Supplementary Material to *"Out-of-Sample Tests for Conditional Quantile Coverage - An Application to Growth-at-Risk"*

Valentina Corradi* Jack Fosten†‡ Daniel Gutknecht§

June 30, 2023

Abstract

This document is divided into eight sections. Section S.1 outlines the nonlinear location scale model. Specifically, Subsection S.1.1 provides the details and assumptions of the nonlinear location scale model, while Subsection S.1.2 outlines the estimation. Section S.2 contains two auxiliary Lemmas and the proofs of all results from the main paper. In particular, Subsection S.2.1 contains two Lemmas that establish an asymptotic linear representation for the quantile regression and the location scale model (see Lemma Q.1 and L.1), Subsection S.2.2 gathers the proofs of all Theorems from the paper, while Subsection S.2.3 contains the proofs of the auxiliary Lemma Q.1 and L.1. Sections S.3 and S.4 on the other hand outline the differences in the construction of the bootstrap statistic in the case of a two-sided interval $[\tau_L, \tau_U]$ and of nonlinear location scale models (together with differences in the proof of Theorem 2), respectively. Section S.5 contains an outline of the bootstrap statistic in the case of the recursive estimation scheme. Section S.6 outlines the possibility to accommodate predictions from Conditional Autoregressive Value-at-Risk (CAViaR) models in our test. Finally, Section S.7 displays the results of some additional Monte Carlo simulations, while Section S.8 introduces an additional Monte Carlo design for the case of equally mis-specified and overlapping models and Section S.9 provides an additional empirical application to VaR forecasting.

*School of Economics, University of Surrey, GU2 7XH Guildford, UK, E-mail: V.Corradi@surrey.ac.uk

†King's Business School, King's College London, WC2B 4BG London, UK, E-mail: jack.fosten@kcl.ac.uk

‡Data Analytics in Finance and Macro (DAFM) Research Centre, King's College London, UK

§Faculty of Economics and Business, Goethe University Frankfurt, 60629 Frankfurt am Main, Germany. E-mail: Gutknecht@wiwi.uni-frankfurt.de.

S.1 Nonlinear Location Scale Model

S.1.1 Model Set-up

Recall the nonlinear location scale model given by:

$$y_{t+1} = m\left(X_{j,t}, \boldsymbol{\theta}^{\dagger}_{j,m}\right) + \sigma(X_{j,t}, \boldsymbol{\theta}^{\dagger}_{j,\sigma})\epsilon_{j,t+1},$$

where $\epsilon_{j,t+1} = \left(y_{t+1} - m(X_{j,t}, \boldsymbol{\theta}^{\dagger}_{j,m})\right)/\sigma(X_{j,t}, \boldsymbol{\theta}^{\dagger}_{j,\sigma})$, and $m(\cdot, \boldsymbol{\theta}^{\dagger}_{j,m})$ as well as $\sigma(\cdot, \boldsymbol{\theta}^{\dagger}_{j,\sigma})$ are some nonlinear functions indexed by some finite dimensional parameter vectors $\boldsymbol{\theta}^{\dagger}_{j,m}$ and $\boldsymbol{\theta}^{\dagger}_{j,\sigma}$. In this case, we have that the conditional quantile function $q_\tau(\boldsymbol{\theta}^{\dagger}_j; X_{j,t})$ with $\boldsymbol{\theta}^{\dagger}_j = (\boldsymbol{\theta}^{\dagger\prime}_{j,m}, \boldsymbol{\theta}^{\dagger\prime}_{j,\sigma})'$ is given by:

$$q_\tau(\boldsymbol{\theta}^{\dagger}_j; X_{j,t}) = m\left(X_{j,t}, \boldsymbol{\theta}^{\dagger}_{j,m}\right) + \sigma(X_{j,t}, \boldsymbol{\theta}^{\dagger}_{j,\sigma})\beta^{\dagger}_{j,\epsilon}(\tau). \tag{S.1}$$

where $\beta^{\dagger}_{j,\epsilon}(\tau)$ is the τ quantile of $\epsilon_{j,t+1}$. The following condition is a high-level condition for the case where one or more nonlinear location scale model(s) are used in the comparison. Thus, let $\|\cdot\|$ denote the Euclidean norm and $\nabla^{(k)}_{\boldsymbol{\alpha}} g(\cdot;\boldsymbol{\alpha})$ denote the k-th order partial derivative of the function $g(\cdot;\boldsymbol{\alpha})$ with respect to the vector $\boldsymbol{\alpha}$.[1] Then, for every model $j \in \{1, \ldots, J\}$, which can be written as in (S.1), the following holds:

Assumption A.6:
(i) For every $j \in \{1, \ldots, J\}$, the estimators $\widehat{\boldsymbol{\theta}}_{j,m,R}$ and $\widehat{\boldsymbol{\theta}}_{j,\sigma,R}$ are $\sqrt{R}$-consistent for some unique population vectors $\boldsymbol{\theta}^{\dagger}_{j,m}$ and $\boldsymbol{\theta}^{\dagger}_{j,\sigma}$, which lay in the interior of the compact parameter spaces $\boldsymbol{\Theta}_m$ and $\boldsymbol{\Theta}_\sigma$, respectively.
(ii) Let $\nabla^{(1)}_{\boldsymbol{\theta}_l}\zeta(y_{s+1}, X_s, \boldsymbol{\theta}_{j,m}, \boldsymbol{\theta}_{j,\sigma})$ and $\nabla^{(2)}_{\boldsymbol{\theta}_l}\zeta(y_{t+1}, X_s, \boldsymbol{\theta}_{j,m}, \boldsymbol{\theta}_{j,\sigma})$ with $l \in \{m, \sigma\}$ and $j = 1, \ldots, J$ denote the first and second order partial derivatives of some objective function $\zeta(y_{s+1}, X_s, \boldsymbol{\theta}_{j,m}, \boldsymbol{\theta}_{j,\sigma})$ with respect to $\boldsymbol{\theta}_{j,l}$. For every $j \in \{1, \ldots, J\}$, both estimators satisfy the following asymptotic linear representation:

$$\sqrt{R}\left(\widehat{\boldsymbol{\theta}}_{j,m,R} - \boldsymbol{\theta}^{\dagger}_{j,m}\right) = \mathbf{M}^{-1}_{j,m}\frac{1}{\sqrt{R}}\sum_{s=1}^{R-1}\nabla^{(1)}_{\boldsymbol{\theta}_m}\zeta(y_{s+1}, X_{j,s}, \boldsymbol{\theta}^{\dagger}_{j,m}, \boldsymbol{\theta}^{\dagger}_{j,\sigma}) + o_p(1)$$

and:

$$\sqrt{R}\left(\widehat{\boldsymbol{\theta}}_{j,\sigma,R} - \boldsymbol{\theta}^{\dagger}_{j,\sigma}\right) = \mathbf{M}^{-1}_{j,\sigma}\frac{1}{\sqrt{R}}\sum_{s=1}^{R-1}\nabla^{(1)}_{\boldsymbol{\theta}_\sigma}\zeta(y_{s+1}, X_{j,s}, \boldsymbol{\theta}^{\dagger}_{j,m}, \boldsymbol{\theta}^{\dagger}_{j,\sigma}) + o_p(1)$$

where $\mathbf{M}_{j,m} = \mathrm{E}\left(\nabla^{(2)}_{\boldsymbol{\theta}_m}\zeta(y_{t+1}, X_{j,t}, \boldsymbol{\theta}^{\dagger}_{j,m}, \boldsymbol{\theta}^{\dagger}_{j,\sigma})\right)$ and $\mathbf{M}_{j,\sigma} = \mathrm{E}\left(\nabla^{(2)}_{\boldsymbol{\theta}_\sigma}\zeta(y_{t+1}, X_{j,t}, \boldsymbol{\theta}^{\dagger}_{j,m}, \boldsymbol{\theta}^{\dagger}_{j,\sigma})\right)$ are positive definite while $\mathrm{E}\left(\nabla^{(1)}_{\boldsymbol{\theta}_m}\zeta(y_{s+1}, X_{j,s}, \boldsymbol{\theta}^{\dagger}_{j,m}, \boldsymbol{\theta}^{\dagger}_{j,\sigma})\,|X_{j,s}\right) = 0$ and $\mathrm{E}\left(\nabla^{(1)}_{\boldsymbol{\theta}_\sigma}\zeta(y_{s+1}, X_{j,s}, \boldsymbol{\theta}^{\dagger}_{j,m}, \boldsymbol{\theta}^{\dagger}_{j,\sigma})\mid X_{j,s}\right)$ $= 0$ almost surely. Finally, assume that:

$$\mathrm{E}\left(\|\nabla^{(1)}_{\boldsymbol{\theta}_m}\zeta(y_{s+1}, X_{j,s}, \boldsymbol{\theta}^{\dagger}_{j,m}, \boldsymbol{\theta}^{\dagger}_{j,\sigma})\|^2\right) < \infty \quad \text{and} \quad \mathrm{E}\left(\|\nabla^{(1)}_{\boldsymbol{\theta}_\sigma}\zeta(y_{s+1}, X_{j,s}, \boldsymbol{\theta}^{\dagger}_{j,m}, \boldsymbol{\theta}^{\dagger}_{j,\sigma})\|^2\right) < \infty.$$

(iii) For every $j \in \{1, \ldots, J\}$, it holds that:

$$0 < \sup_{X\in\mathcal{X}}|\sigma(X, \boldsymbol{\theta}^{\dagger}_{j,\sigma})| < \infty \quad \text{and} \quad \sup_{X\in\mathcal{X}}|m(X, \boldsymbol{\theta}^{\dagger}_{j,\sigma})| < \infty.$$

In addition, for $j = 1, \ldots, J$ and every $\boldsymbol{\theta}_{j,m} \in \boldsymbol{\Theta}_m$ and $\boldsymbol{\theta}_{j,\sigma} \in \boldsymbol{\Theta}_\sigma$, the functions $m(\cdot, \boldsymbol{\theta}_{j,m})$ and $\sigma(\cdot, \boldsymbol{\theta}_{j,\sigma})$ are continuously differentiable in the parameter vectors $\boldsymbol{\theta}_{j,m}$ and $\boldsymbol{\theta}_{j,\sigma}$ (a.s.), respectively, with uniformly bounded derivatives.

