left(\frac{\widehat{\xi}_{\tau_N}(e)}{\xi_{\tau_N}(e)} \right)$$
$$+ p^{\gamma}\gamma(\gamma^{-1}-1)^{\gamma} \left(\mathbb{E}\left[\varepsilon \,\middle|\, \varepsilon > \frac{y_0 - g(\boldsymbol{X})}{\sigma(\boldsymbol{X})}, \boldsymbol{X} \in K_0 \right] \right) \frac{\mu}{\sqrt{N(1-\tau_N)}}$$
$$+ \frac{p^{-\rho}-1}{\rho} \left(1 + \rho \left[\frac{(\gamma^{-1}-1)^{-\rho}}{1-\gamma-\rho} + \frac{(\gamma^{-1}-1)^{-\rho}-1}{\rho} \right] \right) \frac{\lambda}{\sqrt{N(1-\tau_N)}}$$
$$+ \mathrm{o}_{\mathbb{P}}\left(\frac{1}{\sqrt{N(1-\tau_N)}} \right). \tag{55}$$

It remains to analyse the joint convergence of $\widehat{\gamma}_{\lfloor N(1-\tau_N) \rfloor}$ and $\widehat{\xi}_{\tau_N}(e)$. First, clearly

$$\max_{1 \le i \le N} \frac{|\widehat{e}_i^{(n)} - e_i|}{1 + |e_i|} = \mathrm{O}_{\mathbb{P}}(1/v_n) = \mathrm{o}_{\mathbb{P}}(1/\sqrt{N(1-\tau_N)}),$$

which is (2) adapted to the random number N of noncensored observations (see Lemma C.6). Here the v_n-uniform consistency of $\widehat{g}$ and $\widehat{\sigma}$ on K_0 and boundedness of $1/\sigma$ on the support of $\boldsymbol{X}$ were used, along with again $n^{1-a}/v_n^2 \to 0$, and the identity $N(1-\tau_N) = N^{1-a} \le n^{1-a}$. Set then

$$\widehat{\gamma}_{\lfloor N(1-\tau_N) \rfloor} = \frac{1}{\lfloor N(1-\tau_N) \rfloor} \sum_{i=1}^{\lfloor N(1-\tau_N) \rfloor} \log \frac{\widehat{e}^{(n)}_{N-i+1,N}}{\widehat{e}^{(n)}_{N-\lfloor N(1-\tau_N) \rfloor, N}}$$

$$\text{and } \widetilde{\gamma}_{\lfloor N(1-\tau_N) \rfloor} = \frac{1}{\lfloor N(1-\tau_N) \rfloor} \sum_{i=1}^{\lfloor N(1-\tau_N) \rfloor} \log \frac{e_{N-i+1,N}}{e_{N-\lfloor N(1-\tau_N) \rfloor, N}}.$$

By Lemma C.6(i),

$$\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor} = \widetilde{\gamma}_{\lfloor N(1-\tau_N)\rfloor} + \mathrm{o}_{\mathbb{P}}(1/\sqrt{N(1-\tau_N)})$$

and therefore

$$\sqrt{N(1-\tau_N)}(\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor} - \gamma) = \sqrt{N(1-\tau_N)}(\widetilde{\gamma}_{\lfloor N(1-\tau_N)\rfloor} - \gamma) + \mathrm{o}_{\mathbb{P}}(1). \tag{56}$$

Let further

$$\widehat{\xi}_{\tau_N}(e) = \operatorname*{arg\,min}_{u\in\mathbb{R}} \sum_{i=1}^{N} \eta_{\tau_N}(\widehat{e}_i^{(n)} - u) \text{ and } \widetilde{\xi}_{\tau_N}(e) = \operatorname*{arg\,min}_{u\in\mathbb{R}} \sum_{i=1}^{N} \eta_{\tau_N}(e_i - u)$$

along with

$$\psi_N(u) = \frac{1}{2\xi^2_{\tau_N}(e)} \sum_{i=1}^{N} \left[\eta_{\tau_N}\left(e_i - \xi_{\tau_N}(e) - \frac{u\xi_{\tau_N}(e)}{\sqrt{N(1-\tau_N)}}\right) - \eta_{\tau_N}(e_i - \xi_{\tau_N}(e))\right]$$

$$\text{and } \chi_N(u) = \frac{1}{2\xi^2_{\tau_N}(e)} \sum_{i=1}^{N} \left[\eta_{\tau_N}\left(\widehat{e}_i^{(n)} - \xi_{\tau_N}(e) - \frac{u\xi_{\tau_N}(e)}{\sqrt{N(1-\tau_N)}}\right) - \eta_{\tau_N}(\widehat{e}_i^{(n)} - \xi_{\tau_N}(e))\right].$$

Lemma C.5 entails $\chi_N(u) = \psi_N(u) + \mathrm{o}_{\mathbb{P}}(1)$. Recall the notation $\varphi_\tau(y) = |\tau - \mathbb{1}\{y \le 0\}|y$ and write, as in the proof of Theorem 2 in [4], $\psi_N(u) = -u\mathcal{T}_{1,N} + \mathcal{T}_{2,N}(u)$ with

$$\mathcal{T}_{1,N} = \frac{1}{\sqrt{N(1-\tau_N)}} \sum_{i=1}^{N} \frac{1}{\xi_{\tau_N}(e)} \varphi_{\tau_N}(e_i - \xi_{\tau_N}(e))$$

and

$$\mathcal{T}_{2,N}(u) = -\frac{1}{\xi^2_{\tau_N}(e)} \sum_{i=1}^{N} \int_0^{u\xi_{\tau_N}(e)/\sqrt{N(1-\tau_N)}} (\varphi_{\tau_N}(e_i - \xi_{\tau_N}(e) - z) - \varphi_{\tau_N}(e_i - \xi_{\tau_N}(e)))\, dz.$$

The distribution of the e_i, $1 \le i \le N$, given $N = m$, is the distribution of m independent copies of e. Using the arguments of the proof of Theorem 2 in [4] and Lemma C.4(i) and (ii), we obtain $\mathcal{T}_{1,N} = \mathrm{O}_{\mathbb{P}}(1)$ and $\mathcal{T}_{2,N}(u) \xrightarrow{\mathbb{P}} u^2/2\gamma$. It follows that

$$\chi_N(u) = \psi_N(u) + \mathrm{o}_{\mathbb{P}}(1) = \frac{u^2}{2\gamma} - u\mathcal{T}_{1,N} + \mathrm{o}_{\mathbb{P}}(1).$$

Conclude, by the basic corollary on p.2 in [13], that the minimisers of χ_N and ψ_N are both only a $\mathrm{o}_{\mathbb{P}}(1)$ away from the minimiser of the right-hand side, and thus only a $\mathrm{o}_{\mathbb{P}}(1)$ away from each other. This can be rephrased as

$$\sqrt{N(1-\tau_N)}\left(\frac{\widehat{\xi}_{\tau_N}(e)}{\xi_{\tau_N}(e)} - 1\right) = \sqrt{N(1-\tau_N)}\left(\frac{\widetilde{\xi}_{\tau_N}(e)}{\xi_{\tau_N}(e)} - 1\right) + \mathrm{o}_{\mathbb{P}}(1). \tag{57}$$

Finally, the distribution of the pair $(\widetilde{\gamma}_{\lfloor N(1-\tau_N)\rfloor}, \widetilde{\xi}_{\tau_N}(e))$ given $N = m$ is equal to the distribution of their counterparts $(\breve{\gamma}_{\lfloor m(1-\tau_m)\rfloor}, \breve{\xi}_{\tau_m}(e))$ based on m independent copies of e. Combine then Theorem 3 in [5], which provides the bivariate asymptotic distribution of $(\breve{\gamma}_{\lfloor m(1-\tau_m)\rfloor}, \breve{\xi}_{\tau_m}(e))$, with Lemma C.4(i) to get

$$\sqrt{N(1-\tau_N)}\left(\widetilde{\gamma}_{\lfloor N(1-\tau_N)\rfloor} - \gamma, \frac{\widetilde{\xi}_{\tau_N}(e)}{\xi_{\tau_N}(e)} - 1\right) \xrightarrow{d} \mathcal{N}(\mathcal{B}(\rho, p), \mathcal{V}(\gamma)) \tag{58}$$

with $\mathcal{B}(\rho,p)=(p^{-\rho}\lambda/(1-\rho),0)$ (recall that e satisfies condition $\mathcal{C}_2(\gamma,\rho,p^{-\rho}A)$) and

$$\mathcal{V}(\gamma)=\begin{pmatrix}\gamma^2 & \dfrac{\gamma^3(\gamma^{-1}-1)^\gamma}{(1-\gamma)^2}\\ \dfrac{\gamma^3(\gamma^{-1}-1)^\gamma}{(1-\gamma)^2} & \dfrac{2\gamma^3}{1-2\gamma}\end{pmatrix}.$$

Combining (54), (55), (56), (57), (58) with the delta method completes the proof of (i).

(ii) Define

$$\widehat{\xi}^{\star}_{\tau_N'}(\varepsilon):=\left(\frac{1-\tau_N'}{1-\tau_N}\right)^{-\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor}}\left(\frac{N}{N_0}\right)^{\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor}}\widehat{\xi}_{\tau_N}(e)$$

so that $\widehat{\xi}^{\star}_{\tau_N'}(Y|\boldsymbol{x})=\widehat{g}(\boldsymbol{x})+\widehat{\sigma}(\boldsymbol{x})\widehat{\xi}^{\star}_{\tau_N'}(\varepsilon)$. Then

$$\begin{aligned}\frac{\widehat{\xi}^{\star}_{\tau_N'}(Y|\boldsymbol{x})}{\xi_{\tau_N'}(Y|\boldsymbol{x})}-1&=\left(\frac{\widehat{\xi}^{\star}_{\tau_N'}(\varepsilon)}{\xi_{\tau_N'}(\varepsilon)}-1\right)(1+\mathrm{o}_{\mathbb{P}}(1))+\mathrm{o}_{\mathbb{P}}(|\widehat{g}(\boldsymbol{x})-g(\boldsymbol{x})|)+\mathrm{O}_{\mathbb{P}}\left(\left|\frac{\widehat{\sigma}(\boldsymbol{x})}{\sigma(\boldsymbol{x})}-1\right|\right)\\&\quad+\mathrm{O}_{\mathbb{P}}(1/q_{\tau_N'}(\varepsilon))+\mathrm{o}_{\mathbb{P}}(|A((1-\tau_N')^{-1})|)\end{aligned}$$

by Lemma C.7(vi), the consistency assumption on $\widehat{g}$ and $\widehat{\sigma}$, and $N=N(n)\xrightarrow{\mathbb{P}}\infty$. Our bias conditions combined with the regular variation properties of $t\mapsto q_{1-t^{-1}}(\varepsilon)$ and $t\mapsto|A(t)|$ and the v_n-uniform consistency of $\widehat{g}$ and $\widehat{\sigma}$ on K_0 yield

$$\frac{\widehat{\xi}^{\star}_{\tau_N'}(Y|\boldsymbol{x})}{\xi_{\tau_N'}(Y|\boldsymbol{x})}-1=\left(\frac{\widehat{\xi}^{\star}_{\tau_N'}(\varepsilon)}{\xi_{\tau_N'}(\varepsilon)}-1\right)(1+\mathrm{o}_{\mathbb{P}}(1))+\mathrm{o}_{\mathbb{P}}(1/\sqrt{N(1-\tau_N)}).$$

Since, from the proof of (i),

$$\sqrt{N(1-\tau_N)}(\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor}-\gamma)\xrightarrow{d}\mathcal{N}(p^{-\rho}\lambda/(1-\rho),\gamma^2)$$

$$\text{and}\quad\sqrt{N(1-\tau_N)}\left(\frac{\widehat{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)}-1\right)=\mathrm{O}_{\mathbb{P}}(1),$$

a direct adaptation of the proof of Theorem 5 of [5] produces

$$\begin{aligned}\frac{\sqrt{N(1-\tau_N)}}{\log[(1-\tau_N)/(1-\tau_N')]}\left(\frac{\widehat{\xi}^{\star}_{\tau_N'}(\varepsilon)}{\xi_{\tau_N'}(\varepsilon)}-1\right)&=\sqrt{N(1-\tau_N)}(\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor}-\gamma)+\mathrm{o}_{\mathbb{P}}(1)\\&\xrightarrow{d}\mathcal{N}\left(p^{-\rho}\frac{\lambda}{1-\rho},\gamma^2\right).\end{aligned}$$

We omit the details. □

PROOF OF THEOREM 3.3. (i) Write first

$$\begin{aligned}&\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}(Y_{n+1}\,|\,\mathcal{F}_n)}{\xi_{\tau_n}(Y_{n+1}\,|\,\mathcal{F}_n)}-1\right)\\&=\frac{\xi_{\tau_n}(\varepsilon)}{\sum_{j=1}^p\phi_jY_{n+1-j}+\sum_{j=1}^q\theta_j\varepsilon_{n+1-j}+\xi_{\tau_n}(\varepsilon)}\times\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}-1\right)\end{aligned}$$

$$+\sqrt{n(1-\tau_n)}\frac{\sum_{j=1}^{p}(\widehat{\phi}_{j,n}-\phi_j)Y_{n+1-j}}{\sum_{j=1}^{p}\phi_j Y_{n+1-j}+\sum_{j=1}^{q}\theta_j\varepsilon_{n+1-j}+\xi_{\tau_n}(\varepsilon)}$$
$$+\sqrt{n(1-\tau_n)}\frac{\sum_{j=1}^{q}(\widehat{\theta}_{j,n}-\theta_j)\varepsilon_{n+1-j}}{\sum_{j=1}^{p}\phi_j Y_{n+1-j}+\sum_{j=1}^{q}\theta_j\varepsilon_{n+1-j}+\xi_{\tau_n}(\varepsilon)}$$
$$+\sqrt{n(1-\tau_n)}\frac{\sum_{j=1}^{q}\widehat{\theta}_{j,n}(\widehat{\varepsilon}_{n+1-j}^{(n)}-\varepsilon_{n+1-j})}{\sum_{j=1}^{p}\phi_j Y_{n+1-j}+\sum_{j=1}^{q}\theta_j\varepsilon_{n+1-j}+\xi_{\tau_n}(\varepsilon)}.$$

To control the gap between residuals and unobserved innovations (and hence check the high-level condition (2)), we rewrite the ARMA model in vector form, namely as $\boldsymbol{Y}_{t,p}=\boldsymbol{A}\boldsymbol{Y}_{t-1,p}-\boldsymbol{B}\boldsymbol{\varepsilon}_{t-1,q}+\boldsymbol{\varepsilon}_{t,q}$ with

$$\boldsymbol{Y}_{t,p}=\begin{pmatrix}Y_t\\ Y_{t-1}\\ \vdots\\ Y_{t-p+1}\end{pmatrix}\text{ and }\boldsymbol{A}=\begin{pmatrix}\phi_1 & \cdots & \cdots & \cdots & \phi_p\\ 1 & 0 & \cdots & \cdots & 0\\ 0 & 1 & \cdots & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & \cdots & 1 & 0\end{pmatrix},$$

$$\boldsymbol{\varepsilon}_{t,q}=\begin{pmatrix}\varepsilon_t\\ \varepsilon_{t-1}\\ \vdots\\ \varepsilon_{t-q+1}\end{pmatrix}\text{ and }\boldsymbol{B}=\begin{pmatrix}-\theta_1 & \cdots & \cdots & \cdots & -\theta_q\\ 1 & 0 & \cdots & \cdots & 0\\ 0 & 1 & \cdots & \cdots & 0\\ \vdots & \ddots & \ddots & \ddots & \vdots\\ 0 & \cdots & \cdots & 1 & 0\end{pmatrix}.$$