[1] More generally, if the derivative is taken with respect to the whole argument of the function, we omit the subscript in this document.

(iv) For every $j = 1, \ldots, J$, it holds that $\mathrm{E}\left(\epsilon_{j,t+1}^2\right) < \infty$, and that the Lebesgue density of $\epsilon_{j,t+1}$, say $f_{\epsilon_j}(\cdot)$, is strictly positive and continuously differentiable with bounded derivatives.

As mentioned before, Assumption A.6 is a high-level condition, which requires that the estimators of the parameter vectors $\boldsymbol{\theta}_{j,m}^{\dagger}$ and $\boldsymbol{\theta}_{j,\sigma}^{\dagger}$ of the nonlinear location scale model admit an asymptotic linear representation, which is satisfied for several commonly-used extremum estimators.

S.1.2 Estimation

For nonlinear parametric location scale models, we first estimate the conditional mean and variance parameters, say $\widehat{\boldsymbol{\theta}}_{j,m}$ and $\widehat{\boldsymbol{\theta}}_{j,\sigma}$, via quasi-maximum likelihood to get $m(X_{j,t}, \widehat{\boldsymbol{\theta}}_{j,m})$ and $\sigma(X_t, \widehat{\boldsymbol{\theta}}_{j,\sigma})$. In a second step, we then estimate the unconditional quantile of the error term ϵ_{t+1}, using the residuals:

$$\widehat{\beta}_{j,\epsilon,R}(\tau) = \arg\min_{\beta_\epsilon \in B_\epsilon} \frac{1}{R} \sum_{s=1}^{R-1} \rho_\tau \left(\frac{y_{s+1} - m\left(X_{j,s}, \widehat{\boldsymbol{\theta}}_{j,m,R}\right)}{\sigma(X_{j,s}, \widehat{\boldsymbol{\theta}}_{j,\sigma,R})} - \beta_\epsilon \right), \tag{S.2}$$

where $\widehat{\beta}_{j,\epsilon,R}(\tau)$ is an estimator of the τ-level quantile of ϵ_{j+1}.[2] In this case, the conditional quantile of the model is constructed as:

$$q_\tau(\widehat{\boldsymbol{\psi}}_{j,R}; X_{j,t}) = m(X_{j,t}, \widehat{\boldsymbol{\theta}}_{j,m,R}) + \sigma(X_{j,t}, \widehat{\boldsymbol{\theta}}_{j,\sigma,R}) \widehat{\beta}_{j,\epsilon,R}(\tau) \tag{S.3}$$

with $\widehat{\boldsymbol{\psi}}_{j,R} = (\widehat{\boldsymbol{\theta}}'_{j,m,R}, \widehat{\boldsymbol{\theta}}'_{j,\sigma,R}, \widehat{\beta}_{j,\epsilon,R}(\tau))'$.

S.2 Proofs

S.2.1 Auxiliary Lemmas

The proofs of the Theorems stated in the main text rely on the the following Lemmas for linear quantile regression and nonlinear location scale models.

Lemma Q.1: Let $q_\tau(\boldsymbol{\psi}^\dagger; X_{j,t}) = X'_{j,t}\boldsymbol{\beta}_j^\dagger(\tau)$. Under Assumptions A.1 and A.3, it holds that:

(i) For each $\tau \in \mathcal{T}$ and all $j = 1, \ldots, J$:

$$\left\| \widehat{\boldsymbol{\beta}}_{j,R}(\tau) - \boldsymbol{\beta}_j^\dagger(\tau) \right\| = o_p(1),$$

where $\boldsymbol{\beta}_j^\dagger(\tau)$ is defined in Equation (3) and $\widehat{\boldsymbol{\beta}}_{j,R}(\tau)$ in Equation (4).

(ii) For any $j = 1, \ldots, J$, the empirical process:

$$\frac{1}{\sqrt{R}} \sum_{t=1}^{R-1} \left(X_{j,t} \left(1\left\{ y_{t+1} \leq X'_{j,t}\boldsymbol{\beta} \right\} - \tau \right) - \mathrm{E}\left(X_{j,t} \left(1\left\{ y_{t+1} \leq X'_{j,t}\boldsymbol{\beta} \right\} - \tau \right) \right) \right)$$

is stochastically equicontinuous in $\boldsymbol{\beta} \in \mathcal{B}$ and $\tau \in \mathcal{T}$ w.r.t. the L_2 pseudo-metric:

$$\rho_{\mathcal{B}\times\mathcal{T}}((\tau, \boldsymbol{\beta}), (\tau', \boldsymbol{\beta}'))^2 = \max_{l \in d_j} \mathrm{E}\left(\left(X_{lj,t} \left(1\left\{ y_{t+1} \leq X'_{j,t}\boldsymbol{\beta} \right\} - \tau \right) - X_{lj,t} \left(1\left\{ y_{t+1} \leq X'_{j,t}\boldsymbol{\beta}' \right\} - \tau' \right) \right)^2 \right)$$

where $X_{lj,t}$ denotes the l-th element of $X_{j,t}$ and d_j is the dimension of $X_{j,t}$.

(iii) For each $j = 1, \ldots, J$ and $\tau \in \mathcal{T}$:

$$\sqrt{R}\left(\widehat{\boldsymbol{\beta}}_{j,R}(\tau) - \boldsymbol{\beta}_j^\dagger(\tau) \right)$$

[2] Throughout, we assume that $B_\epsilon \subset \mathcal{B}$.

$$= \quad H_j(\tau)^{-1}\left(\frac{1}{\sqrt{R}}\sum_{t=1}^{R-1}X_{j,t}\left(1\left\{y_{t+1}\leq X_{j,t}'\boldsymbol{\beta}_j^{\dagger}(\tau)\right\}-\tau\right)\right)+o_p(1),$$

where $H_j(\tau)$ is defined in Assumption A.3.

For the sake of notational brevity, we drop the model subscript $j = 1, \ldots, J$ in the following statement as well as in the corresponding proof in Subsection S.2.3.

Lemma L.1: Let $\widehat{\boldsymbol{\psi}}_R(\tau) = \left(\widehat{\boldsymbol{\theta}}_{m,R}', \widehat{\boldsymbol{\theta}}_{\sigma,R}', \widehat{\beta}_{\epsilon,R}(\tau)\right)'$ and $\boldsymbol{\psi}^{\dagger}(\tau) = \left(\boldsymbol{\theta}_m^{\dagger\prime}, \boldsymbol{\theta}_\sigma^{\dagger\prime}, \beta_\epsilon^{\dagger}(\tau)\right)'$. Also, with slight abuse of notation, let $\mathcal{X}$ denote a compact subset of R_X, the support of X_t. Under Assumptions A.1, A.3, and A.6, it holds that:

(i) For each $\tau \in \mathcal{T}$:

$$\left\|\widehat{\boldsymbol{\psi}}_R(\tau) - \boldsymbol{\psi}^{\dagger}(\tau)\right\| = o_p(1).$$

(ii) For each $\tau \in \mathcal{T}$ and uniformly in $X \in \mathcal{X}$:

$$\begin{aligned}
& \sqrt{R}\left(q_\tau(\widehat{\boldsymbol{\psi}}_R; X) - q_\tau(\boldsymbol{\psi}^{\dagger}; X)\right) \\
= \; & \sqrt{R}\left(\left(m(X, \widehat{\boldsymbol{\theta}}_{m,R}) + \sigma(X, \widehat{\boldsymbol{\theta}}_{\sigma,R})\widehat{\beta}_{\epsilon,R}(\tau)\right) - \left(m(X, \boldsymbol{\theta}_m^{\dagger}) + \sigma(X, \boldsymbol{\theta}_\sigma^{\dagger})\beta_\epsilon^{\dagger}(\tau)\right)\right) \\
= \; & \nabla_{\boldsymbol{\theta}_m} m(X, \boldsymbol{\theta}_m^{\dagger})\left(\mathbf{M}_m^{-1}\frac{1}{\sqrt{R}}\sum_{s=1}^{R-1}\nabla_{\boldsymbol{\theta}_m}\zeta(y_{s+1}, X_s, \boldsymbol{\theta}_m^{\dagger}, \boldsymbol{\theta}_\sigma^{\dagger})\right) \\
& + \nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X, \boldsymbol{\theta}_\sigma^{\dagger}\right)\beta_\epsilon^{\dagger}(\tau)\left(\mathbf{M}_\sigma^{-1}\frac{1}{\sqrt{R}}\sum_{s=1}^{R-1}\nabla_{\boldsymbol{\theta}_\sigma}\zeta(y_{s+1}, X_s, \boldsymbol{\theta}_m^{\dagger}, \boldsymbol{\theta}_\sigma^{\dagger})\right) \\
& + \sigma\left(X, \boldsymbol{\theta}_\sigma^{\dagger}\right)\Bigg(H(\tau)^{-1}\frac{1}{\sqrt{R}}\sum_{t=1}^{R-1}\left(1\left\{\epsilon_{t+1}\leq\beta_\epsilon^{\dagger}(\tau)\right\}-\tau\right) \\
& - H(\tau)^{-1}\,\mathrm{E}\left(f_\epsilon(\beta_\epsilon^{\dagger}(\tau))\frac{\nabla_{\boldsymbol{\theta}_m} m(X_t, \boldsymbol{\theta}_m^{\dagger})}{\sigma\left(X_t, \boldsymbol{\theta}_\sigma^{\dagger}\right)}\right)\left(\mathbf{M}_m^{-1}\frac{1}{\sqrt{R}}\sum_{s=1}^{R-1}\nabla_{\boldsymbol{\theta}_m}\zeta(y_{s+1}, X_s, \boldsymbol{\theta}_m^{\dagger}, \boldsymbol{\theta}_\sigma^{\dagger})\right) \\
& - H(\tau)^{-1}\,\mathrm{E}\left(f_\epsilon(\beta_\epsilon^{\dagger}(\tau))\epsilon_{t+1}\frac{\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X_t, \boldsymbol{\theta}_\sigma^{\dagger}\right)}{\sigma^2\left(X_t, \boldsymbol{\theta}_\sigma^{\dagger}\right)}\right)\left(\mathbf{M}_\sigma^{-1}\frac{1}{\sqrt{R}}\sum_{s=1}^{R-1}\nabla_{\boldsymbol{\theta}_\sigma}\zeta(y_{s+1}, X_s, \boldsymbol{\theta}_m^{\dagger}, \boldsymbol{\theta}_\sigma^{\dagger})\right)\Bigg) + o_p(1)
\end{aligned}$$

with:

$$\epsilon_{t+1} = \frac{y_{t+1} - m(X_t, \boldsymbol{\theta}_m^{\dagger})}{\sigma(X_t, \boldsymbol{\theta}_\sigma^{\dagger})} \quad \text{and} \quad \widehat{\epsilon}_{t+1} = \frac{y_{t+1} - m(X_t, \widehat{\boldsymbol{\theta}}_{m,R})}{\sigma(X_t, \widehat{\boldsymbol{\theta}}_{\sigma,R})},$$

while $H(\tau)$, $\mathbf{M}_m$, $\mathbf{M}_\sigma$, and $\zeta(y_{t+1}, X_t, \boldsymbol{\theta}_\sigma^{\dagger}, \boldsymbol{\theta}_\sigma^{\dagger})$ are defined in Assumptions A.3 and A.6, respectively.

S.2.2 Proofs of Theorems

Without loss of generality, we will examine the one-sided case with $\mathcal{E}_j\left((0,\tau]; \mathbf{X}_t\right)$, $j = 1, 2$, only since the two-sided case follows by analogous arguments. In addition, as we will apply linearisation arguments around population parameters $\boldsymbol{\psi}_j^{\dagger}(\tau)$, we adopt a slightly different notation w.r.t. the main text to make the parameter dependence also in the conditional coverage explicit. We write the empirical coverage as:

$$\widehat{C}_{j,P}\left(\widehat{\boldsymbol{\psi}}_{j,R}(\tau); \mathbf{X}_t\right) = \frac{\sum_{s=R}^{T-1} 1\left\{y_{s+1}\leq q_\tau(\widehat{\boldsymbol{\psi}}_{j,R}; X_{j,t})\right\}\mathbf{K}\left(\frac{\mathbf{X}_s - \mathbf{X}_t}{h}\right)}{\sum_{s=R}^{T-1}\mathbf{K}\left(\frac{\mathbf{X}_s - \mathbf{X}_t}{h}\right)},$$

while the population counterpart evaluated at the estimated conditional quantile $q_\tau(\widehat{\boldsymbol{\psi}}_{j,R};X_{j,t})$ becomes:

$$C_j\left(\widehat{\boldsymbol{\psi}}_{j,R}(\tau);\mathbf{X}_t\right)=F_{t+1}(q_\tau(\widehat{\boldsymbol{\psi}}_{j,R};X_{j,t})|\mathbf{X}_t).$$

Proof of Theorem 1:
(i) This part of the proof deals with CASE I under H_0. First, note that by a second order Taylor expansion:

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\left(L\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-\tau\right)-L\left(\widehat{C}_{2,P}\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_j\right)-\tau\right)\right)\\
=&\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\left(L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)-L\left(C_2\left(\boldsymbol{\psi}_2^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)\right)\\
&+\left(\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)\right)\right.\\
&\left.-\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_2\left(\boldsymbol{\psi}_2^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{2,P}\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_j\right)-C_2\left(\boldsymbol{\psi}_2^\dagger(\tau);\mathbf{X}_j\right)\right)\right)\\
&+\left(\frac{1}{2\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(2)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)\right)^2\right.\\
&\left.-\frac{1}{2\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(2)}L\left(C_2\left(\boldsymbol{\psi}_2^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{2,P}\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_j\right)-C_2\left(\boldsymbol{\psi}_2^\dagger(\tau);\mathbf{X}_j\right)\right)^2\right)\\
&+\left(\frac{1}{3!\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(3)}L\left(\overline{C}_{1,P}\left(\tau;\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)\right)^3\right.\\
&\left.-\frac{1}{3!\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(3)}L\left(\overline{C}_{2,P}\left(\tau;\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{2,P}\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_j\right)-C_2\left(\boldsymbol{\psi}_2^\dagger(\tau);\mathbf{X}_j\right)\right)^3\right)\\
=&\mathcal{T}_{1,P}+\mathcal{T}_{2,R,P}+\mathcal{T}_{3,R,P}+\mathcal{T}_{4,R,P},
\end{aligned}$$

where $\nabla^{(1)}L(\cdot)$, $\nabla^{(2)}L(\cdot)$, and $\nabla^{(3)}L(\cdot)$ denote the first, second, and third order derivative of $L(\cdot)$, respectively, while $\overline{C}_{l,P}(\tau;\mathbf{X}_j)$, $l=1,2$, lie between $\widehat{C}_{l,P}\left(\widehat{\boldsymbol{\psi}}_{l,R}(\tau);\mathbf{X}_j\right)$ and $C_l\left(\boldsymbol{\psi}_l^\dagger(\tau);\mathbf{X}_j\right)$. We organise the proof of part (i) into three steps, each of which deals with one of the terms from the above expansion in isolation. In particular, we will show that:

Step 1: Pointwise in τ:

$$\begin{aligned}
&\mathcal{T}_{1,P}+\mathcal{T}_{2,R,P}\\
=&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}\left(L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)-L\left(C_2\left(\boldsymbol{\psi}_2^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)\right)\\
&+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)\left(1\{y_{t+1}\leq q_\tau(\boldsymbol{\psi}_1^\dagger(\tau);X_{1,t})\}-F_{t+1}(q_\tau(\boldsymbol{\psi}_1^\dagger(\tau);X_{1,t})|\mathbf{X}_{1,t})\right)\\
&-\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_2\left(\boldsymbol{\psi}_2^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)\left(1\{y_{t+1}\leq q_\tau(\boldsymbol{\psi}_2^\dagger(\tau);X_{2,t})\}-F_{t+1}(q_\tau(\boldsymbol{\psi}_2^\dagger(\tau);X_{2,t})|\mathbf{X}_{2,t})\right)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}\varphi(\boldsymbol{\psi}_1^{\dagger}(\tau);y_{t+1},X_{1,t})-\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}\varphi(\boldsymbol{\psi}_2^{\dagger}(\tau);y_{t+1},X_{2,t})+o_p(1)\\
\equiv&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(A_{1,t}(\tau)-A_{2,t}(\tau))+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(B_{1,t}(\tau)-B_{2,t}(\tau))+\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}(D_{1,t}(\tau)-D_{2,t}(\tau))\\
&\xrightarrow{d} N(0,\Omega(\tau))
\end{aligned}
$$

where:

$$
\begin{aligned}
\Omega(\tau) &= \Omega_{AA}(\tau)+\Omega_{BB}(\tau)+\Omega_{DD}(\tau)+2\Omega_{AB}(\tau)\\
&= \mathrm{Avar}\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(A_{1,t}(\tau)-A_{2,t}(\tau))\right)+\mathrm{Avar}\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(B_{1,t}(\tau)-B_{2,t}(\tau))\right)\\
&\quad+\mathrm{Avar}\left(\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}(D_{1,t}(\tau)-D_{2,t}(\tau))\right) \qquad \text{(S.4)}\\
&\quad+2\mathrm{Acov}\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(A_{1,t}(\tau)-A_{2,t}(\tau)),\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(B_{1,t}(\tau)-B_{2,t}(\tau))\right).
\end{aligned}
$$