Set $r=\max(p,q)$. Since $\widehat{\varepsilon}_t^{(n)}=Y_t-\sum_{j=1}^{p}\widehat{\phi}_{j,n}Y_{t-j}-\sum_{j=1}^{q}\widehat{\theta}_{j,n}\widehat{\varepsilon}_{t-j}^{(n)}$ for $r+1\le t\le n$, we have $\boldsymbol{Y}_{t,p}=\widehat{\boldsymbol{A}}_n\boldsymbol{Y}_{t-1,p}-\widehat{\boldsymbol{B}}_n\widehat{\boldsymbol{\varepsilon}}_{t-1,q}^{(n)}+\widehat{\boldsymbol{\varepsilon}}_{t,q}^{(n)}$, where the notation is defined by replacing the ε_t, ϕ_j and θ_j by the $\widehat{\varepsilon}_t^{(n)}$, $\widehat{\phi}_{j,n}$ and $\widehat{\theta}_{j,n}$. It follows that for such t

$$\begin{aligned}\widehat{\boldsymbol{\varepsilon}}_{t,q}^{(n)}-\boldsymbol{\varepsilon}_{t,q}&=(\boldsymbol{A}-\widehat{\boldsymbol{A}}_n)\boldsymbol{Y}_{t-1,p}-(\boldsymbol{B}-\widehat{\boldsymbol{B}}_n)\widehat{\boldsymbol{\varepsilon}}_{t-1,q}^{(n)}+\boldsymbol{B}(\widehat{\boldsymbol{\varepsilon}}_{t-1,q}^{(n)}-\boldsymbol{\varepsilon}_{t-1,q})\\ &=\sum_{j=1}^{t-r}\boldsymbol{B}^{j-1}(\boldsymbol{A}-\widehat{\boldsymbol{A}}_n)\boldsymbol{Y}_{t-j,p}-\sum_{j=1}^{t-r}\boldsymbol{B}^{j-1}(\boldsymbol{B}-\widehat{\boldsymbol{B}}_n)\boldsymbol{\varepsilon}_{t-j,q}\\ &\quad-\sum_{j=1}^{t-r}\boldsymbol{B}^{j-1}(\boldsymbol{B}-\widehat{\boldsymbol{B}}_n)(\widehat{\boldsymbol{\varepsilon}}_{t-j,q}^{(n)}-\boldsymbol{\varepsilon}_{t-j,q})-\boldsymbol{B}^{t-r}\boldsymbol{\varepsilon}_{r,q}\end{aligned}\tag{59}$$

because $\widehat{\boldsymbol{\varepsilon}}_{r,q}^{(n)}=0$. Observe now that by causality of $(Y_t)_{t\in\mathbb{Z}}$, the Y_t have the linear representation $Y_t=\sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$, and it is a consequence of the arguments in the proof of Theorem 3.1.1 in [3] that the ψ_j define a summable series and decay geometrically fast, *i.e.* $|\psi_j|\le C\,R^j$ for real constants $C>0$ and $R\in(0,1)$. Write, for $1\le t\le n$,

$$|Y_t|\le\sum_{j=0}^{t-1}|\psi_j||\varepsilon_{t-j}|+\sum_{j=t}^{\infty}|\psi_j||\varepsilon_{t-j}|\le\left(\sum_{j=0}^{\infty}|\psi_j|\right)\max_{1\le t\le n}|\varepsilon_t|+CR\sum_{l=0}^{\infty}R^l|\varepsilon_{-l}|.$$

The last sum on the right-hand side is finite with probability 1 because ε has a finite first moment. Conclude that $\max_{1\le t\le n}|Y_t|=\mathrm{O}_{\mathbb{P}}(1+\max_{1\le t\le n}|\varepsilon_t|)$. Since the ε_t are independent and satisfy $\mathcal{C}_1(\gamma)$, we find

$$\max_{1\le t\le n}|\varepsilon_t|=\mathrm{O}_{\mathbb{P}}(n^{\gamma+\iota})\ \text{ and then }\ \max_{1\le t\le n}|Y_t|=\mathrm{O}_{\mathbb{P}}(n^{\gamma+\iota})\ \text{ for any }\iota>0,\tag{60}$$

by condition $\mathbb{P}(\varepsilon > x)/\mathbb{P}(|\varepsilon| > x) \to \ell \in (0,1]$ as $x \to \infty$, combined with Theorem 1.1.6 and Lemma 1.2.9 in [6], and Potter bounds (see *e.g.* Proposition B.1.9.5 in [6]). Notice now that $\boldsymbol{B}$ is essentially the companion matrix of the polynomial $Q(z) = 1 + \sum_{j=1}^{q} \theta_j z^j$. It is a standard exercise in linear algebra to show that $\boldsymbol{B}$ has characteristic polynomial

$$\det(\lambda \mathrm{I}_p - \boldsymbol{B}) = \lambda^q + \sum_{j=1}^{q} \theta_j \lambda^{q-j} = \lambda^q Q(1/\lambda).$$

Since Q has no root z such that $|z| \le 1$, all eigenvalues of $\boldsymbol{B}$ must then have a modulus smaller than 1, *i.e.* its spectral radius $\rho(\boldsymbol{B})$ is smaller than 1. Let $\|\cdot\|$ denote indifferently the supremum norm on $\mathbb{R}^d$ spaces and the induced operator norm on square matrices, and recall that $\|\boldsymbol{B}^j\|^{1/j} \to \rho(\boldsymbol{B})$ as $j \to \infty$ (this is in fact true for any operator norm), which means in particular that the series $\sum_{j\ge 0} \|\boldsymbol{B}^j\|$ is summable. Defining $\widehat{\varepsilon}_1^{(n)} = \cdots = \widehat{\varepsilon}_{r-q}^{(n)} = 0$ for the sake of convenience, we obtain

$$\begin{aligned}\max_{1\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| &\le \max_{r+1\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| + \max_{1\le t\le r} |\varepsilon_t| \\ &\le \|\boldsymbol{A} - \widehat{\boldsymbol{A}}_n\| \sum_{j=0}^{\infty} \|\boldsymbol{B}^j\| \max_{1\le t\le n} |Y_t| + \|\boldsymbol{B} - \widehat{\boldsymbol{B}}_n\| \sum_{j=0}^{\infty} \|\boldsymbol{B}^j\| \max_{1\le t\le n} |\varepsilon_t| \\ &\quad + \|\boldsymbol{B} - \widehat{\boldsymbol{B}}_n\| \sum_{j=0}^{\infty} \|\boldsymbol{B}^j\| \max_{1\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| + \left(1 + \sup_{j\ge 0} \|\boldsymbol{B}^j\|\right) \max_{1\le t\le r} |\varepsilon_t|.\end{aligned}$$

By $\sqrt{n}-$consistency of the $\widehat{\phi}_{j,n}$ and $\widehat{\theta}_{j,n}$, $\|\boldsymbol{A} - \widehat{\boldsymbol{A}}_n\| = \mathrm{O}_{\mathbb{P}}(n^{-1/2})$ and $\|\boldsymbol{B} - \widehat{\boldsymbol{B}}_n\| = \mathrm{O}_{\mathbb{P}}(n^{-1/2})$. Isolate then $\max_{1\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|$ to conclude that

$$\max_{1\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| = \mathrm{O}_{\mathbb{P}}\left(1 + n^{-1/2}\left[\max_{1\le t\le n} |Y_t| + \max_{1\le t\le n} |\varepsilon_t|\right]\right) = \mathrm{O}_{\mathbb{P}}(1)$$

by (60) and the assumption $\gamma < 1/2$. We now use (59) again, this time to control $\max_{t_n\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|$ to apply Theorem 2.1 (for the sample size $n - t_n + 1 = n(1 + \mathrm{o}(1))$, since the estimator $\widehat{\xi}_{\tau_n}(\varepsilon)$ is based upon the last $n - t_n + 1$ residuals). For $t \ge t_n \to \infty$, $\|\boldsymbol{B}^{t-r} \boldsymbol{\varepsilon}_{r,q}\| \le \|\boldsymbol{B}^{t-r}\| \|\boldsymbol{\varepsilon}_{r,q}\|$ and $t - r \ge t_n/2$ for n large enough; hence, by (59), the bound

$$\max_{t_n\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| = \mathrm{O}_{\mathbb{P}}\left(n^{-1/2}\left[1 + \max_{1\le t\le n} |Y_t| + \max_{1\le t\le n} |\varepsilon_t|\right] + \sup_{j\ge t_n/2} \|\boldsymbol{B}^j\|\right).$$

Since $\|\boldsymbol{B}^j\|^{1/j} \to \rho(\boldsymbol{B}) \in [0,1)$ as $j \to \infty$, we have for n large enough

$$\max_{t_n\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t| = \mathrm{O}_{\mathbb{P}}\left(n^{-1/2}\left[1 + \max_{1\le t\le n} |Y_t| + \max_{1\le t\le n} |\varepsilon_t|\right] + (1-\kappa)^{t_n}\right)$$

for some $\kappa \in (0,1)$. We have $\sqrt{n}(1-\kappa)^{t_n} \to 0$ because $t_n/\log n \to \infty$. Conclude that

$$\max_{t_n\le t\le n} \frac{|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|}{1 + |\varepsilon_t|} = \mathrm{O}_{\mathbb{P}}\left(\max_{t_n\le t\le n} |\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|\right) = \mathrm{O}_{\mathbb{P}}(n^{\gamma - 1/2 + \iota}) \text{ for all } \iota > 0$$

and therefore (2) is proved:

$$\sqrt{n(1-\tau_n)} \max_{t_n\le t\le n} \frac{|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|}{1 + |\varepsilon_t|} = \mathrm{o}_{\mathbb{P}}(1).$$

Complete the proof by combining the $\sqrt{n}-$consistency of the estimators $\widehat{\phi}_{j,n}$ and $\widehat{\theta}_{j,n}$, Lemma C.8 (an extension of Theorem 2.1 necessary here since at each step, the indices of the relevant ε_i may not be contained in those relevant to the previous step and thus, strictly speaking, we do not work with a single i.i.d. sequence) and the convergence $\xi_{\tau_n}(\varepsilon) \to \infty$.

(ii) Set

$$\widehat{\xi}^{\star}_{\tau'_n}(\varepsilon) = \left(\frac{1-\tau'_n}{1-\tau_n}\right)^{-\overline{\gamma}} \widehat{\xi}_{\tau_n}(\varepsilon)$$

and write

$$\begin{aligned}
&\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]}\left(\frac{\widehat{\xi}^{\star}_{\tau'_n}(Y_{n+1}\,|\,\mathcal{F}_n)}{\xi_{\tau'_n}(Y_{n+1}\,|\,\mathcal{F}_n)} - 1\right) \\
&= \frac{\xi_{\tau'_n}(\varepsilon)}{\sum_{j=1}^p \phi_j Y_{n+1-j} + \sum_{j=1}^q \theta_j \varepsilon_{n+1-j} + \xi_{\tau'_n}(\varepsilon)} \times \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]}\left(\frac{\widehat{\xi}^{\star}_{\tau'_n}(\varepsilon)}{\xi_{\tau'_n}(\varepsilon)} - 1\right) \\
&+ \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]} \times \frac{\sum_{j=1}^p (\widehat{\phi}_{j,n} - \phi_j) Y_{n+1-j}}{\sum_{j=1}^p \phi_j Y_{n+1-j} + \sum_{j=1}^q \theta_j \varepsilon_{n+1-j} + \xi_{\tau'_n}(\varepsilon)} \\
&+ \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]} \times \frac{\sum_{j=1}^q (\widehat{\theta}_{j,n} - \theta_j) \varepsilon_{n+1-j}}{\sum_{j=1}^p \phi_j Y_{n+1-j} + \sum_{j=1}^q \theta_j \varepsilon_{n+1-j} + \xi_{\tau'_n}(\varepsilon)} \\
&+ \frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau'_n)]} \times \frac{\sum_{j=1}^q \widehat{\theta}_{j,n} (\widehat{\varepsilon}^{(n)}_{n+1-j} - \varepsilon_{n+1-j})}{\sum_{j=1}^p \phi_j Y_{n+1-j} + \sum_{j=1}^q \theta_j \varepsilon_{n+1-j} + \xi_{\tau'_n}(\varepsilon)}.
\end{aligned}$$

Combine then what was obtained in (i) with the first convergence in Theorem 2.3. □

PROOF OF THEOREM 3.4. (i) Recall that $\widehat{\omega}_n$, the $\widehat{\alpha}_{j,n}$ and the $\widehat{\beta}_{j,n}$ are consistent estimators of (strictly) positive parameters, and thus are positive with arbitrarily high probability as $n \to \infty$. In what follows we implicitly work on this high probability event. Write

$$\begin{aligned}
\sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}(Y_{n+1}\,|\,\mathcal{F}_n)}{\xi_{\tau_n}(Y_{n+1}\,|\,\mathcal{F}_n)} - 1\right) &= \sqrt{n(1-\tau_n)}\left(\frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)} - 1\right) \\
&\quad + \sqrt{n(1-\tau_n)}\left(\frac{\widehat{\sigma}_{n+1}}{\sigma_{n+1}} - 1\right)\frac{\widehat{\xi}_{\tau_n}(\varepsilon)}{\xi_{\tau_n}(\varepsilon)}.
\end{aligned}$$

Let us first check the high-level condition (2). Define $r = \max(p,q)$. For any t with $r+1 \le t \le n$,

$$\frac{|\widehat{\varepsilon}^{(n)}_t - \varepsilon_t|}{1+|\varepsilon_t|} \le \left|\frac{\sigma_t}{\widehat{\sigma}^{(n)}_t} - 1\right| = \left|\frac{\sigma_t^2 - (\widehat{\sigma}^{(n)}_t)^2}{\widehat{\sigma}^{(n)}_t(\sigma_t + \widehat{\sigma}^{(n)}_t)}\right| \le \left|\frac{\sigma_t^2}{(\widehat{\sigma}^{(n)}_t)^2} - 1\right|. \tag{61}$$

We focus on $|(\widehat{\sigma}^{(n)}_t)^2 - \sigma_t^2|$. Note that $\boldsymbol{v}_{t,p} = \boldsymbol{Z}_{t,q} + \boldsymbol{B}\boldsymbol{v}_{t-1,p}$ with

$$\boldsymbol{v}_{t,p} = \begin{pmatrix} \sigma_t^2 \\ \sigma_{t-1}^2 \\ \vdots \\ \sigma_{t-p+1}^2 \end{pmatrix}, \quad \boldsymbol{Z}_{t,q} = \begin{pmatrix} \omega + \sum_{j=1}^q \alpha_j Y_{t-j}^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \text{and} \quad \boldsymbol{B} = \begin{pmatrix} \beta_1 & \cdots & \cdots & \cdots & \beta_p \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & 0 \end{pmatrix}.$$