Here, for the linear quantile regression model:

$$
\varphi(\boldsymbol{\psi}_j^{\dagger}(\tau);y_{t+1},X_{j,t})=\Lambda_j(\tau)\left(H_j(\tau)^{-1}X_{j,t}\left(1\left\{y_{t+1}-X_{j,t}\boldsymbol{\beta}_j^{\dagger}(\tau)\le 0\right\}-\tau\right)\right) \qquad \text{(S.5)}
$$

and $\Lambda_j(\tau)=\mathrm{E}\left(1\{\mathbf{X}_t\in\mathcal{X}\}\left(\nabla L\left(C_j\left((0,\tau];\mathbf{X}_t\right)-\tau\right)f_{t+1}\left(X'_{j,t}\boldsymbol{\beta}_j^{\dagger}(\tau)|\mathbf{X}_t\right)X'_{j,t}\right)\right)$. On the other hand, in the location scale case we have for $j=1,2$:

$$
\begin{aligned}
&\varphi(\boldsymbol{\psi}_j^{\dagger}(\tau);y_{t+1},X_{j,t}) \qquad \text{(S.6)}\\
=\;&\widetilde{\Lambda}_{j,1}(\tau)\left(\mathbf{M}_{j,m}^{-1}\nabla_{\boldsymbol{\theta}_m}\zeta(y_{t+1},X_{j,t},\boldsymbol{\theta}_{m_j}^{\dagger},\boldsymbol{\theta}_{\sigma_j}^{\dagger})\right)+\widetilde{\Lambda}_{j,2}(\tau)\left(\mathbf{M}_{j,\sigma}^{-1}\nabla_{\boldsymbol{\theta}_\sigma}\zeta(y_{t+1},X_{j,t},\boldsymbol{\theta}_{m_j}^{\dagger},\boldsymbol{\theta}_{\sigma_j}^{\dagger})\right)\\
&+\widetilde{\Lambda}_{j,3}(\tau)\Bigg(H_j(\tau)^{-1}\left(1\left\{\epsilon_{j,t+1}\le q_\tau\left(\epsilon_{j,t+1}\right)\right\}-\tau\right)\\
&-H_j(\tau)^{-1}\,\mathrm{E}\left(f_{\epsilon_j}(q_\tau(\epsilon_{j,t+1}))\frac{\nabla_{\boldsymbol{\theta}_m}m(X_t,\boldsymbol{\theta}_{m_j}^{\dagger})}{\sigma\left(X_t,\boldsymbol{\theta}_{\sigma_j}^{\dagger}\right)}\right)\left(\mathbf{M}_{j,m}^{-1}\nabla_{\boldsymbol{\theta}_m}\zeta(y_{t+1},X_{j,t},\boldsymbol{\theta}_{m_j}^{\dagger},\boldsymbol{\theta}_{\sigma_j}^{\dagger})\right)\\
&-H_j(\tau)^{-1}\,\mathrm{E}\left(f_{\epsilon_j}(q_\tau(\epsilon_{j,t+1}))\epsilon_{j,t+1}\frac{\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X_{j,t},\boldsymbol{\theta}_{\sigma_j}^{\dagger}\right)}{\sigma\left(X_{j,t},\boldsymbol{\theta}_{\sigma_j}^{\dagger}\right)}\right)\left(\mathbf{M}_{j,\sigma}^{-1}\nabla_{\boldsymbol{\theta}_\sigma}\zeta(y_{t+1},X_{j,t},\boldsymbol{\theta}_{m_j}^{\dagger},\boldsymbol{\theta}_{\sigma_j}^{\dagger})\right)\Bigg)
\end{aligned}
$$

with terms $\widetilde{\Lambda}_{j,1}(\tau)=\mathrm{E}\left(1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_j\left(\boldsymbol{\psi}_j^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)f_{t+1}(q_\tau(\boldsymbol{\psi}_j^{\dagger}(\tau);X_{j,t})|\mathbf{X}_t)\nabla_{\boldsymbol{\theta}_m}m(X_{j,t},\boldsymbol{\theta}_{m_j}^{\dagger})\right)$, $\widetilde{\Lambda}_{j,2}(\tau)=\mathrm{E}\left(1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_j\left(\boldsymbol{\psi}_j^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)f_{t+1}(q_\tau(\boldsymbol{\psi}_j^{\dagger}(\tau);X_{j,t})|\mathbf{X}_t)\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X_{j,t},\boldsymbol{\theta}_{\sigma_j}^{\dagger}\right)q_\tau(\epsilon_{j,t+1})\right)$ and $\widetilde{\Lambda}_{j,3}(\tau)=\mathrm{E}\left(1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_j\left(\boldsymbol{\psi}_j^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)f_{t+1}(q_\tau(\boldsymbol{\psi}_j^{\dagger}(\tau);X_{j,t})|\mathbf{X}_t)\sigma\left(X_{j,t},\boldsymbol{\theta}_{\sigma_j}^{\dagger}\right)\right)$.

Step 2: Pointwise in τ:

$$
\mathcal{T}_{3,R,P}=O_p\left(\frac{\sqrt{P}}{R}\right)=o_p(1).
$$

Step 3: Pointwise in τ:

$$
\mathcal{T}_{4,R,P}=o_p\left(\frac{\sqrt{P}}{R}\right)=o_p(1).
$$

Thus, in CASE I under H_0, we have that:

$$\begin{aligned}
& \mathcal{T}_{1,P}+\mathcal{T}_{2,R,P}+\mathcal{T}_{3,R,P}+\mathcal{T}_{4,R,P} \\
= & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(A_{1,t}(\tau)-A_{2,t}(\tau)\right)+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(B_{1,t}(\tau)-B_{2,t}(\tau)\right)+\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}\left(D_{1,t}(\tau)-D_{2,t}(\tau)\right) \\
& +o_p(1),
\end{aligned} \tag{S.7}$$

where the RHS of (S.7) converges to a mean zero Gaussian distribution with variance-covariance kernel $\Omega(\tau)$. We now proceed by proving each step in turn.

Proof of Step 1: We decompose $\mathcal{T}_{1,P}$ as follows:

$$\mathcal{T}_{1,P}=\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\left(L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)-L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\right)$$

$$\begin{aligned}
= & \frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}\left(1\{\mathbf{X}_j\in\mathcal{X}\}\left(L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)-L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\right)\right. \\
& \left.-\mathrm{E}\left(1\{\mathbf{X}_j\in\mathcal{X}\}\left(L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)-L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\right)\right)\right) \\
& +\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}\mathrm{E}\left(1\{\mathbf{X}_j\in\mathcal{X}\}\left(L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)-L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\right)\right)
\end{aligned}$$

In CASE I, under both hypotheses, we have that:

$$\begin{aligned}
& \frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}\left(1\{\mathbf{X}_j\in\mathcal{X}\}\left(L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)-L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\right)\right. \\
& \left.-\mathrm{E}\left(1\{\mathbf{X}_j\in\mathcal{X}\}\left(L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)-L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\right)\right)\right)
\end{aligned}$$

converges weakly to a zero mean Gaussian process by a CLT for strong mixing, bounded observations, see e.g. Corollary 5.1 in Hall and Heyde (1980). Moreover, under the null hypothesis:

$$\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}\mathrm{E}\left(1\{\mathbf{X}_j\in\mathcal{X}\}\left(L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)-L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\right)\right)=0,$$

and so:

$$\mathcal{T}_{1,P}=\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(A_{1,t}(\tau)-A_{2,t}(\tau)\right).$$

Turning to $\mathcal{T}_{2,R,P}$, note that we can write this term as:

$$\begin{aligned}
& \left(\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)\right)\right. \\
& \left.-\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{2,P}\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_j\right)-C_2\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_j\right)\right)\right) \\
& +\left(\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\left(C_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)\right)\right. \\
& \left.-\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\left(C_2\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_j\right)-C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)\right)\right)
\end{aligned} \tag{S.8}$$

$$= \mathcal{T}_{2,R,P}(A) + \mathcal{T}_{2,R,P}(B).$$

We start with $\mathcal{T}_{2,R,P}(A)$ and focus exclusively on the part involving model 1 (the arguments for model 2 will be analogous). Next, define $N_{\boldsymbol{\psi}_1,R} = \left\{\boldsymbol{\psi}_1 : \|\boldsymbol{\psi}_1 - \boldsymbol{\psi}_1^\dagger(\tau)\| \leq CR^{-\frac{1}{2}},\, \boldsymbol{\psi}_1 \in \boldsymbol{\Psi}\right\}$ for some constant $C > 0$, and note that $\widehat{\boldsymbol{\psi}}_{1,R}(\tau) \in N_{\boldsymbol{\psi}_1,R}$ with probability approaching one as $P \to \infty$ by either Lemma Q.1 or Lemma L.1. Then, the part of $\mathcal{T}_{2,R,P}(A)$ that involves model 1 reads as:

$$\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}\Big(1\{\mathbf{X}_j \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1;\mathbf{X}_j\right)\right)\\
&-\mathrm{E}_T\left(1\{\mathbf{X}_j \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1;\mathbf{X}_j\right)\right)\right)\Big) \qquad \text{(S.9)}\\
&+\sqrt{P}\mathrm{E}_T\left(1\{\mathbf{X}_j \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1;\mathbf{X}_j\right)\right)\right),
\end{aligned}$$

where E_T denotes the expectation conditional on the original sample $\{y_{t+1},\mathbf{X}_t\}_{t=1}^{T-1}$. To show that the first term of (S.9) is $o_p(1)$ we need to verify that this term is (i) stochastically equicontinuous for the metric space $\mathcal{C}$ equipped with pseudo-metric $\rho_{\mathcal{C}}(\cdot)$, and that for all $\mathbf{X} \in \mathcal{X}$ and $\boldsymbol{\psi}_1 \in N_{\boldsymbol{\psi}_1,R}$, (ii) $\Pr\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}\right) \in \mathcal{C}\right) \to 1$ as well as (iii) $\rho_{\mathcal{C}}\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}\right), C_1\left(\boldsymbol{\psi}_1;\mathbf{X}\right)\right) \xrightarrow{p} 1$ (cf. Andrews, 1994, p.2265), where we use the L_2 pseudo-metric:

$$\rho_{\mathcal{C}}\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}\right), C_1\left(\boldsymbol{\psi}_1;\mathbf{X}\right)\right)^2 = \sup_{\boldsymbol{\psi}_1 \in N_{\boldsymbol{\psi}_1,R},\mathbf{X}\in\mathcal{X}} \mathrm{E}\left(\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}\right) - C_1\left(\boldsymbol{\psi}_1;\mathbf{X}\right)\right)^2 \big| \{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right),$$

where $\mathrm{E}\left(\cdot\,\big|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)$ denotes the expectation conditional on the sample, while $\mathrm{Var}\left(\cdot\,\big|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)$ denotes the corresponding variance. Starting with (ii), this follows from Theorem 1 of Andrews (1995) using A.1, A.4, A.5. Moreover, the results of Theorem 1 of Andrews (1995) also imply convergence w.r.t. the L_2 pseudo-metric since:

$$\begin{aligned}
\rho_{\mathcal{C}}\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}\right), C_1\left(\boldsymbol{\psi}_1;\mathbf{X}\right)\right)^2 &\leq \sup_{\boldsymbol{\psi}_1 \in N_{\boldsymbol{\psi}_1,R},\mathbf{X}\in\mathcal{X}} \mathrm{Var}\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}\right)\big|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)\\
&\quad+ \sup_{\boldsymbol{\psi}_1 \in N_{\boldsymbol{\psi}_1,R},\mathbf{X}\in\mathcal{X}} \left\{\left(\mathrm{E}\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}\,\big|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)\right) - C_1\left(\boldsymbol{\psi}_1;\mathbf{X}\right)\right)^2\right\}\\
&= o_p(1).
\end{aligned}$$

Finally, to verify (i), note that the function class:

$$\mathcal{F} = \left\{\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);x\right)-\tau\right)C_1\left(\boldsymbol{\psi}_1;x\right):\ \tau \in \mathcal{T},\ \boldsymbol{\psi}_1 \in N_{\boldsymbol{\psi}_1,R},\ x \in \mathcal{X}\right\}$$

is uniformly bounded by A.2 and A.4, and satisfies an L_2 continuity condition with bound:

$$C \sup_{\tau'\in\mathcal{T},|\tau'-\tau|\leq r_1}|\tau'-\tau|^2 + \widetilde{C}\sup_{\boldsymbol{\psi}_1'\in N_{\boldsymbol{\psi}_1,R},\|\boldsymbol{\psi}_1'-\boldsymbol{\psi}_1\|\leq r_2}\|\boldsymbol{\psi}_1'-\boldsymbol{\psi}_1\|^2 + \widetilde{\widetilde{C}}\sup_{x\in\mathcal{X},\|x'-x\|\leq r_3}\|x'-x\|^2$$

for some generic positive constants $C, \widetilde{C}, \widetilde{\widetilde{C}}$ such that $\sqrt{r_1^2+r_2^2+r_3^2} \leq r$. It follows that the bracketing condition of Theorem 2.2 in Andrews and Pollard (1994) holds because the L_2 continuity condition implies that the bracketing number satisfies:

$$N(\eta,\mathcal{F}) \leq \overline{C}\left(\frac{1}{\eta}\right)^{2d+1},$$

see Andrews and Pollard (1994, p.121). Moreover, setting $Q = 2$ and $\gamma = \varepsilon = 1$, we have that the mixing condition of Theorem 2.2 therein is satisfied, and hence the first term of (S.9) satisfies the stochastic equicontinuity condition. On the other hand, for the second term of (S.9), note that:

$$\sqrt{P}\mathrm{E}_T\left(1\{\mathbf{X}_j \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)\left(\widehat{C}_{1,P}\left(\boldsymbol{\psi}_1;\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1;\mathbf{X}_j\right)\right)\right)$$

$$
\begin{aligned}
= & \frac{1}{\sqrt{P}}\sum_{s=R}^{T-1}\int \frac{1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)}{h^d f_X(\mathbf{X}_j)}(1\left\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,j})\right\} \\
& -F_{j+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,j})|\mathbf{X}_j))\mathbf{K}\left(\frac{\mathbf{X}_s-\mathbf{X}_j}{h}\right)f_X(\mathbf{X}_j)\mathrm{d}\mathbf{X}_j \qquad \text{(S.10)}\\
& +\frac{1}{\sqrt{P}}\sum_{s=R}^{T-1}\int\left(\frac{1}{\widehat{f}_X(\mathbf{X}_j)}-\frac{1}{f_X(\mathbf{X}_j)}\right)\frac{1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)}{h^d} \\
& (1\left\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,j})\right\}-F_{j+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,j})|\mathbf{X}_j))\mathbf{K}\left(\frac{\mathbf{X}_s-\mathbf{X}_j}{h}\right)f_X(\mathbf{X}_j)\mathrm{d}\mathbf{X}_j,
\end{aligned}
$$

where we use the fact that $\mathrm{E}[1\{y_{j+1}<q_\tau(\boldsymbol{\psi}_j;X_{1,j})\}|\mathbf{X}_j]=F_{j+1}(q_\tau(\boldsymbol{\psi}_j;X_{1,j})|\mathbf{X}_j)$. Since $\sup_{x\in\mathcal{X}}|\widehat{f}_X(x)-f_X(x)|=o_p(1)$ by the bandwidth conditions as well as A.1, A.4, and A.5 (Andrews, 1995, Theorem 1), the second term in (S.10) is of smaller probability order than the first one, and hence we will focus on the first one in what follows. That is, by change of variables with $u=(\mathbf{X}_s-\mathbf{X}_j)/h$ (let $u_1=(X_{1,s}-X_{1,j})/h$ denote its first element), the first term on the RHS of (S.10) equals:

$$
\begin{aligned}
& \frac{1}{\sqrt{P}}\sum_{s=R}^{T-1}\int 1\left\{(\mathbf{X}_s+hu)\in\mathcal{X}\right\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s+hu\right)-\tau\right)(1\left\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s}+hu_1)\right\} \\
& -F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s}+hu_1)|\mathbf{X}_s+hu))\mathbf{K}\left(u\right)\mathrm{d}u \\
= & \mathcal{A}_{1,P}(\boldsymbol{\psi}_1)+\mathcal{A}_{2,P}(\boldsymbol{\psi}_1)+\mathcal{A}_{3,P}(\boldsymbol{\psi}_1)+\mathcal{A}_{4,P}(\boldsymbol{\psi}_1)+\mathcal{A}_{5,P}(\boldsymbol{\psi}_1),
\end{aligned}
$$

where:

$$
\begin{aligned}
\mathcal{A}_{1,P}(\boldsymbol{\psi}_1)= & \frac{1}{\sqrt{P}}\sum_{s=R}^{T-1}1\{\mathbf{X}_s\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\right)-\tau\right) \\
& (1\left\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s})\right\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s)),
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{A}_{2,P}(\boldsymbol{\psi}_1)= & \frac{1}{\sqrt{P}}\sum_{s=R}^{T-1}\int 1\{\mathbf{X}_s\in\mathcal{X}\}\left(\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s+hu\right)-\tau\right)-\nabla L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\right)-\tau\right)\right) \\
& (1\left\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s})\right\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s))\mathbf{K}\left(u\right)\mathrm{d}u, \qquad \text{(S.11)}\\
\mathcal{A}_{3,P}(\boldsymbol{\psi}_1)= & \frac{1}{\sqrt{P}}\sum_{s=R}^{T-1}\int\left(1\{(\mathbf{X}_s+hu)\in\mathcal{X}\}-1\{\mathbf{X}_s\in\mathcal{X}\}\right)\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\right)-\tau\right) \\
& (1\left\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s})\right\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s))\,\mathbf{K}\left(u\right)\mathrm{d}u, \qquad \text{(S.12)}\\
\mathcal{A}_{4,P}(\boldsymbol{\psi}_1)= & \frac{1}{\sqrt{P}}\sum_{s=R}^{T-1}\int 1\{\mathbf{X}_s\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\right)-\tau\right)((1\left\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s}+hu_1)\right\} \\
& \qquad \text{(S.13)}\\
& -F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s}+hu_1)|\mathbf{X}_s+hu))-(1\left\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s})\right\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s)))\,\mathbf{K}\left(u\right)\mathrm{d}u, \\
\mathcal{A}_{5,P}(\boldsymbol{\psi}_1)= & \frac{1}{\sqrt{P}}\sum_{s=R}^{T-1}\int\left(1\{(\mathbf{X}_s+hu)\in\mathcal{X}\}-1\{\mathbf{X}_s\in\mathcal{X}\}\right) \\
& \times\left(\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s+hu\right)-\tau\right)-\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\right)-\tau\right)\right) \\
& \times((1\left\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s}+hu_1)\right\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s}+hu_1)|\mathbf{X}_s+hu) \\
& -(1\left\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s})\right\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s)))\,\mathbf{K}\left(u\right)\mathrm{d}u. \qquad \text{(S.14)}
\end{aligned}
$$