Similarly $\widehat{\boldsymbol{v}}_{t,p}^{(n)}=\widehat{\boldsymbol{Z}}_{t,q}^{(n)}+\widehat{\boldsymbol{B}}_n\widehat{\boldsymbol{v}}_{t-1,p}^{(n)}$ where the notation is defined by replacing the σ_t^2, ω, the α_j and β_j by the $(\widehat{\sigma}_t^{(n)})^2$, $\widehat{\omega}_n$, the $\widehat{\alpha}_{j,n}$ and $\widehat{\beta}_{j,n}$. For $r+1\le t\le n$ then,

$$\boldsymbol{v}_{t,p}=\sum_{j=0}^{t-r-1}\boldsymbol{B}^j\boldsymbol{Z}_{t-j,q}+\boldsymbol{B}^{t-r}\boldsymbol{v}_{r,p},\ \ \widehat{\boldsymbol{v}}_{t,p}^{(n)}=\sum_{j=0}^{t-r-1}\widehat{\boldsymbol{B}}_n^j\widehat{\boldsymbol{Z}}_{t-j,q}^{(n)}+\widehat{\boldsymbol{B}}_n^{t-r}\widehat{\boldsymbol{v}}_{r,p}^{(n)}$$

and therefore

$$\begin{aligned}&\widehat{\boldsymbol{v}}_{t,p}^{(n)}-\boldsymbol{v}_{t,p}\\&=\sum_{j=0}^{t-r-1}\widehat{\boldsymbol{B}}_n^j(\widehat{\boldsymbol{Z}}_{t-j,q}^{(n)}-\boldsymbol{Z}_{t-j,q})+\sum_{j=0}^{t-r-1}(\widehat{\boldsymbol{B}}_n^j-\boldsymbol{B}^j)\boldsymbol{Z}_{t-j,q}+\widehat{\boldsymbol{B}}_n^{t-r}\widehat{\boldsymbol{v}}_{r,p}^{(n)}-\boldsymbol{B}^{t-r}\boldsymbol{v}_{r,p}.\end{aligned}$$

This readily provides

$$(\widehat{\sigma}_t^{(n)})^2=\sum_{j=0}^{t-r-1}\widehat{\boldsymbol{B}}_n^j(1,1)\left(\widehat{\omega}_n+\sum_{i=1}^{q}\widehat{\alpha}_{i,n}Y_{t-j-i}^2\right)+(\widehat{\boldsymbol{B}}_n^{t-r}\widehat{\boldsymbol{v}}_{r,p}^{(n)})(1)$$

where $\boldsymbol{u}(1)$ denotes the first element of a vector $\boldsymbol{u}$ and $\boldsymbol{A}(1,1)$ the top left element of a matrix $\boldsymbol{A}$, and similarly

$$\begin{aligned}(\widehat{\sigma}_t^{(n)})^2-\sigma_t^2=&\sum_{j=0}^{t-r-1}\widehat{\boldsymbol{B}}_n^j(1,1)\left(\widehat{\omega}_n-\omega+\sum_{i=1}^{q}(\widehat{\alpha}_{i,n}-\alpha_i)Y_{t-j-i}^2\right)\\&+\sum_{j=0}^{t-r-1}(\widehat{\boldsymbol{B}}_n^j(1,1)-\boldsymbol{B}^j(1,1))\left(\omega+\sum_{i=1}^{q}\alpha_iY_{t-j-i}^2\right)\\&+(\widehat{\boldsymbol{B}}_n^{t-r}\widehat{\boldsymbol{v}}_{r,p}^{(n)}-\boldsymbol{B}^{t-r}\boldsymbol{v}_{r,p})(1).\end{aligned}\tag{62}$$

We compare each term in (62) to $(\widehat{\sigma}_t^{(n)})^2$. First of all

$$\begin{aligned}&\frac{1}{(\widehat{\sigma}_t^{(n)})^2}\left|\sum_{j=0}^{t-r-1}\widehat{\boldsymbol{B}}_n^j(1,1)\left(\widehat{\omega}_n-\omega+\sum_{i=1}^{q}(\widehat{\alpha}_{i,n}-\alpha_i)Y_{t-j-i}^2\right)\right|\\&\le\left|\frac{\widehat{\omega}_n-\omega}{\widehat{\omega}_n}\right|+\sum_{i=1}^{q}\left|\frac{\widehat{\alpha}_{i,n}-\alpha_i}{\widehat{\alpha}_{i,n}}\right|=\mathrm{O}_{\mathbb{P}}(n^{-1/2}).\end{aligned}\tag{63}$$

Now $\boldsymbol{B}$ and $\widehat{\boldsymbol{B}}_n$ are positive matrices, so that if $\kappa_n:=\max_{1\le i\le p}|\widehat{\beta}_{i,n}-\beta_i|$, clearly $\widehat{\boldsymbol{B}}_n\le(1+\kappa_n)\boldsymbol{B}$ elementwise and thus $\widehat{\boldsymbol{B}}_n^j\le(1+\kappa_n)^j\boldsymbol{B}^j$ elementwise for any j. In particular $\widehat{\boldsymbol{B}}_n^j(1,1)\le(1+\kappa_n)^j\boldsymbol{B}^j(1,1)$ and likewise $\boldsymbol{B}^j(1,1)\le(1+\kappa_n)^j\widehat{\boldsymbol{B}}_n^j(1,1)$. Hence the bound

$$\begin{aligned}|\widehat{\boldsymbol{B}}_n^j(1,1)-\boldsymbol{B}^j(1,1)|&\le[(1+\kappa_n)^j-1]\max(\boldsymbol{B}^j(1,1),\widehat{\boldsymbol{B}}_n^j(1,1))\\&\le j\kappa_n(1+\kappa_n)^{j-1}\max(\boldsymbol{B}^j(1,1),\widehat{\boldsymbol{B}}_n^j(1,1))\\&\le j\kappa_n(1+\kappa_n)^{2j-1}\boldsymbol{B}^j(1,1).\end{aligned}\tag{64}$$

Like in the proof of Theorem 3.3, let $\|\cdot\|$ denote indifferently the supremum norm on $\mathbb{R}^d$ spaces and the induced operator norm on square matrices. Notice that $|\boldsymbol{B}^j(1,1)|\le\|\boldsymbol{B}^j\|$;

since $\|\boldsymbol{B}^j\|^{1/j} \to \rho(\boldsymbol{B}) \in [0,1)$ as $j \to \infty$ (to check that indeed the spectral radius $\rho(\boldsymbol{B}) \in [0,1)$, use Corollary 2.2 in [9]) and $\kappa_n = \mathrm{O}_{\mathbb{P}}(1/\sqrt{n})$, we have

$$\sum_{j=0}^{\infty} |\widehat{\boldsymbol{B}}_n^j(1,1) - \boldsymbol{B}^j(1,1)| = \mathrm{O}_{\mathbb{P}}(\kappa_n) = \mathrm{O}_{\mathbb{P}}(n^{-1/2}).$$

Recalling that $(\widehat{\sigma}_t^{(n)})^2 \geq \widehat{\omega}_n \xrightarrow{\mathbb{P}} \omega > 0$, we therefore obtain

$$\max_{r+1 \leq t \leq n} \frac{1}{(\widehat{\sigma}_t^{(n)})^2} \left| \sum_{j=0}^{t-r-1} (\widehat{\boldsymbol{B}}_n^j(1,1) - \boldsymbol{B}^j(1,1)) \right| = \mathrm{O}_{\mathbb{P}}(n^{-1/2}). \tag{65}$$

Next we write, for any $i \in \{1, \ldots, q\}$,

$$\frac{1}{(\widehat{\sigma}_t^{(n)})^2} \left| \sum_{j=0}^{t-r-1} (\widehat{\boldsymbol{B}}_n^j(1,1) - \boldsymbol{B}^j(1,1)) \alpha_i Y_{t-j-i}^2 \right| \leq \sum_{j=0}^{t-r-1} \frac{|\widehat{\boldsymbol{B}}_n^j(1,1) - \boldsymbol{B}^j(1,1)| \alpha_i Y_{t-j-i}^2}{\widehat{\omega}_n + \widehat{\boldsymbol{B}}_n^j(1,1) \widehat{\alpha}_{i,n} Y_{t-j-i}^2}.$$

Similarly to (64), $|\widehat{\boldsymbol{B}}_n^j(1,1) - \boldsymbol{B}^j(1,1)| \leq j\kappa_n (1+\kappa_n)^{2j-1} \widehat{\boldsymbol{B}}_n^j(1,1)$. Thus, for any $s > 0$,

$$\begin{aligned}
&\frac{1}{(\widehat{\sigma}_t^{(n)})^2} \left| \sum_{j=0}^{t-r-1} (\widehat{\boldsymbol{B}}_n^j(1,1) - \boldsymbol{B}^j(1,1)) \alpha_i Y_{t-j-i}^2 \right| \\
&\leq \kappa_n \frac{\alpha_i}{\widehat{\alpha}_{i,n}} \sum_{j=0}^{t-r-1} j(1+\kappa_n)^{2j-1} \frac{\widehat{\boldsymbol{B}}_n^j(1,1) \widehat{\alpha}_{i,n} Y_{t-j-i}^2 / \widehat{\omega}_n}{1 + \widehat{\boldsymbol{B}}_n^j(1,1) \widehat{\alpha}_{i,n} Y_{t-j-i}^2 / \widehat{\omega}_n} \\
&\leq \kappa_n \frac{\alpha_i}{\widehat{\alpha}_{i,n}} \sum_{j=0}^{t-r-1} j(1+\kappa_n)^{2j-1} \left(\frac{\widehat{\boldsymbol{B}}_n^j(1,1) \widehat{\alpha}_{i,n} Y_{t-j-i}^2}{\widehat{\omega}_n} \right)^s \\
&\leq \kappa_n \times \frac{\alpha_i}{\widehat{\alpha}_{i,n}} \left(\frac{\widehat{\alpha}_{i,n}}{\widehat{\omega}_n} \right)^s \times \sum_{j=0}^{t-r-1} j(1+\kappa_n)^{2j-1} \left((1+\kappa_n)^j \boldsymbol{B}^j(1,1) \right)^s Y_{t-j-i}^{2s}
\end{aligned}$$

where the inequality $x/(1+x) \leq x^s$, valid for any s and $x > 0$, was used. Because $|\boldsymbol{B}^j(1,1)| \leq \|\boldsymbol{B}^j\|$ and $\|\boldsymbol{B}^j\|^{1/j} \to \rho(\boldsymbol{B}) \in [0,1)$ as $j \to \infty$, as well as $\kappa_n \to 0$ in probability, we have

$$\sum_{j=0}^{\infty} j(1+\kappa_n)^{2j-1} ((1+\kappa_n)^j \boldsymbol{B}^j(1,1))^s < \infty$$

with arbitrarily high probability as $n \to \infty$. Hence the bound

$$\max_{r+1 \leq t \leq n} \frac{1}{(\widehat{\sigma}_t^{(n)})^2} \left| \sum_{j=0}^{t-r-1} (\widehat{\boldsymbol{B}}_n^j(1,1) - \boldsymbol{B}^j(1,1)) \alpha_i Y_{t-j-i}^2 \right| = \mathrm{O}_{\mathbb{P}} \left(n^{-1/2} \max_{1 \leq t \leq n} Y_t^{2s} \right)$$

valid for any $s > 0$. Recall that there is $s_0 > 0$ such that $\mathbb{E}(Y_1^{2s_0}) < \infty$ (see Corollary 2.3 p.36 in [9]). Using the identity

$$\mathbb{E}(Y_1^{2s_0}) = \int_0^{\infty} \mathbb{P}(Y_1^{2s_0} > y) \, dy < \infty$$

and noting that the function $y \mapsto \mathbb{P}(Y_1^{2s_0} > y)$ is nonnegative and nonincreasing, it is a standard exercise to show that $\mathbb{P}(Y_1^{2s_0} > y) = \mathrm{o}(y^{-1})$ as $y \to \infty$. Conclude that, for any $s \le s_0$, $n\,\mathbb{P}(Y_1^{2s} > n^{s/s_0}) = n\,\mathbb{P}(Y_1^{2s_0} > n) = \mathrm{o}(1)$, and then that

$$\mathbb{P}\left(\max_{1\le t\le n} Y_t^{2s} > n^{s/s_0}\right) \le n\,\mathbb{P}(Y_1^{2s} > n^{s/s_0}) = \mathrm{o}(1),$$

of which a consequence is that $\max_{1\le t\le n} Y_t^{2s} = \mathrm{O}_{\mathbb{P}}(n^{s/s_0})$ for any $s \le s_0$. In particular, since s can be chosen arbitrarily small,

$$\max_{r+1\le t\le n} \frac{1}{(\widehat{\sigma}_t^{(n)})^2}\left|\sum_{j=0}^{t-r-1}(\widehat{\boldsymbol{B}}_n^j(1,1) - \boldsymbol{B}^j(1,1))\alpha_i Y_{t-j-i}^2\right| = \mathrm{O}_{\mathbb{P}}(n^{\iota-1/2}) \text{ for all } \iota > 0. \tag{66}$$

Finally, for $t \ge t_n$ and n large enough,

$$\begin{aligned}&\max_{t_n\le t\le n} \frac{1}{(\widehat{\sigma}_t^{(n)})^2}|(\widehat{\boldsymbol{B}}_n^{t-r}\widehat{\boldsymbol{v}}_{r,p}^{(n)} - \boldsymbol{B}^{t-r}\boldsymbol{v}_{r,p})(1)|\\ &\le \frac{1}{\widehat{\omega}_n}\sup_{j\ge t_n/2}\left\{\|\widehat{\boldsymbol{B}}_n^j\| + \|\boldsymbol{B}^j\|\right\}\max_{r-p+1\le t\le r}\left\{\widehat{\sigma}_t^{(n)} + \sigma_t\right\} = \mathrm{O}_{\mathbb{P}}\left(\sup_{j\ge t_n/2}\left\{\|\widehat{\boldsymbol{B}}_n^j\| + \|\boldsymbol{B}^j\|\right\}\right)\end{aligned}$$

by consistency of $\widehat{\omega}_n$, definition of $\widehat{\sigma}_{r-p+1}^{(n)}, \ldots, \widehat{\sigma}_r^{(n)}$ and finiteness of at least a fractional moment of the σ_t (and hence finiteness of the σ_t with probability 1; see Corollary 2.3 p.36 in [9]). Besides, it is a simple exercise in linear algebra to show that for a $d \times d$ matrix with nonnegative elements, $\|A\| = \max_{1\le i\le d}\sum_{j=1}^d A(i,j)$; consequently

$$\max_{t_n\le t\le n} \frac{1}{(\widehat{\sigma}_t^{(n)})^2}|(\widehat{\boldsymbol{B}}_n^{t-r}\widehat{\boldsymbol{v}}_{r,p}^{(n)} - \boldsymbol{B}^{t-r}\boldsymbol{v}_{r,p})(1)| = \mathrm{O}_{\mathbb{P}}\left(\sup_{j\ge t_n/2}\left\{(1+\kappa_n)^j\|\boldsymbol{B}^j\|\right\}\right).$$

Recall that $\|\boldsymbol{B}^j\|^{1/j} \to \rho(\boldsymbol{B}) \in [0,1)$ as $j \to \infty$ and $\kappa_n \to 0$ in probability, so that

$$\max_{t_n\le t\le n} \frac{1}{(\widehat{\sigma}_t^{(n)})^2}|(\widehat{\boldsymbol{B}}_n^{t-r}\widehat{\boldsymbol{v}}_{r,p}^{(n)} - \boldsymbol{B}^{t-r}\boldsymbol{v}_{r,p})(1)| = \mathrm{O}_{\mathbb{P}}(n^{-1/2}) \tag{67}$$

because $t_n/\log n \to \infty$. Combine (61), (62), (63), (65), (66), (67) and recall that $\tau_n = 1 - n^{-a}$ to find

$$\sqrt{n(1-\tau_n)}\max_{t_n\le t\le n}\left|\frac{\sigma_t^2}{(\widehat{\sigma}_t^{(n)})^2} - 1\right| \xrightarrow{\mathbb{P}} 0 \text{ and then } \sqrt{n(1-\tau_n)}\max_{t_n\le t\le n}\frac{|\widehat{\varepsilon}_t^{(n)} - \varepsilon_t|}{1+|\varepsilon_t|} \xrightarrow{\mathbb{P}} 0$$

by (61). Condition (2) thus holds. Second, the inequality $|\widehat{\sigma}_{n+1}/\sigma_{n+1} - 1| \le |\widehat{\sigma}_{n+1}^2/\sigma_{n+1}^2 - 1|$ and a similar argument yield

$$\sqrt{n(1-\tau_n)}\left|\frac{\widehat{\sigma}_{n+1}}{\sigma_{n+1}} - 1\right| \xrightarrow{\mathbb{P}} 0.$$

Conclude by applying Lemma C.8 (for the sample size $s_n = n - t_n + 1 = n(1+\mathrm{o}(1))$, since the estimator $\widehat{\xi}_{\tau_n}(\varepsilon)$ is based upon the last $n - t_n + 1$ residuals; this array version of Theorem 2.1 is necessary once again here).

(ii) Set

$$\widehat{\xi}_{\tau_n'}^{\star}(\varepsilon) = \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\overline{\gamma}}\widehat{\xi}_{\tau_n}(\varepsilon)$$

and write

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{\xi}^{\star}_{\tau_n'}(Y_{n+1}\,|\,\mathcal{F}_n)}{\xi_{\tau_n'}(Y_{n+1}\,|\,\mathcal{F}_n)}-1\right)$$
$$=\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{\xi}^{\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}-1\right)$$
$$+\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{\sigma}_{n+1}}{\sigma_{n+1}}-1\right)\frac{\widehat{\xi}^{\star}_{\tau_n'}(\varepsilon)}{\xi_{\tau_n'}(\varepsilon)}.$$

To conclude, combine (i) with the relationship $\widehat{\sigma}_{n+1}/\sigma_{n+1}=1+\mathrm{O}_{\mathbb{P}}(1/\sqrt{n(1-\tau_n)})$ and the first convergence in Theorem 2.3. □

APPENDIX E: ADDITIONAL RESULTS ON INDIRECT ESTIMATORS AND THEIR PROOFS

This section focuses on the indirect versions of our extreme expectile estimators. The first result is an analogue of Corollary 3.1 in the heteroscedastic linear regression model (M_1), for the indirect estimators $\widetilde{\xi}_{\tau_n}(Y|\boldsymbol{x})$ and $\widetilde{\xi}^{\star}_{\tau_n'}(Y|\boldsymbol{x})$ defined as

$$\widetilde{\xi}_{\tau_n}(Y|\boldsymbol{x})=\widehat{\alpha}+\widehat{\boldsymbol{\beta}}^{\top}\boldsymbol{x}+(1+\widehat{\boldsymbol{\theta}}^{\top}\boldsymbol{x})(\overline{\gamma}^{-1}-1)^{-\overline{\gamma}}\widehat{\varepsilon}^{(n)}_{n-\lfloor n(1-\tau_n)\rfloor,n}$$
$$\text{and }\ \widetilde{\xi}^{\star}_{\tau_n'}(Y|\boldsymbol{x})=\widehat{\alpha}+\widehat{\boldsymbol{\beta}}^{\top}\boldsymbol{x}+(1+\widehat{\boldsymbol{\theta}}^{\top}\boldsymbol{x})\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\overline{\gamma}}(\overline{\gamma}^{-1}-1)^{-\overline{\gamma}}\widehat{\varepsilon}^{(n)}_{n-\lfloor n(1-\tau_n)\rfloor,n}.$$

Here $\overline{\gamma}=\widehat{\gamma}_{\lfloor n(1-\tau_n)\rfloor}$ is assumed to be the Hill estimator based on residuals, as in Section 2.2.

COROLLARY E.1. *Assume that the setup is that of the heteroscedastic linear model* (M_1). *Suppose that* $\mathbb{E}|\varepsilon_-|^2<\infty$. *Assume further that* ε *satisfies condition* $\mathcal{C}_2(\gamma,\rho,A)$ *with* $0<\gamma<1/2$, $\rho<0$, *and that* $\tau_n,\tau_n'\uparrow 1$ *satisfy* (3) *and* (4). *Then for any* $\boldsymbol{x}\in K$,

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}(Y|\boldsymbol{x})}{\xi_{\tau_n}(Y|\boldsymbol{x})}-1\right)\xrightarrow{d}\mathcal{N}\left(\lambda\left[\frac{m(\gamma)}{1-\rho}-b(\gamma,\rho)\right],\gamma^2\left[1+[m(\gamma)]^2\right]\right),$$

with the notation of Corollary 2.1, and

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{\xi}^{\star}_{\tau_n'}(Y|\boldsymbol{x})}{\xi_{\tau_n'}(Y|\boldsymbol{x})}-1\right)\xrightarrow{d}\mathcal{N}\left(\frac{\lambda}{1-\rho},\gamma^2\right).$$

PROOF OF COROLLARY E.1. To obtain the first convergence, repeat the proof of Corollary 3.1, with $\widehat{\xi}_{\tau_n}(\varepsilon)$ replaced by $\widetilde{\xi}_{\tau_n}(\varepsilon)=(\widehat{\gamma}^{-1}_{\lfloor n(1-\tau_n)\rfloor}-1)^{-\widehat{\gamma}_{\lfloor n(1-\tau_n)\rfloor}}\,\widehat{\varepsilon}^{(n)}_{n-\lfloor n(1-\tau_n)\rfloor,n}$, and apply Corollary 2.1 rather than Theorem 2.1. The second convergence is obtained by combining the first convergence with Theorem 2.3. □

The second result considers, in the context of the heteroscedastic single-index model (M_2), the indirect estimators $\widetilde{\xi}_{\tau_N}(Y|\boldsymbol{x})$ and $\widetilde{\xi}^{\star}_{\tau_N'}(Y|\boldsymbol{x})$ defined, for an $\boldsymbol{x}\in K_0$, as

$$\widetilde{\xi}_{\tau_N}(Y|\boldsymbol{x})=\widehat{g}_{h_n,t_n}(\widehat{\boldsymbol{\beta}}^{\top}\boldsymbol{x})+\widehat{\sigma}_{h_n,t_n}(\widehat{\boldsymbol{\beta}}^{\top}\boldsymbol{x})(\overline{\gamma}^{-1}-1)^{-\overline{\gamma}}\widehat{\varepsilon}^{(n)}_{N-\lfloor N(1-\tau_N)\rfloor,N,K_0}$$

at the intermediate level, and

$$\widetilde{\xi}^{\star}_{\tau_N'}(Y|\boldsymbol{x})=\widehat{g}_{h_n,t_n}(\widehat{\boldsymbol{\beta}}^{\top}\boldsymbol{x})+\widehat{\sigma}_{h_n,t_n}(\widehat{\boldsymbol{\beta}}^{\top}\boldsymbol{x})\left(\frac{1-\tau_N'}{1-\tau_N}\right)^{-\overline{\gamma}}(\overline{\gamma}^{-1}-1)^{-\overline{\gamma}}\widehat{\varepsilon}^{(n)}_{N-\lfloor N(1-\tau_N)\rfloor,N,K_0}$$

at the extreme level. Here $\overline{\gamma} = \widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor}$ is assumed to be the Hill estimator based on the random number of residuals $\lfloor N(1-\tau_N)\rfloor$ where $N = \sum_{i=1}^n \mathbb{1}\{\boldsymbol{X}_i \in K_0\}$.

THEOREM E.1. *Work in model (M_2). Assume that ε satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$ with $0<\gamma<1/2$ and $\rho<0$ and that the conditions of Proposition C.1 in Appendix C hold. Let K_0 be a compact subset of K° such that $\mathbb{P}(\boldsymbol{X}\in K_0)>0$, and $N=N(K_0,n)$. In addition, suppose that the sequences $\tau_n = 1-n^{-a}$ with $a\in(1/5,1)$ and $\tau_n'\uparrow 1$ satisfy* (3) *and* (4). *Then, for any $\boldsymbol{x}\in K_0$,*

$$\sqrt{N(1-\tau_N)}\left(\frac{\widetilde{\xi}_{\tau_N}(Y|\boldsymbol{x})}{\xi_{\tau_N}(Y|\boldsymbol{x})}-1\right) \xrightarrow{d} \mathcal{N}\left(\lambda\left[\frac{m(\gamma)}{1-\rho}-b(\gamma,\rho)\right], \gamma^2\left[1+[m(\gamma)]^2\right]\right),$$

with the notation of Corollary 2.1, and

$$\frac{\sqrt{N(1-\tau_N)}}{\log[(1-\tau_N)/(1-\tau_N')]}\left(\frac{\widetilde{\xi}^\star_{\tau_N'}(Y|\boldsymbol{x})}{\xi_{\tau_N'}(Y|\boldsymbol{x})}-1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

PROOF OF THEOREM E.1. Combine Corollary 2.1 with the de-conditioning Lemma C.4(i) to obtain

$$\sqrt{N(1-\tau_N)}\left(\frac{\widetilde{\xi}_{\tau_N}(\varepsilon)}{\xi_{\tau_N}(\varepsilon)}-1\right) \xrightarrow{d} \mathcal{N}\left(\lambda\left[\frac{m(\gamma)}{1-\rho}-b(\gamma,\rho)\right], \gamma^2\left[1+[m(\gamma)]^2\right]\right),$$

where $\widetilde{\xi}_{\tau_N}(\varepsilon) = (\widehat{\gamma}^{-1}_{\lfloor N(1-\tau_N)\rfloor}-1)^{-\widehat{\gamma}_{\lfloor N(1-\tau_N)\rfloor}}\,\widehat{\varepsilon}^{(n)}_{N-\lfloor N(1-\tau_N)\rfloor,N,K_0}$. Complete the proof by following the final four lines of the proof of Theorem 3.1(i) (this crucially relies on the assumptions of Proposition C.1) and the proof of Theorem 3.1(ii). □

The third result focuses on the indirect estimators

$$\widetilde{\xi}_{\tau_n}(Y_{n+1}\,|\,\mathcal{F}_n) = \sum_{j=1}^p \widehat{\phi}_{j,n}Y_{n+1-j} + \sum_{j=1}^q \widehat{\theta}_{j,n}\widehat{\varepsilon}^{(n)}_{n+1-j} + (\overline{\gamma}^{-1}-1)^{-\overline{\gamma}}\overline{q}_{\tau_n}(\varepsilon)$$

and $$\widetilde{\xi}^\star_{\tau_n'}(Y_{n+1}\,|\,\mathcal{F}_n) = \sum_{j=1}^p \widehat{\phi}_{j,n}Y_{n+1-j} + \sum_{j=1}^q \widehat{\theta}_{j,n}\widehat{\varepsilon}^{(n)}_{n+1-j} + \left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\overline{\gamma}}(\overline{\gamma}^{-1}-1)^{-\overline{\gamma}}\overline{q}_{\tau_n}(\varepsilon)$$

in the ARMA(p,q) model (T_1). Here $\overline{q}_{\tau_n}(\varepsilon) = \widehat{\varepsilon}^{(n)}_{n-t_n+1-\lfloor(n-t_n+1)(1-\tau_n)\rfloor,n-t_n+1}$ is a top order statistic of the last $n-t_n+1$ residuals $\widehat{\varepsilon}^{(n)}_{t_n},\widehat{\varepsilon}^{(n)}_{t_n+1},\dots,\widehat{\varepsilon}^{(n)}_n$, with $t_n/\log n\to\infty$ and $t_n/n\to 0$, and $\overline{\gamma}$ is assumed to be the Hill estimator based on these residuals.

THEOREM E.2. *Work in model (T_1). Assume further that ε satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$ with $0<\gamma<1/2$ and $\rho<0$, and that $\tau_n,\tau_n'\uparrow 1$ satisfy* (3) *and* (4). *If moreover $n^{2\gamma+\iota}(1-\tau_n)\to 0$ for some $\iota>0$, then*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}(Y_{n+1}\,|\,\mathcal{F}_n)}{\xi_{\tau_n}(Y_{n+1}\,|\,\mathcal{F}_n)}-1\right) \xrightarrow{d} \mathcal{N}\left(\lambda\left[\frac{m(\gamma)}{1-\rho}-b(\gamma,\rho)\right], \gamma^2\left[1+[m(\gamma)]^2\right]\right),$$

with the notation of Corollary 2.1, and

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{\xi}^\star_{\tau_n'}(Y_{n+1}\,|\,\mathcal{F}_n)}{\xi_{\tau_n'}(Y_{n+1}\,|\,\mathcal{F}_n)}-1\right) \xrightarrow{d} \mathcal{N}\left(\frac{\lambda}{1-\rho}, \gamma^2\right).$$

PROOF OF THEOREM E.2. Mimic the proof of Theorem 3.3, applying (an array version of) Corollary 2.1 rather than Lemma C.8. □

The fourth and final result gives the asymptotic properties of the indirect estimators

$$\widetilde{\xi}_{\tau_n}(Y_{n+1}\,|\,\mathcal{F}_n) = \widehat{\sigma}_{n+1}(\overline{\gamma}^{-1}-1)^{-\overline{\gamma}}\overline{q}_{\tau_n}(\varepsilon)$$
$$\text{and}\ \ \widetilde{\xi}^{\star}_{\tau_n'}(Y_{n+1}\,|\,\mathcal{F}_n) = \widehat{\sigma}_{n+1}\times\left(\frac{1-\tau_n'}{1-\tau_n}\right)^{-\overline{\gamma}}(\overline{\gamma}^{-1}-1)^{-\overline{\gamma}}\overline{q}_{\tau_n}(\varepsilon)$$

in the GARCH(p,q) model (T_2), where again $\overline{q}_{\tau_n}(\varepsilon) = \widehat{\varepsilon}^{(n)}_{n-t_n+1-\lfloor (n-t_n+1)(1-\tau_n)\rfloor,n-t_n+1}$ is a top order statistic of the last $n-t_n+1$ residuals $\widehat{\varepsilon}^{(n)}_{t_n},\widehat{\varepsilon}^{(n)}_{t_n+1},\ldots,\widehat{\varepsilon}^{(n)}_{n}$, with $t_n/\log n\to\infty$ and $t_n/n\to 0$, and $\overline{\gamma}$ is assumed to be the Hill estimator based on these residuals.