As for $\mathcal{A}_{1,P}(\boldsymbol{\psi}_1)$, note that for every $\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}$, this term is mean zero by iterated expectations. Therefore, by a CLT for strong mixing, bounded observations (see Hall and Heyde, 1980, Corollary

5.1), we have again that $\mathcal{A}_{1,P}(\boldsymbol{\psi}_1)$ converges pointwise in $N_{\boldsymbol{\psi}_1,R}$ to a zero mean Gaussian random variable. Moreover, similar to before, note that the function class:

$$\mathcal{F} = \Big\{1\{\mathbf{X}_s \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\right)-\tau\right)(1\{y_{s+1}\le q_\tau(\boldsymbol{\psi}_1;X_{1,s})\} \\ -F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s)) : \ \boldsymbol{\psi}_1 \in N_{\boldsymbol{\psi}_1,R}\Big\}$$

is uniformly bounded by A.2 and A.4, and can be shown to satisfy the L_2 continuity condition with bound:

$$C \sup_{\boldsymbol{\psi}_1'\in N_{\boldsymbol{\psi}_1,R},\|\boldsymbol{\psi}_1'-\boldsymbol{\psi}_1\|\le r_2} \|\boldsymbol{\psi}_1'-\boldsymbol{\psi}_1\|^2.$$

It follows again that the bracketing condition of Theorem 2.2 in Andrews and Pollard (1994) holds. Thus, setting again $Q=2$ and $\gamma=\epsilon=1$ for A.1, we obtain that the empirical process $\mathcal{A}_{1,P}(\boldsymbol{\psi}_1)$ is also stochastically equicontinuous in $\boldsymbol{\psi}_1$. Since we also show below that $\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}|\mathcal{A}_{j,P}(\boldsymbol{\psi}_1)| = o_p(1)$, $j=2,\ldots,5$, it follows that uniformly in $\boldsymbol{\psi}_1$:

$$\mathcal{T}_{2,R,P}(A) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(B_{1,t}(\tau)-B_{2,t}(\tau)) + o_p(1).$$

Next, we analyse $\mathcal{A}_{2,P}(\boldsymbol{\psi}_1)$ and $\mathcal{A}_{3,P}(\boldsymbol{\psi}_1)$ from (S.11) and (S.12), respectively. Note that, by Fubini's Theorem and iterated expectations, both terms are mean zero. As for the variance, note that:

$$\begin{aligned}
&\mathrm{Var}\left(\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}|\mathcal{A}_{2,P}(\boldsymbol{\psi}_1)|\right)\\
\le\ & \frac{1}{P}\sum_{s=R}^{T-1}\mathrm{Var}\Bigg(\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}\Bigg|\left(\int 1\{\mathbf{X}_s\in\mathcal{X}\}\left(\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s+hu\right)-\tau\right)-\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\right)-\tau\right)\right)\mathbf{K}(u)\,\mathrm{d}u\right)\\
&\times(1\{y_{s+1}\le q_\tau(\boldsymbol{\psi}_1;X_{1,s})\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s)\Bigg)\Bigg|\\
&+\frac{2}{P}\sum_{s=R}^{T-1}\sum_{t>s}\mathrm{Cov}\Bigg(\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}\Bigg|\left(\int\left(\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s+hu\right)-\tau\right)-\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\right)-\tau\right)\right)\mathbf{K}(u)\,\mathrm{d}u\right)\\
&\times 1\{\mathbf{X}_s\in\mathcal{X}\}(1\{y_{s+1}\le q_\tau(\boldsymbol{\psi}_1;X_{1,s})\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s)\\
&\times\left(\int\left(\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t+hu\right)-\tau\right)-\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)\right)\mathbf{K}(u)\,\mathrm{d}u\right)\\
&\times 1\{\mathbf{X}_t\in\mathcal{X}\}(1\{y_{t+1}\le q_\tau(\boldsymbol{\psi}_1;X_{1,t})\}-F_{t+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,t})|\mathbf{X}_t)\Bigg)\Bigg|\\
=\ & \mathcal{A}_{2,P}(A)+\mathcal{A}_{2,P}(B)
\end{aligned}$$

We start with $\mathcal{A}_{2,P}(A)$, which can be bounded by:

$$\begin{aligned}
&\mathrm{E}\Bigg(\left(\int\nabla^{(2)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\overline{\mathbf{X}}_s\right)-\tau\right)\nabla_X F_{s+1}(q_\tau(\boldsymbol{\psi}_1^{\dagger};\overline{X}_{1,s})|\overline{\mathbf{X}}_s)uh\mathbf{K}(u)\,\mathrm{d}u\right)^2\\
&\times\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}|1\{\mathbf{X}_s\in\mathcal{X}\}(1\{y_{s+1}\le q_\tau(\boldsymbol{\psi}_1;X_{1,s})\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s))|^2\Bigg)\\
=\ &\mathrm{E}\Bigg(\left(\int\nabla^{(2)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\overline{\mathbf{X}}_s\right)-\tau\right)\nabla_X F_{s+1}(q_\tau(\boldsymbol{\psi}_1^{\dagger};\overline{X}_{1,s})|\overline{\mathbf{X}}_s)1\{\overline{\mathbf{X}}_s\in\mathcal{X}\}uh\mathbf{K}(u)\,\mathrm{d}u\right)^2\\
&\times\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}|1\{\mathbf{X}_s\in\mathcal{X}\}(1\{y_{s+1}\le q_\tau(\boldsymbol{\psi}_1;X_{1,s})\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s))|^2\Bigg)(1+o(1))
\end{aligned}$$

Using Jensen's inequality, the lead term can be bounded by:

$$Ch^2\left(\int u^2\mathbf{K}(u)^2\,\mathrm{d}u\right)\mathrm{E}\left(\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}|1\{\mathbf{X}_s\in\mathcal{X}\}(1\{y_{s+1}\le q_\tau(\boldsymbol{\psi}_1;X_{1,s})\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s))|^2\right)=O(h^2),$$

which holds for all $\boldsymbol{\psi}_1 \in N_{\boldsymbol{\psi}_1,R}$. For $\mathcal{A}_{2,P}(B)$ on the other hand, since $\beta-$mixing processes are also strong mixing of the same size, note that for some $\kappa > 2$ satisfying $(\kappa-2)/\kappa > \varepsilon/(2+\varepsilon)$ and positive constants C, C':

$$
\begin{aligned}
& |\mathcal{A}_{2,P}(B)| \\
\leq\ & \frac{C}{P}\sum_{s=R}^{T-1}\sum_{t>s}\beta(t-s)^{1-\frac{2}{\kappa}}\Bigg(\mathrm{E}\Bigg(\Bigg(\int\Big(\nabla^{(1)}L\Big(C_1\Big(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s+hu\Big)-\tau\Big) \\
& -\nabla^{(1)}L\Big(C_1\Big(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\Big)-\tau\Big)\Big)\mathbf{K}(u)\,\mathrm{d}u\Bigg)\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}|1\{\mathbf{X}_s\in\mathcal{X}\}(1\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s})\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s))|\Bigg)^{\kappa}\Bigg)^{\frac{1}{\kappa}} \\
\leq\ & C'h^2\sum_{j=1}^{\infty}\beta(j)^{\frac{\varepsilon}{2+\varepsilon}}
\end{aligned}
$$

where the first inequality follows from Corollary A.2 in Hall and Heyde (1980) and the last bound is again independent of $\boldsymbol{\psi}_1$. Thus, by A.1, we have that this term is of order $O(h^2)$. Turning to $\mathcal{A}_{3,P}(\boldsymbol{\psi}_1)$, since this term is also mean zero, we may again bound the lead term of the variance using Cauchy-Schwarz's and Jensen's inequality:

$$
\begin{aligned}
& \mathrm{E}\Bigg(\Bigg(\int\left(1\{(\mathbf{X}_s+hu)\in\mathcal{X}\}-1\{\mathbf{X}_s\in\mathcal{X}\}\right)\mathbf{K}(u)\,\mathrm{d}u\Bigg)^2\nabla^{(1)}L\Big(C_1\Big(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\Big)-\tau\Big)^2 \\
& \times\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}|1\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s})\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s)|^2\Bigg) \\
\leq\ & \Bigg(\mathrm{E}\Bigg(\int\left(1\{(\mathbf{X}_s+hu)\in\mathcal{X}\}-1\{\mathbf{X}_s\in\mathcal{X}\}\right)^4\mathbf{K}(u)^4\,\mathrm{d}u\Bigg)\Bigg)^{\frac{1}{2}}\Bigg(\mathrm{E}\Bigg(\Big(\nabla^{(1)}L\Big(C_1\Big(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\Big)-\tau\Big)\Big)^4 \\
& \times\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}|1\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s})\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s)|^4\Bigg)\Bigg)^{\frac{1}{2}}
\end{aligned}
$$

The second expectation on the RHS is of order $O(1)$ by A.2 and the boundedness of the indicator and the conditional distribution function, and thus holds for all $\boldsymbol{\psi}_1 \in N_{\boldsymbol{\psi}_1,R}$. For the first expectation on the other hand, observe that $1\{(\mathbf{X}_s+hu)\in\mathcal{X}\}-1\{\mathbf{X}_s\in\mathcal{X}\} = 1\{(\mathbf{X}_s+hu)\in\mathcal{X},\mathbf{X}_s\notin\mathcal{X}\}+1\{\mathbf{X}_s\in\mathcal{X},(\mathbf{X}_s+hu)\notin\mathcal{X}\}$ a.s.. Thus, by Fubini's Theorem we have for this expression that:

$$
\begin{aligned}
& \int\left(\Pr\left((\mathbf{X}_s+hu)\in\mathcal{X},\mathbf{X}_s\notin\mathcal{X}\right)+\Pr\left(\mathbf{X}_s\in\mathcal{X},(\mathbf{X}_s+hu)\notin\mathcal{X}\right)\right)\mathbf{K}(u)^4\,\mathrm{d}u \\
\leq\ & \left(\sup_{u\in[-1,1]}\Pr\left((x+hu)\in\mathcal{X},x\notin\mathcal{X}\right)+\sup_{u\in[-1,1]}\Pr\left((x+hu)\in\mathcal{X},x\notin\mathcal{X}\right)\right)\int\mathbf{K}(u)^4\,\mathrm{d}u\to 0
\end{aligned}
$$

by A.4, A.5, since $h\to 0$ as $P\to\infty$. Finally, for $\mathcal{A}_{4,P}(\boldsymbol{\psi}_1)$ in (S.13), we may decompose the term as follows:

$$
\begin{aligned}
\mathcal{A}_{4,P}(\boldsymbol{\psi}_1) =\ & \frac{1}{\sqrt{P}}\sum_{s=R}^{T-1}\int 1\{\mathbf{X}_s\in\mathcal{X}\}\nabla^{(1)}L\Big(C_1\Big(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\Big)-\tau\Big)\left((1\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s}+hu_1)\}\right. \\
& \left.-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s}+hu_1)|\mathbf{X}_s))-(1\{y_{s+1}\leq q_\tau(\boldsymbol{\psi}_1;X_{1,s})\}-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s})|\mathbf{X}_s))\right)\mathbf{K}(u)\,\mathrm{d}u \\
& +\frac{1}{\sqrt{P}}\sum_{s=R}^{T-1}\int 1\{\mathbf{X}_s\in\mathcal{X}\}\nabla^{(1)}L\Big(C_1\Big(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_s\Big)-\tau\Big)\left(F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s}+hu_1)|\mathbf{X}_s)\right. \\
& \left.-F_{s+1}(q_\tau(\boldsymbol{\psi}_1;X_{1,s}+hu_1)|\mathbf{X}_s+hu)\right)\mathbf{K}(u)\,\mathrm{d}u \\
=\ & \mathcal{A}_{4,P}(A;\boldsymbol{\psi}_1)+\mathcal{A}_{4,P}(B,\boldsymbol{\psi}_1)
\end{aligned} \tag{S.15}
$$

Since $\mathcal{A}_{4,P}(A;\boldsymbol{\psi}_1)$ is mean zero by iterated expectations, we can address $\mathcal{A}_{4,P}(A;\boldsymbol{\psi}_1)$ by similar arguments to before to show that the lead term of the variance is of order $O(h)$ uniformly in $\boldsymbol{\psi}_1 \in N_{\boldsymbol{\psi}_1,R}$. For the bias term $\mathcal{A}_{4,P}(B;\boldsymbol{\psi}_1)$ on the other hand, using A.5 and standard Taylor expansion arguments first around $X_{1,s}$ and subsequently around $\mathbf{X}_s$ yield that $\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}|\mathcal{A}_{4,P}(B;\boldsymbol{\psi}_1)| = O(\sqrt{P}h^r) = o(1)$ since $Ph^{2r}\to 0$. Finally, similar arguments to the ones above may also be used to show that $\sup_{\boldsymbol{\psi}_1\in N_{\boldsymbol{\psi}_1,R}}|\mathcal{A}_{5,P}(\boldsymbol{\psi}_1)|$ is of order $o(h) = o(1)$.

We next move to $\mathcal{T}_{2,R,P}(B)$ from (S.8), where we focus again on model 1 exclusively. More specifically:

$$
\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\Big(C_1\Big(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\Big)-\tau\Big)\Big(C_1\Big(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\Big)-C_1\Big(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\Big)\Big)
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{P}}\sum_{j=R}^{T-1} 1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)f_{j+1}(\overline{q}_{\tau}(\boldsymbol{\psi}_{1,R}(\tau);X_{1,j})|\mathbf{X}_j)\\
&\quad\times\left(q_{\tau}(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);X_{1,j})-q_{\tau}(\boldsymbol{\psi}_1^{\dagger}(\tau);X_{1,j})\right),
\end{aligned}
$$

where $\overline{q}_{\tau}(\boldsymbol{\psi}_{1,R}(\tau);X_{1,j})$ denotes an intermediate value. Focusing on the case of quantile regression and inserting the linear representation from Lemma Q.1 (as well as substituting $\boldsymbol{\beta}_1^{\dagger}(\tau)$ for $\boldsymbol{\psi}_1^{\dagger}(\tau)$), one can show that:

$$
\begin{aligned}
&\sqrt{P}\mathrm{E}\left(1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)f_{j+1}(X_{1,j}'\boldsymbol{\beta}_1^{\dagger}(\tau)|\mathbf{X}_j)X_{1,j}'\right)\left(H_1(\tau)^{-1}\right.\\
&\left.\times\frac{1}{R}\sum_{s=1}^{R-1}X_{1,s}\left(1\left\{y_{s+1}-X_{1,s}'\boldsymbol{\beta}_1^{\dagger}(\tau)\le 0\right\}-\tau\right)\right)(1+o_p(1))\\
=&\ \frac{\sqrt{P}}{R}\sum_{s=1}^{R-1}\Lambda_1(\tau)\left(H_1(\tau)^{-1}X_{1,s}\left(1\left\{y_{s+1}-X_{1,s}\boldsymbol{\beta}_1^{\dagger}(\tau)\le 0\right\}-\tau\right)\right)(1+o_p(1))
\end{aligned}
$$

where:

$$
\Lambda_1(\tau)\equiv\mathrm{E}\left(1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)f_{j+1}(X_{1,j}'\boldsymbol{\beta}_1^{\dagger}(\tau)|\mathbf{X}_j)X_{1,j}'\right). \tag{S.16}
$$

On the other hand, using the representation from Lemma L.1 yields the expression for the location scale model. We therefore have that:

$$
\mathcal{T}_{2,R,P}(B)=\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}\left(D_{1,t}(\tau)-D_{2,t}(\tau)\right)+o_p(1).
$$

Summarizing the above, we obtain the statement from Step 1, where $\Omega(\tau)$ follows since

$$
\begin{aligned}
&\mathrm{Acov}\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(A_{1,t}(\tau)-A_{2,t}(\tau)\right),\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}\left(D_{1,t}(\tau)-D_{2,t}(\tau)\right)\right)\\
=&\ \mathrm{Acov}\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(B_{1,t}(\tau)-B_{2,t}(\tau)\right),\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}\left(D_{1,t}(\tau)-D_{2,t}(\tau)\right)\right)\\
=&\ 0
\end{aligned}
$$

due to the use of a fixed estimation scheme.

Proof of Step 2: We now turn to $\mathcal{T}_{3,R,P}$ and show that provided $Ph^{2d}\to\infty$, $\mathcal{T}_{3,R,P}=o_p(1)$. For brevity, we only consider the squared estimation error component for model 1, say $\mathcal{T}_{3,R,P}^{(1)}$.
Letting $\omega_1(\tau;\mathbf{X}_j)=1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(2)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)$, we write:

$$
\begin{aligned}
\mathcal{T}_{3,R,P}^{(1)} &= \frac{1}{2\sqrt{P}}\sum_{j=R}^{T-1}\omega_1(\tau;\mathbf{X}_j)\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)\right)^2\\
&= \frac{1}{2\sqrt{P}}\sum_{j=R}^{T-1}\omega_1(\tau;\mathbf{X}_j)\left(C_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)\right)^2\\
&\quad+\frac{1}{2\sqrt{P}}\sum_{j=R}^{T-1}\omega_1(\tau;\mathbf{X}_j)\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)\right)^2 \qquad \text{(S.17)}\\
&\quad-\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}\omega_1(\tau;\mathbf{X}_j)\left(C_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)\right)\\
&\quad\times\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)\right)\\
&= \mathcal{T}_{3,R,P}^{(1),A}+\mathcal{T}_{3,R,P}^{(1),B}+\mathcal{T}_{3,R,P}^{(1),C}
\end{aligned}
$$

For notational simplicity and ease of reading, we focus again on the case of quantile regression, i.e. $\widehat{\boldsymbol{\psi}}_{1,R}(\tau)=\widehat{\boldsymbol{\beta}}_{1,R}(\tau)$, and $\widehat{q}_{\tau}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);X_{1,t}\right)=X_{1,t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)$, and divide this part of the proof into further sub-steps.