THEOREM E.3. *Work in model (T_2). Assume further that ε satisfies condition $\mathcal{C}_2(\gamma,\rho,A)$ with $0<\gamma<1/2$ and $\rho<0$. Suppose also that $\tau_n,\tau_n'\uparrow 1$ satisfy* (3) *and* (4) *with $\tau_n = 1-n^{-a}$ for $a\in(0,1)$. Then*

$$\sqrt{n(1-\tau_n)}\left(\frac{\widetilde{\xi}_{\tau_n}(Y_{n+1}\,|\,\mathcal{F}_n)}{\xi_{\tau_n}(Y_{n+1}\,|\,\mathcal{F}_n)}-1\right)\stackrel{d}{\longrightarrow}\mathcal{N}\left(\lambda\left[\frac{m(\gamma)}{1-\rho}-b(\gamma,\rho)\right],\gamma^2\left[1+[m(\gamma)]^2\right]\right),$$

with the notation of Corollary 2.1, and

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widetilde{\xi}^{\star}_{\tau_n'}(Y_{n+1}\,|\,\mathcal{F}_n)}{\xi_{\tau_n'}(Y_{n+1}\,|\,\mathcal{F}_n)}-1\right)\stackrel{d}{\longrightarrow}\mathcal{N}\left(\frac{\lambda}{1-\rho},\gamma^2\right).$$

PROOF OF THEOREM E.3. Mimic the proof of Theorem 3.4, applying (an array version of) Corollary 2.1 rather than Lemma C.8. □

APPENDIX F: FINITE-SAMPLE STUDY: DETAILS ON COMPUTATIONAL PROCEDURES AND FURTHER FINITE-SAMPLE RESULTS

F.1. Optimal choice of the intermediate level τ_n. In the calculation of our extreme value estimates, the intermediate level τ_n is a tuning parameter that has to be chosen. This is of course essentially equivalent to choosing the parameter $k_n = \lfloor n(1-\tau_n)\rfloor$ representing the effective sample size in the Hill estimator used for the extrapolation. There are various ways of choosing k_n; we briefly discuss here a procedure based on an asymptotic mean-squared error minimisation criterion. As highlighted in Equation (3.2.13) p.77 in [6], the asymptotic mean-squared error of the Hill estimator under $\mathcal{C}_2(\gamma,\rho,A)$ is:

$$\text{AMSE}(k_n) := \frac{1}{(1-\rho)^2}\left[A\left(\frac{n}{k_n}\right)\right]^2+\frac{\gamma^2}{k_n}.$$

Let us consider the typical case of an auxiliary function $A(t) = b\gamma t^{\rho}$, as in our simulation study. Minimising the AMSE with respect to k_n yields an optimal value k_n^* given by

$$k_n^* = \left\lfloor\left(\frac{(1-\rho)^2}{-2\rho b^2}\right)^{1/(1-2\rho)} n^{-2\rho/(1-2\rho)}\right\rfloor.$$

This optimal value of k_n fulfills the well-known bias-variance trade-off in extreme value analysis, by balancing in an optimal way the variance increasing with low k_n and the bias increasing with high k_n. In practice, this value of k_n^* is of course unavailable because it depends on the unknown values of γ, b and ρ. In our simulation study where a sample of

$n = 1{,}000$ data points is available, we therefore suggest to use the sample counterpart $\widehat{k}_n^*$ of k_n^* obtained through plugging in a prior estimate of γ calculated using the bias-reduced Hill estimator with $k_n = n/10 = 100$, along with estimates of b and ρ obtained using the function `mop` from the R package `evt0`, all based of course on residuals of the model rather than the unobservable noise variables.

To check the quality of the estimation with this choice $\widehat{k}_n^*$ of k_n, we repeated our simulation studies in Sections 4.1 and 4.2, with the same parameters but with $\widehat{k}_n^*$ in place of $k_n = 100$. Results are reported in Tables F.2 and F.4. It is readily seen there that there is no obvious advantage in using a data-driven criterion for the choice of k_n, and in fact results tend to be slightly worse. This is most likely because a data-driven choice of k_n is itself random and therefore may contribute to estimation uncertainty.

Model	Procedure	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
Linear (G1)	(S1)	$2.29 \cdot 10^{-2}$	$3.56 \cdot 10^{-2}$	$6.46 \cdot 10^{-2}$	$1.13 \cdot 10^{-1}$
	(S1i)	$1.37 \cdot 10^{-2}$	$3.14 \cdot 10^{-2}$	$6.51 \cdot 10^{-2}$	$1.21 \cdot 10^{-1}$
	(S2)	$2.73 \cdot 10^{-2}$	$3.76 \cdot 10^{-2}$	$6.17 \cdot 10^{-2}$	$9.86 \cdot 10^{-2}$
	(S2i)	$3.11 \cdot 10^{-2}$	$3.57 \cdot 10^{-2}$	$5.93 \cdot 10^{-2}$	$1.05 \cdot 10^{-1}$
	(B1)	$1.26 \cdot 10^{-1}$	$8.06 \cdot 10^{-2}$	$9.89 \cdot 10^{-2}$	$1.93 \cdot 10^{-1}$
	(B1i)	$1.58 \cdot 10^{-1}$	$7.85 \cdot 10^{-2}$	$9.75 \cdot 10^{-2}$	$1.96 \cdot 10^{-1}$
	(B2)	$1.22 \cdot 10^{-1}$	$1.09 \cdot 10^{-1}$	$9.90 \cdot 10^{-2}$	$1.08 \cdot 10^{-1}$
	(B3)	$2.52 \cdot 10^{-2}$	$3.93 \cdot 10^{-2}$	$6.78 \cdot 10^{-2}$	$1.16 \cdot 10^{-1}$
	(B4)	$4.82 \cdot 10^{-2}$	$4.13 \cdot 10^{-2}$	$6.34 \cdot 10^{-2}$	$1.04 \cdot 10^{-1}$
	(B4i)	$8.15 \cdot 10^{-3}$	$2.73 \cdot 10^{-2}$	$6.23 \cdot 10^{-2}$	$1.18 \cdot 10^{-1}$
	(B5)	$2.26 \cdot 10^{-2}$	$3.53 \cdot 10^{-2}$	$6.23 \cdot 10^{-2}$	$1.06 \cdot 10^{-1}$
	(B5i)	$9.47 \cdot 10^{-3}$	$3.09 \cdot 10^{-2}$	$6.38 \cdot 10^{-2}$	$1.12 \cdot 10^{-1}$
Single index (G2)	(S1)	$1.83 \cdot 10^{-1}$	$1.10 \cdot 10^{-1}$	$8.13 \cdot 10^{-2}$	$1.09 \cdot 10^{-1}$
	(S1i)	$1.96 \cdot 10^{-1}$	$1.18 \cdot 10^{-1}$	$6.97 \cdot 10^{-2}$	$1.01 \cdot 10^{-1}$
	(S2)	$3.90 \cdot 10^{-2}$	$4.38 \cdot 10^{-2}$	$6.89 \cdot 10^{-2}$	$1.08 \cdot 10^{-1}$
	(S2i)	$5.75 \cdot 10^{-2}$	$4.27 \cdot 10^{-2}$	$6.53 \cdot 10^{-2}$	$1.08 \cdot 10^{-1}$
	(B1)	$1.43 \cdot 10^{-1}$	$8.89 \cdot 10^{-2}$	$1.18 \cdot 10^{-1}$	$2.06 \cdot 10^{-1}$
	(B1i)	$1.74 \cdot 10^{-1}$	$7.64 \cdot 10^{-2}$	$1.14 \cdot 10^{-1}$	$2.05 \cdot 10^{-1}$
	(B2)	$3.46 \cdot 10^{-1}$	$2.79 \cdot 10^{-1}$	$2.37 \cdot 10^{-1}$	$1.95 \cdot 10^{-1}$
	(B3)	$2.97 \cdot 10^{-2}$	$4.20 \cdot 10^{-2}$	$7.20 \cdot 10^{-2}$	$1.20 \cdot 10^{-1}$
	(B4)	$5.84 \cdot 10^{-2}$	$4.82 \cdot 10^{-2}$	$7.14 \cdot 10^{-2}$	$1.13 \cdot 10^{-1}$
	(B4i)	$9.86 \cdot 10^{-3}$	$3.19 \cdot 10^{-2}$	$7.01 \cdot 10^{-2}$	$1.28 \cdot 10^{-1}$
	(B5)	$2.73 \cdot 10^{-2}$	$4.12 \cdot 10^{-2}$	$7.01 \cdot 10^{-2}$	$1.15 \cdot 10^{-1}$
	(B5i)	$1.15 \cdot 10^{-2}$	$3.61 \cdot 10^{-2}$	$7.18 \cdot 10^{-2}$	$1.22 \cdot 10^{-1}$

TABLE F.1

RMAD of methods (S1), (S2), (S1i) and (S2i), and of benchmarks (B1)–(B5i), in models (G1)–(G2). Estimators based on the fixed intermediate level $k_n = n/10 = 100$.

F.2. Pointwise confidence interval construction. We have explained, following our simulation studies in Sections 4.1 and 4.2, that most of the uncertainty in the problem of estimating extreme conditional expectiles appears indeed to come from the extreme value step. This seems to be particularly the case as soon as $\gamma \geq 0.2$. One may then use the asymptotic results developed in this paper to carry out pointwise inference about extreme conditional quantiles. Indeed, in typical cases the limit law in Theorem 2.3 is standard, and in fact is even Gaussian, because it is the limiting distribution of the extreme value index estimator $\overline{\gamma}$; under their respective suitable conditions, all common extreme value index estimators are asymptotically Gaussian. This is the case for the Hill estimator, of course, as we state in our Corollary 2.1, but also for, among others, the Pickands estimator, the Maximum Likelihood estimator constructed using the Generalised Pareto approximation, the moment estimator

Model	Procedure	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
Linear (G1)	(S1)	$2.30 \cdot 10^{-2}$	$3.82 \cdot 10^{-2}$	$6.55 \cdot 10^{-2}$	$1.18 \cdot 10^{-1}$
	(S1i)	$1.45 \cdot 10^{-2}$	$3.36 \cdot 10^{-2}$	$6.75 \cdot 10^{-2}$	$1.25 \cdot 10^{-1}$
	(S2)	$2.87 \cdot 10^{-2}$	$3.98 \cdot 10^{-2}$	$6.39 \cdot 10^{-2}$	$1.06 \cdot 10^{-1}$
	(S2i)	$3.25 \cdot 10^{-2}$	$3.62 \cdot 10^{-2}$	$6.11 \cdot 10^{-2}$	$1.08 \cdot 10^{-1}$
	(B1)	$1.26 \cdot 10^{-1}$	$8.06 \cdot 10^{-2}$	$9.89 \cdot 10^{-2}$	$1.93 \cdot 10^{-1}$
	(B1i)	$1.58 \cdot 10^{-1}$	$7.85 \cdot 10^{-2}$	$9.75 \cdot 10^{-2}$	$1.96 \cdot 10^{-1}$
	(B2)	$1.27 \cdot 10^{-1}$	$1.09 \cdot 10^{-1}$	$1.01 \cdot 10^{-1}$	$1.15 \cdot 10^{-1}$
	(B3)	$2.43 \cdot 10^{-2}$	$3.98 \cdot 10^{-2}$	$7.31 \cdot 10^{-2}$	$1.26 \cdot 10^{-1}$
	(B4)	$4.82 \cdot 10^{-2}$	$4.58 \cdot 10^{-2}$	$5.97 \cdot 10^{-2}$	$1.07 \cdot 10^{-1}$
	(B4i)	$9.07 \cdot 10^{-3}$	$3.12 \cdot 10^{-2}$	$6.90 \cdot 10^{-2}$	$1.31 \cdot 10^{-1}$
	(B5)	$2.39 \cdot 10^{-2}$	$3.67 \cdot 10^{-2}$	$6.39 \cdot 10^{-2}$	$1.04 \cdot 10^{-1}$
	(B5i)	$9.65 \cdot 10^{-3}$	$3.15 \cdot 10^{-2}$	$6.41 \cdot 10^{-2}$	$1.11 \cdot 10^{-1}$
Single index (G2)	(S1)	$1.84 \cdot 10^{-1}$	$1.11 \cdot 10^{-1}$	$7.96 \cdot 10^{-2}$	$1.10 \cdot 10^{-1}$
	(S1i)	$1.96 \cdot 10^{-1}$	$1.19 \cdot 10^{-1}$	$7.08 \cdot 10^{-2}$	$1.03 \cdot 10^{-1}$
	(S2)	$4.04 \cdot 10^{-2}$	$4.43 \cdot 10^{-2}$	$6.91 \cdot 10^{-2}$	$1.11 \cdot 10^{-1}$
	(S2i)	$5.86 \cdot 10^{-2}$	$4.37 \cdot 10^{-2}$	$6.51 \cdot 10^{-2}$	$1.09 \cdot 10^{-1}$
	(B1)	$1.43 \cdot 10^{-1}$	$8.89 \cdot 10^{-2}$	$1.18 \cdot 10^{-1}$	$2.06 \cdot 10^{-1}$
	(B1i)	$1.74 \cdot 10^{-1}$	$7.64 \cdot 10^{-2}$	$1.14 \cdot 10^{-1}$	$2.05 \cdot 10^{-1}$
	(B2)	$3.48 \cdot 10^{-1}$	$2.79 \cdot 10^{-1}$	$2.32 \cdot 10^{-1}$	$1.91 \cdot 10^{-1}$
	(B3)	$2.92 \cdot 10^{-2}$	$4.38 \cdot 10^{-2}$	$7.85 \cdot 10^{-2}$	$1.32 \cdot 10^{-1}$
	(B4)	$5.84 \cdot 10^{-2}$	$5.35 \cdot 10^{-2}$	$6.72 \cdot 10^{-2}$	$1.17 \cdot 10^{-1}$
	(B4i)	$1.10 \cdot 10^{-2}$	$3.64 \cdot 10^{-2}$	$7.76 \cdot 10^{-2}$	$1.43 \cdot 10^{-1}$
	(B5)	$2.89 \cdot 10^{-2}$	$4.28 \cdot 10^{-2}$	$7.18 \cdot 10^{-2}$	$1.14 \cdot 10^{-1}$
	(B5i)	$1.17 \cdot 10^{-2}$	$3.68 \cdot 10^{-2}$	$7.21 \cdot 10^{-2}$	$1.20 \cdot 10^{-1}$

TABLE F.2

RMAD of methods (S1), (S2), (S1i) and (S2i), and of benchmarks (B1)–(B5i), in models (G1)–(G2). Estimators based on the data-driven intermediate level $\widehat{k}_n^*$.

of [7] and probability weighted moment estimators (see respectively Theorems 3.3.5, 3.4.2, 3.5.4 and 3.6.1 in [6]). Asymptotic bias terms depend on γ, the second-order parameter ρ and the auxiliary function A, while asymptotic variances are functions of γ only. For instance, if $\overline{\gamma}$ is the Hill estimator $\widehat{\gamma}_{\lfloor n(1-\tau_n)\rfloor}$ as in Corollary 2.1, Theorem 2.3 reads, in model (1),

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\overline{\xi}^{\star}_{\tau_n'}(Y|\boldsymbol{x})}{\xi_{\tau_n'}(Y|\boldsymbol{x})}-1\right)\stackrel{d}{\longrightarrow}\mathcal{N}\left(\frac{\lambda}{1-\rho},\gamma^2\right).$$

Consistent estimators of ρ and A are available from the work of [11], adapted here by using residuals instead of the unobserved errors. In each case the asymptotic bias and variance terms can then be estimated, and carrying out inference on the extreme conditional expectile of interest is, in principle, straightforward.

For consistency with our finite-sample studies and especially our real data analyses, we discuss the implementation of such confidence intervals based on the bias-reduced estimators $\widehat{\gamma}_k^{\mathrm{RB}}$, obtained by a bias reduction of the Hill estimator $\widehat{\gamma}_k$ (where throughout $k=\lfloor n(1-\tau_n)\rfloor$) and $\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)$, obtained by a bias reduction of the direct extrapolated estimator $\widehat{\xi}^{\star}_{\tau_n'}(\varepsilon)$, whose expression can be found at the beginning of Section 4. Combined with appropriate model structure estimators converging quickly enough, these naturally give rise to an estimator $\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(Y|\boldsymbol{x})$ which, by Theorem 2.3, should satisfy

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]}\left(\frac{\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(Y|\boldsymbol{x})}{\xi_{\tau_n'}(Y|\boldsymbol{x})}-1\right)\stackrel{d}{\longrightarrow}\mathcal{N}(0,\gamma^2).$$

Model	Parameters	Estimator	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
ARMA	$(\phi, \theta) = (0.1, 0.1)$	Direct	$4.75 \cdot 10^{-2}$	$6.31 \cdot 10^{-2}$	$9.57 \cdot 10^{-2}$	$1.37 \cdot 10^{-1}$
	(estimated)	Indirect	$3.00 \cdot 10^{-2}$	$5.43 \cdot 10^{-2}$	$9.47 \cdot 10^{-2}$	$1.50 \cdot 10^{-1}$
	$(\phi, \theta) = (0.1, 0.1)$	Direct	$4.49 \cdot 10^{-2}$	$6.06 \cdot 10^{-2}$	$9.30 \cdot 10^{-2}$	$1.37 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.96 \cdot 10^{-2}$	$5.32 \cdot 10^{-2}$	$9.62 \cdot 10^{-2}$	$1.48 \cdot 10^{-1}$
	$(\phi, \theta) = (0.1, 0.5)$	Direct	$4.69 \cdot 10^{-2}$	$6.25 \cdot 10^{-2}$	$9.57 \cdot 10^{-1}$	$1.38 \cdot 10^{-1}$
	(estimated)	Indirect	$3.09 \cdot 10^{-2}$	$5.36 \cdot 10^{-2}$	$9.88 \cdot 10^{-2}$	$1.49 \cdot 10^{-1}$
	$(\phi, \theta) = (0.1, 0.5)$	Direct	$4.46 \cdot 10^{-2}$	$6.30 \cdot 10^{-2}$	$9.37 \cdot 10^{-2}$	$1.36 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$2.04 \cdot 10^{-2}$	$5.45 \cdot 10^{-2}$	$9.51 \cdot 10^{-2}$	$1.45 \cdot 10^{-1}$
	$(\phi, \theta) = (0.5, 0.1)$	Direct	$4.93 \cdot 10^{-2}$	$6.51 \cdot 10^{-2}$	$9.59 \cdot 10^{-2}$	$1.37 \cdot 10^{-1}$
	(estimated)	Indirect	$3.14 \cdot 10^{-2}$	$5.79 \cdot 10^{-2}$	$1.01 \cdot 10^{-1}$	$1.50 \cdot 10^{-1}$
	$(\phi, \theta) = (0.5, 0.1)$	Direct	$4.53 \cdot 10^{-2}$	$6.28 \cdot 10^{-2}$	$9.30 \cdot 10^{-2}$	$1.36 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$2.06 \cdot 10^{-2}$	$5.47 \cdot 10^{-2}$	$9.57 \cdot 10^{-2}$	$1.46 \cdot 10^{-1}$
	$(\phi, \theta) = (0.5, 0.5)$	Direct	$4.51 \cdot 10^{-2}$	$6.62 \cdot 10^{-2}$	$9.87 \cdot 10^{-2}$	$1.42 \cdot 10^{-1}$
	(estimated)	Indirect	$3.06 \cdot 10^{-2}$	$5.91 \cdot 10^{-2}$	$1.02 \cdot 10^{-1}$	$1.57 \cdot 10^{-1}$
	$(\phi, \theta) = (0.5, 0.5)$	Direct	$4.17 \cdot 10^{-2}$	$6.28 \cdot 10^{-2}$	$9.55 \cdot 10^{-2}$	$1.35 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.96 \cdot 10^{-2}$	$5.53 \cdot 10^{-2}$	$9.72 \cdot 10^{-2}$	$1.47 \cdot 10^{-1}$
GARCH	$(\alpha, \beta) = (0.1, 0.1)$	Direct	$4.42 \cdot 10^{-2}$	$6.03 \cdot 10^{-2}$	$9.01 \cdot 10^{-2}$	$1.31 \cdot 10^{-1}$
	(estimated)	Indirect	$1.92 \cdot 10^{-2}$	$5.22 \cdot 10^{-2}$	$9.42 \cdot 10^{-2}$	$1.39 \cdot 10^{-1}$
	$(\alpha, \beta) = (0.1, 0.1)$	Direct	$4.44 \cdot 10^{-2}$	$6.03 \cdot 10^{-2}$	$9.34 \cdot 10^{-2}$	$1.35 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.88 \cdot 10^{-2}$	$5.23 \cdot 10^{-2}$	$9.49 \cdot 10^{-2}$	$1.45 \cdot 10^{-1}$
	$(\alpha, \beta) = (0.1, 0.45)$	Direct	$4.44 \cdot 10^{-2}$	$5.99 \cdot 10^{-2}$	$9.00 \cdot 10^{-2}$	$1.25 \cdot 10^{-1}$
	(estimated)	Indirect	$1.87 \cdot 10^{-2}$	$5.15 \cdot 10^{-2}$	$9.06 \cdot 10^{-2}$	$1.33 \cdot 10^{-1}$
	$(\alpha, \beta) = (0.1, 0.45)$	Direct	$4.44 \cdot 10^{-2}$	$6.03 \cdot 10^{-2}$	$9.34 \cdot 10^{-2}$	$1.35 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.88 \cdot 10^{-2}$	$5.23 \cdot 10^{-2}$	$9.49 \cdot 10^{-2}$	$1.45 \cdot 10^{-1}$
	$(\alpha, \beta) = (0.45, 0.1)$	Direct	$4.51 \cdot 10^{-2}$	$6.03 \cdot 10^{-2}$	$9.30 \cdot 10^{-2}$	$1.31 \cdot 10^{-1}$
	(estimated)	Indirect	$1.92 \cdot 10^{-2}$	$5.29 \cdot 10^{-2}$	$9.64 \cdot 10^{-2}$	$1.39 \cdot 10^{-1}$
	$(\alpha, \beta) = (0.45, 0.1)$	Direct	$4.44 \cdot 10^{-2}$	$6.03 \cdot 10^{-2}$	$9.34 \cdot 10^{-2}$	$1.35 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.88 \cdot 10^{-2}$	$5.23 \cdot 10^{-2}$	$9.49 \cdot 10^{-2}$	$1.45 \cdot 10^{-1}$
	$(\alpha, \beta) = (0.1, 0.85)$	Direct	$4.50 \cdot 10^{-2}$	$7.31 \cdot 10^{-2}$	$9.64 \cdot 10^{-2}$	$1.20 \cdot 10^{-1}$
	(estimated)	Indirect	$2.65 \cdot 10^{-2}$	$6.68 \cdot 10^{-2}$	$9.57 \cdot 10^{-2}$	$1.14 \cdot 10^{-1}$
	$(\alpha, \beta) = (0.1, 0.85)$	Direct	$4.44 \cdot 10^{-2}$	$6.03 \cdot 10^{-2}$	$9.34 \cdot 10^{-2}$	$1.35 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.88 \cdot 10^{-2}$	$5.23 \cdot 10^{-2}$	$9.49 \cdot 10^{-2}$	$1.45 \cdot 10^{-1}$

TABLE F.3

RMAD of the (bias-reduced) direct and indirect extreme conditional expectile estimators in ARMA and GARCH models. Estimators based on the fixed intermediate level $k_n = n/10 = 100$.

In line with standard practice in extreme value analysis for heavy tails, we consider instead the equivalent version

$$\frac{\sqrt{n(1-\tau_n)}}{\log[(1-\tau_n)/(1-\tau_n')]} \log\left(\frac{\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(Y|\boldsymbol{x})}{\xi_{\tau_n'}(Y|\boldsymbol{x})}\right) \xrightarrow{d} \mathcal{N}(0,\gamma^2)$$

obtained via the delta-method, as this has been observed several times to yield more reasonable confidence intervals when using Weissman-type extrapolated estimators (see *e.g.* [8] in the context of extreme quantile estimation). This immediately provides an asymptotic pointwise 95% confidence interval for $\xi_{\tau_n'}(Y|\boldsymbol{x})$ as

$$\widehat{I}^{(1)}_{\tau_n'}(\boldsymbol{x}) = \left[\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(Y|\boldsymbol{x}) \exp\left(\pm 1.96 \frac{\log[(1-\tau_n)/(1-\tau_n')]}{\sqrt{n(1-\tau_n)}} \widehat{\gamma}^{\mathrm{RB}}_{\lfloor n(1-\tau_n)\rfloor}\right)\right].$$

A slightly different construction, also motivated by Theorem 2.3, is possible by building the confidence interval directly on the estimator $\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)$ first and combining with location and

Model	Parameters	Estimator	$\gamma = 0.1$	$\gamma = 0.2$	$\gamma = 0.3$	$\gamma = 0.4$
ARMA	$(\phi,\theta) = (0.1, 0.1)$	Direct	$4.91 \cdot 10^{-2}$	$6.55 \cdot 10^{-2}$	$9.72 \cdot 10^{-2}$	$1.42 \cdot 10^{-1}$
	(estimated)	Indirect	$3.05 \cdot 10^{-2}$	$5.61 \cdot 10^{-2}$	$9.79 \cdot 10^{-2}$	$1.51 \cdot 10^{-1}$
	$(\phi,\theta) = (0.1, 0.1)$	Direct	$4.74 \cdot 10^{-2}$	$6.24 \cdot 10^{-2}$	$9.70 \cdot 10^{-2}$	$1.38 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.93 \cdot 10^{-2}$	$5.47 \cdot 10^{-2}$	$9.70 \cdot 10^{-2}$	$1.48 \cdot 10^{-1}$
	$(\phi,\theta) = (0.1, 0.5)$	Direct	$5.07 \cdot 10^{-2}$	$6.64 \cdot 10^{-2}$	$9.51 \cdot 10^{-2}$	$1.38 \cdot 10^{-1}$
	(estimated)	Indirect	$3.17 \cdot 10^{-2}$	$5.74 \cdot 10^{-2}$	$9.80 \cdot 10^{-2}$	$1.48 \cdot 10^{-1}$
	$(\phi,\theta) = (0.1, 0.5)$	Direct	$4.89 \cdot 10^{-2}$	$6.38 \cdot 10^{-2}$	$9.81 \cdot 10^{-2}$	$1.38 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$2.04 \cdot 10^{-2}$	$5.52 \cdot 10^{-2}$	$9.71 \cdot 10^{-2}$	$1.48 \cdot 10^{-1}$
	$(\phi,\theta) = (0.5, 0.1)$	Direct	$5.00 \cdot 10^{-2}$	$6.88 \cdot 10^{-2}$	$9.94 \cdot 10^{-2}$	$1.44 \cdot 10^{-1}$
	(estimated)	Indirect	$3.13 \cdot 10^{-2}$	$5.91 \cdot 10^{-2}$	$1.02 \cdot 10^{-1}$	$1.54 \cdot 10^{-1}$
	$(\phi,\theta) = (0.5, 0.1)$	Direct	$4.93 \cdot 10^{-2}$	$6.47 \cdot 10^{-2}$	$9.79 \cdot 10^{-2}$	$1.38 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$2.13 \cdot 10^{-2}$	$5.55 \cdot 10^{-2}$	$9.80 \cdot 10^{-2}$	$1.50 \cdot 10^{-1}$
	$(\phi,\theta) = (0.5, 0.5)$	Direct	$4.70 \cdot 10^{-2}$	$7.36 \cdot 10^{-2}$	$1.01 \cdot 10^{-1}$	$1.43 \cdot 10^{-1}$
	(estimated)	Indirect	$3.09 \cdot 10^{-2}$	$6.29 \cdot 10^{-2}$	$1.04 \cdot 10^{-1}$	$1.56 \cdot 10^{-1}$
	$(\phi,\theta) = (0.5, 0.5)$	Direct	$4.85 \cdot 10^{-2}$	$6.74 \cdot 10^{-2}$	$1.01 \cdot 10^{-1}$	$1.42 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$2.00 \cdot 10^{-2}$	$5.77 \cdot 10^{-2}$	$1.02 \cdot 10^{-1}$	$1.52 \cdot 10^{-1}$
GARCH	$(\alpha,\beta) = (0.1, 0.1)$	Direct	$4.65 \cdot 10^{-2}$	$6.22 \cdot 10^{-2}$	$9.22 \cdot 10^{-2}$	$1.34 \cdot 10^{-1}$
	(estimated)	Indirect	$1.90 \cdot 10^{-2}$	$5.45 \cdot 10^{-2}$	$9.23 \cdot 10^{-2}$	$1.39 \cdot 10^{-1}$
	$(\alpha,\beta) = (0.1, 0.1)$	Direct	$4.61 \cdot 10^{-2}$	$6.19 \cdot 10^{-2}$	$9.55 \cdot 10^{-2}$	$1.37 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.90 \cdot 10^{-2}$	$5.16 \cdot 10^{-2}$	$9.56 \cdot 10^{-2}$	$1.48 \cdot 10^{-1}$
	$(\alpha,\beta) = (0.1, 0.45)$	Direct	$4.72 \cdot 10^{-2}$	$6.29 \cdot 10^{-2}$	$9.09 \cdot 10^{-2}$	$1.28 \cdot 10^{-1}$
	(estimated)	Indirect	$1.87 \cdot 10^{-2}$	$5.33 \cdot 10^{-2}$	$9.23 \cdot 10^{-2}$	$1.35 \cdot 10^{-1}$
	$(\alpha,\beta) = (0.1, 0.45)$	Direct	$4.61 \cdot 10^{-2}$	$6.19 \cdot 10^{-2}$	$9.55 \cdot 10^{-2}$	$1.37 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.90 \cdot 10^{-2}$	$5.16 \cdot 10^{-2}$	$9.56 \cdot 10^{-2}$	$1.48 \cdot 10^{-1}$
	$(\alpha,\beta) = (0.45, 0.1)$	Direct	$4.71 \cdot 10^{-2}$	$6.30 \cdot 10^{-2}$	$9.80 \cdot 10^{-2}$	$1.35 \cdot 10^{-1}$
	(estimated)	Indirect	$1.93 \cdot 10^{-2}$	$5.50 \cdot 10^{-2}$	$9.86 \cdot 10^{-2}$	$1.41 \cdot 10^{-1}$
	$(\alpha,\beta) = (0.45, 0.1)$	Direct	$4.61 \cdot 10^{-2}$	$6.19 \cdot 10^{-2}$	$9.55 \cdot 10^{-2}$	$1.37 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.90 \cdot 10^{-2}$	$5.16 \cdot 10^{-2}$	$9.56 \cdot 10^{-2}$	$1.48 \cdot 10^{-1}$
	$(\alpha,\beta) = (0.1, 0.85)$	Direct	$4.55 \cdot 10^{-2}$	$7.40 \cdot 10^{-2}$	$9.71 \cdot 10^{-2}$	$1.22 \cdot 10^{-1}$
	(estimated)	Indirect	$2.67 \cdot 10^{-2}$	$6.80 \cdot 10^{-2}$	$9.31 \cdot 10^{-2}$	$1.14 \cdot 10^{-1}$
	$(\alpha,\beta) = (0.1, 0.85)$	Direct	$4.61 \cdot 10^{-2}$	$6.19 \cdot 10^{-2}$	$9.55 \cdot 10^{-2}$	$1.37 \cdot 10^{-1}$
	(known, benchmark)	Indirect	$1.90 \cdot 10^{-2}$	$5.16 \cdot 10^{-2}$	$9.56 \cdot 10^{-2}$	$1.48 \cdot 10^{-1}$

TABLE F.4

RMAD of the (bias-reduced) direct and indirect extreme conditional expectile estimators in ARMA and GARCH models. Estimators based on the data-driven intermediate level $\widehat{k}_n^*$.

scale afterwards. In this case, an asymptotic pointwise 95% confidence interval for $\xi_{\tau_n'}(\varepsilon)$ is

$$\left[\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)\exp\left(\pm 1.96\frac{\log[(1-\tau_n)/(1-\tau_n')]}{\sqrt{n(1-\tau_n)}}\widehat{\gamma}^{\mathrm{RB}}_{\lfloor n(1-\tau_n)\rfloor}\right)\right].$$

In the class of regression models (1) where $\xi_{\tau_n'}(Y|\boldsymbol{x}) = g(\boldsymbol{x}) + \sigma(\boldsymbol{x})\xi_{\tau_n'}(\varepsilon)$, this yields an alternative asymptotic pointwise 95% confidence interval for $\xi_{\tau_n'}(Y|\boldsymbol{x})$ as

$$\widehat{I}^{(2)}_{\tau_n'}(\boldsymbol{x}) = \left[\overline{g}(\boldsymbol{x}) + \overline{\sigma}(\boldsymbol{x})\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)\exp\left(\pm 1.96\frac{\log[(1-\tau_n)/(1-\tau_n')]}{\sqrt{n(1-\tau_n)}}\widehat{\gamma}^{\mathrm{RB}}_{\lfloor n(1-\tau_n)\rfloor}\right)\right]$$

if g and σ are estimated by $\overline{g}$ and $\overline{\sigma}$ sufficiently fast that the asymptotic behaviour of $\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)$ dominates. In a model where the conditional mean is assumed to be 0 (for example GARCH models), the intervals $\widehat{I}^{(1)}_{\tau_n'}$ and $\widehat{I}^{(2)}_{\tau_n'}$ coincide. We illustrate the behaviour of $\widehat{I}^{(1)}_{\tau_n'}(\boldsymbol{x})$ (calculated on the bias-reduced direct estimator) in the top left panel of Figure F.1 below, on the example of the Vehicle Insurance Customer data of Section 4.3.

Finite-sample coverages of these two intervals at the 95% nominal level are compared in the setups of Section 4.1 (see Table F.5) and Section 4.2 (see Table F.6) for an extreme value index equal to $1/4 = 0.25$. Interval $\widehat{I}^{(1)}_{\tau'_n}$ yields sensible results at a central point $\boldsymbol{x}$ in regression models, as can be seen from the leftmost table in Table F.5. Interval $\widehat{I}^{(2)}_{\tau'_n}$ has a lower coverage probability and seems to be too narrow. It is interesting to note that the difference between the performance of intervals constructed using estimated model parameters (ignoring the uncertainty incurred at the model estimation step) and of those obtained with the unrealistic knowledge of model structure is negligible; in the regression case, this can be seen by comparing procedures (S1) and (S1i) with benchmarks (B5) and (B5i) in the linear model (G1), and (S2) and (S2i) with benchmarks (B5) and (B5i) in the single-index model (G2). This illustrates once again that the extreme value step, rather than model estimation, is indeed the major contributor to estimation uncertainty as long as the model can be estimated efficiently. We illustrate this point further in our time series models, where it can be seen that for both intervals, the coverage probabilities obtained by assuming knowledge of the model are essentially identical to those where the model structure has to be estimated. In our time series examples, coverage of the Gaussian confidence intervals is in fact arguably quite poor (around 80% in most models), but this will be due to the fact that the sample size is not yet large enough for the Gaussian approximation to be reasonable for sample expectiles. This is not due to the uncertainty in model estimation not being accounted for, since assuming knowledge of the model does not improve coverage substantially. Issues with finite-sample coverage of Gaussian confidence intervals for the estimation of extreme conditional risk measures such as the Expected Shortfall (closely related to the expectile) have been reported before, see *e.g.* [14].

Model	Procedure	$\widehat{I}^{(1)}_{\tau'_n}$	$\widehat{I}^{(2)}_{\tau'_n}$	Model	Procedure	$\widehat{I}^{(1)}_{\tau'_n}$	$\widehat{I}^{(2)}_{\tau'_n}$
Linear (G1)	(S1)	0.910	0.746	Linear (G1)	(S1)	0.740	0.468
	(S1i)	0.924	0.758		(S1i)	0.740	0.458
	(S2)	0.924	0.764		(S2)	0.236	0.114
	(S2i)	0.942	0.780		(S2i)	0.230	0.120
	(B2)	0.816	0.484		(B2)	0.000	0.000
	(B3)	0.908	0.720		(B3)	0.343	0.154
	(B4)	0.914	0.760		(B4)	0.932	0.760
	(B4i)	0.980	0.840		(B4i)	0.988	0.840
	(B5)	0.932	0.774		(B5)	0.944	0.774
	(B5i)	0.944	0.784		(B5i)	0.962	0.784
Single index (G2)	(S1)	0.844	0.590	Single index (G2)	(S1)	0.034	0.026
	(S1i)	0.862	0.646		(S1i)	0.034	0.024
	(S2)	0.920	0.802		(S2)	0.590	0.442
	(S2i)	0.932	0.836		(S2i)	0.596	0.452
	(B2)	0.158	0.060		(B2)	0.060	0.081
	(B3)	0.858	0.750		(B3)	0.242	0.152
	(B4)	0.872	0.760		(B4)	0.868	0.760
	(B4i)	0.962	0.840		(B4i)	0.952	0.840
	(B5)	0.896	0.774		(B5)	0.888	0.774
	(B5i)	0.920	0.784		(B5i)	0.908	0.784

TABLE F.5

Empirical coverage probabilities of the Gaussian asymptotic confidence intervals (95% nominal level) associated with methods (S1), (S2), (S1i) and (S2i), and benchmarks (B2)–(B5i), in models (G1)–(G2). Estimators based on the fixed intermediate level $k_n = n/10 = 100$, left table: central point $\boldsymbol{x} = (1/2, 1/2, 1/2, 1/3)$, right table: noncentral point $\boldsymbol{x} = (0.1, 0.1, 0.1, 0.1)$. The extreme value index γ is set to the value $1/4 = 0.25$. Benchmarks (B1) and (B1i) are not location-scale approaches and therefore have been excluded from this comparative table.

Model	Parameters	Estimator	$\widehat{I}^{(1)}_{\tau_n'}$	$\widehat{I}^{(2)}_{\tau_n'}$
ARMA	$(\phi,\theta)=(0.1,0.1)$	Direct	0.769	0.776
	(estimated)	Indirect	0.785	0.794
	$(\phi,\theta)=(0.1,0.1)$	Direct	0.806	0.804
	(known, benchmark)	Indirect	0.824	0.822
	$(\phi,\theta)=(0.1,0.5)$	Direct	0.766	0.787
	(estimated)	Indirect	0.779	0.791
	$(\phi,\theta)=(0.1,0.5)$	Direct	0.773	0.804
	(known, benchmark)	Indirect	0.792	0.822
	$(\phi,\theta)=(0.5,0.1)$	Direct	0.756	0.776
	(estimated)	Indirect	0.764	0.794
	$(\phi,\theta)=(0.5,0.1)$	Direct	0.759	0.804
	(known, benchmark)	Indirect	0.783	0.822
	$(\phi,\theta)=(0.5,0.5)$	Direct	0.698	0.783
	(estimated)	Indirect	0.707	0.795
	$(\phi,\theta)=(0.5,0.5)$	Direct	0.697	0.804
	(known, benchmark)	Indirect	0.709	0.822
GARCH	$(\alpha,\beta)=(0.1,0.1)$	Direct	0.800	0.800
	(estimated)	Indirect	0.817	0.817
	$(\alpha,\beta)=(0.1,0.1)$	Direct	0.804	0.804
	(known, benchmark)	Indirect	0.815	0.815
	$(\alpha,\beta)=(0.1,0.45)$	Direct	0.793	0.793
	(estimated)	Indirect	0.806	0.806
	$(\alpha,\beta)=(0.1,0.45)$	Direct	0.795	0.795
	(known, benchmark)	Indirect	0.818	0.818
	$(\alpha,\beta)=(0.45,0.1)$	Direct	0.793	0.793
	(estimated)	Indirect	0.802	0.802
	$(\alpha,\beta)=(0.45,0.1)$	Direct	0.784	0.784
	(known, benchmark)	Indirect	0.803	0.803
	$(\alpha,\beta)=(0.1,0.85)$	Direct	0.710	0.710
	(estimated)	Indirect	0.732	0.732
	$(\alpha,\beta)=(0.1,0.85)$	Direct	0.686	0.686
	(known, benchmark)	Indirect	0.717	0.717

TABLE F.6

Empirical coverage probabilities of the Gaussian asymptotic confidence intervals (95% nominal level) associated with the (bias-reduced) direct and indirect one-step ahead extreme expectile estimators in ARMA and GARCH models. Estimators based on the fixed intermediate level $k_n = n/10 = 100$. The extreme value index γ is set to the value $1/4 = 0.25$.

Situations where trusting these Gaussian confidence intervals might be difficult include regression models featuring the estimation of a nonparametric component (such as the heteroscedastic single-index model in Section 3.2, used for the analysis of the Vehicle Insurance Customer data) whose rate of convergence may be close to the rate of convergence of the extreme value estimator. In such models, disregarding the uncertainty incurred at the model estimation stage may be problematic in regions where data is relatively sparse. This is illustrated in the rightmost table of Table F.5, where it can be seen that a noncentral point $\boldsymbol{x}$ of the regression problem, coverage of the proposed Gaussian asymptotic confidence intervals dramatically decreases, especially in the heteroscedastic single-index model. It may then be more prudent to move away from the asymptotic approximation and use instead an approach that fully takes into account the uncertainty in the estimation. We propose and contrast here a couple of alternatives based on regression bootstrap methods. We develop our ideas in the example of the heteroscedastic single-index model of Section 3.2. Suppose that from a data set $(\boldsymbol{X}_i, Y_i)_{1\le i\le n}$, we have estimated a direction vector $\widehat{\boldsymbol{\beta}}$ along with mean and standard deviation functions $\widehat{g}$ and $\widehat{\sigma}$. One possibility to describe the uncertainty in the estimation of $\xi_{\tau_n'}(Y|\boldsymbol{x})$ is to use the wild bootstrap, widespread in the heteroscedastic regression literature and whose origins can be traced back to [19]. This consists in resampling $(\boldsymbol{X}_i, Y_i^*)_{1\le i\le n}$ as follows:

$$Y_i^* = \widehat{g}(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i) + (Y_i - \widehat{g}(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i))\varepsilon_i^*,$$

where $(\varepsilon_i^*)_{1\le i\le n}$ are i.i.d. copies of a random variable ε^* having mean 0 and variance 1. A natural, possible choice for ε^* is the standard normal distribution. We illustrate this methodology on the example of the Vehicle Insurance Customer data of Section 4.3. We simulated $N = 5{,}000$ such bootstrap samples $(\boldsymbol{X}_i, Y_i^*)_{1\le i\le n}$; in each sample, we kept the direction vector $\widehat{\boldsymbol{\beta}}$ fixed and equal to its estimated value based on the original sample, and we estimated the functions g and σ using the same method as in the real data analysis in Section 4.3. This is sensible because the estimator $\widehat{\boldsymbol{\beta}}$ converges much faster than the nonparametric estimators of g and σ, and therefore keeping the direction fixed is very unlikely to be incorrect as far as uncertainty quantification is concerned. Using residuals and the direct, bias-reduced extreme conditional expectile estimator results in an estimate of $\xi_{\tau_n'}(Y|\boldsymbol{x})$ which, for the jth bootstrap sample, we denote by $\widehat{\xi}^{\star,\mathrm{RB},(j)}_{\tau_n'}(Y|\boldsymbol{x})$. We finally build, for a fixed $\boldsymbol{x}$, pointwise 95% bootstrap confidence intervals calculated by taking the empirical quantiles at levels 2.5% and 97.5% of the $\widehat{\xi}^{\star,\mathrm{RB},(j)}_{\tau_n'}(Y|\boldsymbol{x})$, $1\le j\le N$. These are reported in the top right panel of Figure F.1. At extreme levels (say here $\tau_n' = 1 - 1/(nh^*)$, with $h^* = 0.1$) the confidence intervals look reasonable on the right half of the graph. However, they seem to very substantially overestimate the uncertainty in the left half, where data is sparser; this is especially clear around $\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x} = -0.2$, where the estimated extreme conditional expectile curve already extrapolates far beyond the observations locally relevant, which suggests that the upper bound of the associated confidence interval should be relatively close to the point estimate, but this is not the case. Moreover, the wild bootstrap method appears to be very sensitive to the choice of distribution of ε^* (alternative choices include the Rademacher distribution or asymmetric two-point distributions such as the one on p.257 of [16]). Our interpretation is that the wild bootstrap is too conservative here because it fails to get a good idea of the right tail behaviour in the data.

To remedy this problem we suggest a second, semiparametric bootstrap method. This time, the Y_i^*, $1\le i\le n$, are simulated as

$$Y_i^* = \widehat{g}(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i) + \widehat{\sigma}(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{X}_i)\varepsilon_i^*,$$

where the ε_i^* are obtained by

1. Simulating u_i from the standard uniform distribution on $[0,1]$,
2. If $u_i \in [p, 1-p]$, for a fixed $p \in (0,1)$, taking $\varepsilon_i^* = \widehat{F}^{-1}(u_i)$, where $\widehat{F}$ is the empirical distribution function of the residuals $\widehat{\varepsilon}_i$,
3. If $u_i > 1-p$, taking $\varepsilon_i^* = ((1-u_i)/p)^{-\widehat{\gamma}}\widehat{F}^{-1}(1-p)$, where $\widehat{\gamma} = \widehat{\gamma}^{\mathrm{RB}}$ is the bias-reduced Hill estimator (with $k_n = 200$ as in Section 4.3) based on the residuals $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n$,
4. If $u_i < p$, taking $\varepsilon_i^* = (u_i/p)^{-\widehat{\gamma}_\ell}\widehat{F}^{-1}(p)$, where $\widehat{\gamma}_\ell = \widehat{\gamma}_\ell^{\mathrm{RB}}$ is the bias-reduced Hill estimator (with $k_n = 200$) based on the negative residuals $-\widehat{\varepsilon}_1, \ldots, -\widehat{\varepsilon}_n$.

We chose $p = 0.001$; further investigations, which we do not report here, suggest that results are not too sensitive to the choice of p as long as $p \in [0.001, 0.01]$. The idea of steps 3 and 4 above is to allow the resampling algorithm to give a faithful idea of the right and left tails of the data through the use of the Pareto approximations of these tails. We call this algorithm the *semiparametric Pareto tail bootstrap*. Somewhat similar ideas have appeared before in the literature, see *e.g.* [20] whose aim was to approximate the distribution of extreme order statistics.

We illustrate this methodology again on the example of the Vehicle Insurance Customer data of Section 4.3. We simulate $N = 5{,}000$ bootstrap samples $(\boldsymbol{X}_i, Y_i^*)_{1\le i\le n}$ and, like previously, we keep the direction vector $\widehat{\boldsymbol{\beta}}$ fixed and estimate the functions g and σ using the same method as in Section 4.3. This yields extrapolated direct bias-reduced estimates of $\xi_{\tau_n'}(Y|\boldsymbol{x})$ in each sample and therefore pointwise 95% bootstrap confidence intervals calculated by taking the empirical quantiles at levels 2.5% and 97.5% of these estimates. These intervals are reported in the bottom left panel of Figure F.1; all three intervals are compared to each other on the bottom right panel of this Figure. All intervals are roughly similar on the right part of the graph, but on the left part where data is more sparse, the semiparametric Pareto tail bootstrap intervals appear to give a much better idea of the type of tail the data exhibits. In practice, we therefore recommend reporting the Gaussian confidence intervals along with the semiparametric Pareto tail bootstrap confidence intervals, since the latter may give a more accurate picture of uncertainty where data is sparser. This is the approach we adopt in the real data analyses of Sections 4.3 and 4.4.

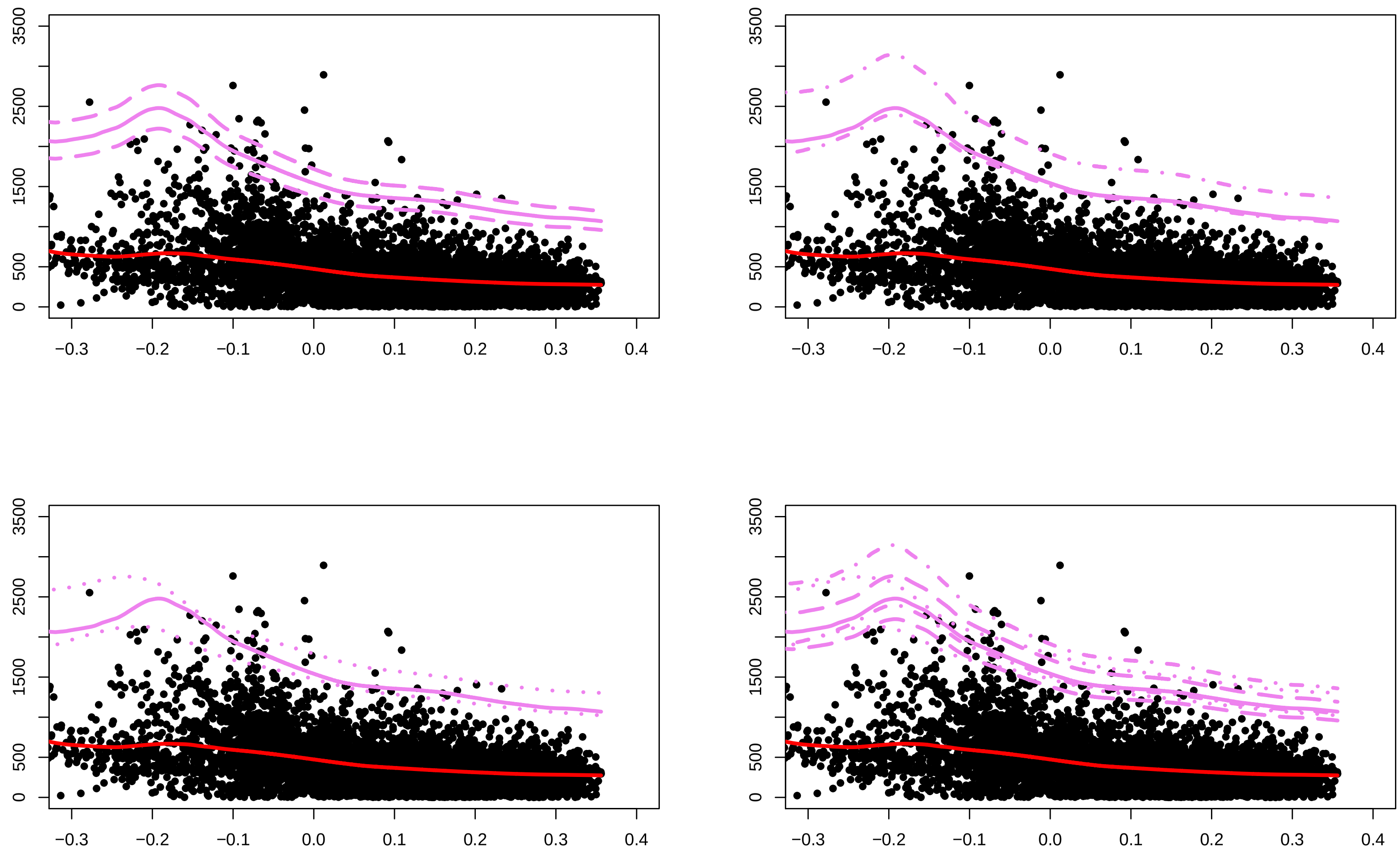

FIG F.1. *Vehicle Insurance Customer data, pointwise 95% confidence intervals. Top left panel: asymptotic Gaussian confidence intervals, top right: wild bootstrap confidence intervals, bottom left: semiparametric Pareto tail bootstrap confidence intervals, bottom right: all three intervals. In all panels, the red line is the regression mean, the solid purple line is the (direct, bias-reduced) extreme conditional expectile estimate at level $\tau_n' = 1 - 1/(nh^*)$ and the dashed, dashed-dotted and dotted lines represent the 95% confidence intervals, in the $(\widehat{\boldsymbol{\beta}}^\top \boldsymbol{x}, y)$ plane.*

A comprehensive analysis of the finite-sample coverage of the proposed semiparametric Pareto tail bootstrap confidence interval is unfortunately not yet feasible in a reasonable amount of time because the calculation of these intervals is computationally very expensive: a rough estimation of the amount of time needed to compute the Pareto tail bootstrap confidence interval in a sample of size $n = 1{,}000$ (from any one of the models we examine in the simulation study) leads to one hour of computational time. Multiplied by the number of replications ($N = 1{,}000$ independent samples in each model), the number of methods and the number of models we consider, a full study in the spirit of Sections 4.1 and 4.2 would require at least several months of calculation even if the code were parallelised. To get an idea of how the proposed bootstrap methodology performs in practice, we suggest the following small simulation experiment inspired from the kind of general model we consider in this paper. Consider a sample of (location-scale) random variables $Y_1, \ldots, Y_n$ defined through

$$Y_i = m + \sigma\varepsilon_i.$$

Here the mean parameter is $m = 2$, the standard deviation parameter is $\sigma = 1$, and the random variables $\varepsilon_1, \ldots, \varepsilon_n$ are $n = 1{,}000$ independent and identically distributed realisations of a symmetric rescaled Burr distribution as in Section 4.2, with $\gamma = 0.25$ and $\rho = -1$. The goal is to infer an extreme expectile of level $\tau_n' = 1 - 5/n = 0.995$ of Y by filtering first the mean and scale components. This very closely resembles the approach adopted throughout the paper in location-scale heteroscedastic regression models. The following estimation methods are compared:

(E1) We estimate first m and σ by the empirical mean $\overline{m}$ and standard deviation $\overline{\sigma}$. We then construct the residuals $\widehat{\varepsilon}_i = (Y_i - \overline{m})/\overline{\sigma}$ and estimate $\xi_{\tau_n'}(\varepsilon)$ using the bias-reduced direct and indirect estimators $\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)$ and $\widetilde{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)$ calculated on the $\widehat{\varepsilon}_i$ with $\tau_n = 1 - 100/n = 0.9$. We finally deduce the two extreme expectile estimators $\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(Y) = \overline{m} + \overline{\sigma}\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)$ and $\widetilde{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(Y) = \overline{m} + \overline{\sigma}\widetilde{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)$.

(E2) Same as in (E1), with $\overline{m}$ and $\overline{\sigma}$ calculated using only the first $n/2$ observations.

(E3) Same as in (E1), with $\overline{m}$ and $\overline{\sigma}$ calculated using only the first $n/4$ observations.

(E4) Same as in (E1), with $\overline{m}$ and $\overline{\sigma}$ calculated using only the first $n/10$ observations.

This is compared to the unrealistic benchmark (BE) where m and σ are assumed to be known and thus the true ε_i are accessible. Note that, following the discussion at the top of p.83 in [6], this benchmark should be seen as enjoying a strong advantage over (E1)–(E4), since the shifted variables Y_i have a second-order parameter $-\gamma = -1/4$, which is much closer to 0 than the original second-order parameter $\rho = -1$ of the ε_i. The latter are, strictly speaking, only accessible in the framework of this unrealistic benchmark (BE). The point of considering the estimation of the mean and scale components using progressively lower sample sizes is to assess the influence of the rate of estimation of location-scale model components; in (E4), there are only 100 variables used to calculate $\overline{m}$ and $\overline{\sigma}$, meaning that the "rate of convergence" of $\overline{m}$ and $\overline{\sigma}$ is $\sqrt{100} = 10$, exactly equal to $\sqrt{n(1-\tau_n)}$ which is the rate of convergence of the extreme value step.

For each method, we compare three confidence intervals. These are, first of all, the two Gaussian asymptotic 95% confidence intervals

$$\widehat{I}^{(1)}_{\tau_n'} = \left[\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(Y)\exp\left(\pm 1.96\frac{\log[(1-\tau_n)/(1-\tau_n')]}{\sqrt{n(1-\tau_n)}}\widehat{\gamma}^{\mathrm{RB}}_{\lfloor n(1-\tau_n)\rfloor}\right)\right]$$

and

$$\widehat{I}^{(2)}_{\tau_n'} = \left[\overline{m} + \overline{\sigma}\widehat{\xi}^{\star,\mathrm{RB}}_{\tau_n'}(\varepsilon)\exp\left(\pm 1.96\frac{\log[(1-\tau_n)/(1-\tau_n')]}{\sqrt{n(1-\tau_n)}}\widehat{\gamma}^{\mathrm{RB}}_{\lfloor n(1-\tau_n)\rfloor}\right)\right].$$

Approach	Expectile estimator	$\widehat{I}^{(1)}_{\tau'_n}$	$\widehat{I}^{(2)}_{\tau'_n}$	$\widehat{I}^{(\mathrm{boot})}_{\tau'_n}$
Benchmark (BE)	Bias-reduced direct	0.994 (1.082)	0.796 (0.516)	0.896 (0.739)
	Bias-reduced indirect	0.998 (1.080)	0.798 (0.514)	0.880 (0.679)
Method (E1)	Bias-reduced direct	0.992 (1.083)	0.790 (0.517)	0.898 (0.742)
	Bias-reduced indirect	0.998 (1.081)	0.804 (0.515)	0.882 (0.681)
Method (E2)	Bias-reduced direct	0.992 (1.082)	0.792 (0.517)	0.900 (0.739)
	Bias-reduced indirect	0.998 (1.080)	0.798 (0.515)	0.880 (0.680)
Method (E3)	Bias-reduced direct	0.996 (1.081)	0.792 (0.516)	0.900 (0.741)
	Bias-reduced indirect	1.000 (1.079)	0.800 (0.514)	0.884 (0.682)
Method (E4)	Bias-reduced direct	0.994 (1.083)	0.790 (0.516)	0.896 (0.741)
	Bias-reduced indirect	0.994 (1.081)	0.800 (0.514)	0.884 (0.681)

TABLE F.7

Empirical coverage probabilities of the Gaussian asymptotic confidence intervals and semiparametric Pareto tail bootstrap confidence intervals (95% nominal level) associated with the (bias-reduced) direct and indirect extreme expectile estimators in the location-scale model $Y = m + \sigma\varepsilon$. Between brackets: associated average lengths of the confidence intervals.

We compare these intervals with the semiparametric Pareto tail bootstrap 95% confidence intervals generated as follows: we simulate $n_b = 500$ bootstrap samples $\varepsilon_1^*, \ldots, \varepsilon_n^*$ by

1. Simulating u_i from the standard uniform distribution on $[0,1]$,
2. If $u_i \in [p, 1-p]$, for $p = 0.001$, taking $\varepsilon_i^* = \widehat{F}^{-1}(u_i)$, where $\widehat{F}$ is the empirical distribution function of the residuals $\widehat{\varepsilon}_i$,
3. If $u_i > 1-p$, taking $\varepsilon_i^* = ((1-u_i)/p)^{-\widehat{\gamma}}\widehat{F}^{-1}(1-p)$, where $\widehat{\gamma} = \widehat{\gamma}^{\mathrm{RB}}$ is the bias-reduced Hill estimator (with $k = 200$) based on the residuals $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n$,
4. If $u_i < p$, taking $\varepsilon_i^* = (u_i/p)^{-\widehat{\gamma}_\ell}\widehat{F}^{-1}(p)$, where $\widehat{\gamma}_\ell = \widehat{\gamma}_\ell^{\mathrm{RB}}$ is the bias-reduced Hill estimator (with $k = 200$) based on the negative residuals $-\widehat{\varepsilon}_1, \ldots, -\widehat{\varepsilon}_n$.

We then deduce bootstrap samples $(Y_1^*, \ldots, Y_n^*) = (\overline{m} + \overline{\sigma}\varepsilon_1^*, \ldots, \overline{m} + \overline{\sigma}\varepsilon_n^*)$. For each sample, we estimate the extreme expectile at level τ'_n (the bias-reduced direct estimator is employed), and take the empirical 0.025 and 0.975 quantiles of the n_b estimates to construct our bootstrap confidence interval $\widehat{I}^{(\mathrm{boot})}_{\tau'_n}$. This is the exact analogue of the construction we proposed above, adapted to this simpler location-scale example. We also compare these intervals with their versions obtained using the bias-reduced indirect estimators. We record empirical coverage probabilities and average lengths of the intervals. Results are presented in Table F.7.

It is readily seen, first of all, that results are almost completely unaffected by the knowledge of the location-scale model structure, and similarly unaffected by the number of data points used for the estimation of the mean and scale parameters. It is also seen that the two Gaussian confidence intervals behave quite poorly, being either too conservative or too narrow and achieving a coverage rate far from the nominal rate. By contrast, the proposed semiparametric Pareto tail bootstrap confidence interval behaves fairly well, with a typical coverage probability of about 90%. This seems to be quite robust to the number of bootstrap replications: a larger number of bootstrap replications was also considered without changing results substantially. This constitutes reasonable grounds for recommending the use of the semiparametric Pareto tail bootstrap confidence interval, although of course a full-scale simulation study should be carried out in future work to assess its accuracy in the regression context (subject to computational improvements that are beyond the scope of this article).

REFERENCES

[1] ABDOUS, B. and REMILLARD, B. (1995). Relating quantiles and expectiles under weighted-symmetry. *Annals of the Institute of Statistical Mathematics* **47** 371–384.

[2] BREIMAN, L. (1965). On some limit theorems similar to the arc-sin law. *Theory of Probability and its Applications* **10** 323–331.

[3] BROCKWELL, P. J. and DAVIS, R. A. (1991). *Time Series: Theory and Methods (second edition)*. Springer.

[4] 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.

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

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

[7] DEKKERS, A. L. M., EINMAHL, J. H. J. and DE HAAN, L. (1989). A moment estimator for the index of an extreme-value distribution. *Annals of Statistics* **17** 1833–1855.

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

[9] FRANCQ, C. and ZAKOÏAN, J. M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley & Sons.

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

[11] 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.

[12] HANSEN, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24** 726–748.

[13] HJORT, N. L. and POLLARD, D. (1993). Asymptotics for minimisers of convex processes. Unpublished manuscript.

[14] HOGA, Y. (2019). Confidence intervals for conditional tail risk measures in ARMA–GARCH models. *Journal of Business & Economic Statistics* **37** 613–624.

[15] KNIGHT, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity. Unpublished manuscript.

[16] MAMMEN, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* **21** 255–285.

[17] MASRY, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis* **17** 571–599.

[18] STUPFLER, G. (2019). On a relationship between randomly and non-randomly thresholded empirical average excesses for heavy tails. *Extremes* **22** 749–769.

[19] WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Annals of Statistics* **14** 1261–1350.

[20] ZELTERMAN, D. (1993). A semiparametric bootstrap technique for simulating extreme order statistics. *Journal of the American Statistical Association* **88** 477–485.