We start with $\mathcal{T}^{(1),A}_{3,R,P}$ from Equation (S.17). Thus:

$$\begin{aligned}
\mathcal{T}^{(1),A}_{3,R,P} &= \frac{1}{2\sqrt{P}}\sum_{j=R}^{T-1}\omega_1(\tau;\mathbf{X}_j)\left(C_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)\right)^2 \qquad \text{(S.18)}\\
&= \frac{1}{2\sqrt{P}}\sum_{j=R}^{T-1}\omega_1(\tau;\mathbf{X}_j)\left(F_{j+1}\left(X_{1,j}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\,|\mathbf{X}_j\right)-F_{j+1}\left(X_{1,j}'\boldsymbol{\beta}_{1,R}^{\dagger}(\tau)\,|\mathbf{X}_j\right)\right)^2\\
&= \frac{1}{2\sqrt{P}}\sum_{j=R}^{T-1}\omega_1(\tau;\mathbf{X}_j)\left(f_{j+1}\left(X_{1j}'\boldsymbol{\beta}_1^{\dagger}(\tau)\,|\mathbf{X}_j\right)X_{1j}'\left(\widehat{\boldsymbol{\beta}}_1(\tau)-\boldsymbol{\beta}_1^{\dagger}(\tau)\right)\right)^2\\
&= \frac{1}{2\sqrt{P}}\sum_{j=R}^{T-1}\underbrace{\omega_1(\tau;\mathbf{X}_j)\,f_{j+1}\left(X_{1j}'\boldsymbol{\beta}_1^{\dagger}(\tau)\,|\mathbf{X}_j\right)^2X_{1j}'}_{A}\underbrace{\left(\widehat{\boldsymbol{\beta}}_1(\tau)-\boldsymbol{\beta}_1^{\dagger}(\tau)\right)\left(\widehat{\boldsymbol{\beta}}_1(\tau)-\boldsymbol{\beta}_1^{\dagger}(\tau)\right)'}_{B}\underbrace{X_{1j}}_{C}
\end{aligned}$$

Now:

$$\begin{aligned}
\operatorname{vec}\left(\mathcal{T}^{(1),A}_{3,R,P}\right) &= \operatorname{vec}(ABC)=(C'\otimes A)\operatorname{vec}(B) \qquad \text{(S.19)}\\
&= \frac{1}{2P}\sum_{j=R}^{T-1}\left(X_{1j}'\otimes\left(\omega_1(\mathbf{X}_j,\tau)\,f\left(X_{1j}'\boldsymbol{\beta}_1^{\dagger}(\tau)\,|\mathbf{X}_j\right)^2X_{1j}'\right)\right)\\
&\quad\times\sqrt{P}\operatorname{vec}\left(\left(\widehat{\boldsymbol{\beta}}_1(\tau)-\boldsymbol{\beta}_1^{\dagger}(\tau)\right)\left(\widehat{\boldsymbol{\beta}}_1(\tau)-\boldsymbol{\beta}_1^{\dagger}(\tau)\right)'\right)\\
&= O_p(1)\times O_p\left(\frac{\sqrt{P}}{R}\right)=o_p(1),
\end{aligned}$$

where the first term follows from Assumptions A.1, A.2, A.4, and a law of large numbers for β-mixing observations, and the second part by Lemma Q(iii) as $\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau)-\boldsymbol{\beta}_1^{\dagger}(\tau)\right)=O_p\left(R^{-1/2}\right)$.

Next we turn to $\mathcal{T}^{(1),B}_{3,R,P}$ from Eq. (S.17), which we can decompose as follows:

$$\begin{aligned}
&\mathcal{T}^{(1),B}_{3,R,P}\\
&= \frac{2}{\sqrt{P}}\sum_{j=R}^{T-1}\frac{\omega_1(\tau;\mathbf{X}_j)}{\widehat{f}_{\mathbf{X}}(\mathbf{X}_j)^2}\\
&\quad\left(\frac{1}{Ph^d}\sum_{s=R}^{T-1}\left(1\left\{y_{s+1}\le X_{1,j}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\right\}-F_{j+1}\left(X_{1,j}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\,|\mathbf{X}_j\right)\right)\mathbf{K}\left(\frac{\mathbf{X}_j-\mathbf{X}_s}{h}\right)\right)^2\\
&= \frac{2}{\sqrt{P}}\sum_{j=R}^{T-1}\frac{\omega_1(\tau;\mathbf{X}_j)}{\widehat{f}_{\mathbf{X}}(\mathbf{X}_j)^2}\left(\frac{1}{Ph^d}\sum_{s=R}^{T-1}\left(1\left\{y_{s+1}\le X_{1,j}'\boldsymbol{\beta}_1^{\dagger}(\tau)\right\}-F_{j+1}\left(X_{1,j}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\,|\mathbf{X}_j\right)\right)\mathbf{K}\left(\frac{\mathbf{X}_j-\mathbf{X}_s}{h}\right)\right)^2\\
&\quad+\frac{2}{\sqrt{P}}\sum_{j=R}^{T-1}\frac{\omega_1(\tau;\mathbf{X}_j)}{\widehat{f}_{\mathbf{X}}(\mathbf{X}_j)^2}\left(\frac{1}{Ph^d}\sum_{s=R}^{T-1}\left(1\left\{y_{s+1}\le X_{1,j}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\right\}-1\left\{y_{s+1}\le X_{1,j}'\boldsymbol{\beta}_1^{\dagger}(\tau)\right\}\right)\mathbf{K}\left(\frac{\mathbf{X}_j-\mathbf{X}_s}{h}\right)\right)^2\\
&\quad+\frac{4}{\sqrt{P}}\sum_{j=R}^{T-1}\frac{\omega_1(\tau;\mathbf{X}_j)}{\widehat{f}_{\mathbf{X}}(\mathbf{X}_j)^2}\left(\frac{1}{Ph^d}\sum_{s=R}^{T-1}\left(1\left\{y_{s+1}\le X_{1,j}'\boldsymbol{\beta}_1^{\dagger}(\tau)\right\}-F_{j+1}\left(X_{1,j}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\,|\mathbf{X}_j\right)\right)\mathbf{K}\left(\frac{\mathbf{X}_j-\mathbf{X}_s}{h}\right)\right.\\
&\quad\left.\frac{1}{Ph^d}\sum_{s=R}^{T-1}\left(1\left\{y_{s+1}\le X_{1,j}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\right\}-1\left\{y_{s+1}\le X_{1,j}'\boldsymbol{\beta}_1^{\dagger}(\tau)\right\}\right)\mathbf{K}\left(\frac{\mathbf{X}_j-\mathbf{X}_s}{h}\right)\right)\\
&= \mathcal{T}^{(1),B1}_{3,R,P}+\mathcal{T}^{(1),B2}_{3,R,P}+\mathcal{T}^{(1),B3}_{3,R,P} \qquad \text{(S.20)}
\end{aligned}$$

Starting with $\mathcal{T}^{(1),B1}_{3,R,P}$ from (S.20), note that by Lemma Q.1(iii) and A.4 we have that:

$$\begin{aligned}
&\mathcal{T}^{(1),B1}_{3,R,P}\\
&= \frac{1}{2\sqrt{P}}\sum_{j=R}^{T-1}\frac{\omega_1(\tau;\mathbf{X}_j)}{\widehat{f}_{\mathbf{X}}(\mathbf{X}_j)^2}\left(\frac{1}{Ph^d}\sum_{s=R}^{T-1}\left(1\left\{y_{s+1}-X_{1,j}'\boldsymbol{\beta}_1^{\dagger}(\tau)\le 0\right\}-F_{j+1}\left(X_{1j}'\boldsymbol{\beta}_1^{\dagger}(\tau)\,|\mathbf{X}_j\right)\right)\mathbf{K}\left(\frac{\mathbf{X}_s-\mathbf{X}_j}{h}\right)\right)^2
\end{aligned}$$

$$\times\left(1+O_p\left(\frac{1}{R}\right)\right).$$

Note that the above expression is the sum over the conditioning variable of the squared difference between the estimated and the 'true' conditional cumulative distribution function. Given Assumption A.1, A.3, A.4, and $Ph^{\frac{3d}{2}}/\log P \to \infty$ as well as $Ph^{\frac{d}{2}+2r} \to 0$ (implied by A.5), similar arguments to the ones used in the proof of Lemma 1A-(i), part (a), in Corradi et al. (2020) yield that:

$$\sqrt{Ph^d}\mathcal{T}^{(1),B1}_{3,R,P} - h^{-d/2}B_1 \xrightarrow{d} N\left(0, V_{\boldsymbol{\beta}_1}\right)$$

where $V_{\boldsymbol{\beta}_1}$ is a positive variance, and the bias is $B_1 = C(\mathbf{K})\int_{\mathcal{X}} 1\{\mathbf{X}\in\mathcal{X}\}\mathrm{d}\mathbf{X} > 0$ for some kernel dependent constant $C(\mathbf{K})$.[3] Thus:

$$\mathcal{T}^{(1),B1}_{3,R,P} - P^{-1/2}h^{-d}B_1 = o_p(1)$$

Next, we turn to $\mathcal{T}^{(1),B2}_{3,R,P}$ from (S.20). Since $\|\widehat{\boldsymbol{\beta}}_{1,R}(\tau) - \boldsymbol{\beta}_1^\dagger(\tau)\| = O_p(R^{-1/2})$ by Lemma Q.1, we consider again a set $N_{\boldsymbol{\beta}_1,R}$ defined in analogy to $N_{\boldsymbol{\psi}_1,R}$. We now proceed by approximating $\mathcal{T}^{(1),B2}_{3,R,P}$ with a third order U-process indexed by $\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}$ noting that, on $\mathcal{X}$, $\widehat{f}_{\mathbf{X}}(\mathbf{X}_j)^2$ in the denominator can be replaced by $f_{\mathbf{X}}(\mathbf{X}_j)^2$ using A.5. That is, introducing some notation, let:

$$\begin{aligned}\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1) \quad \equiv \quad & \frac{\omega_1(\tau;\mathbf{X}_j)}{f_{\mathbf{X}}(\mathbf{X}_j)^2}\left(1\left\{y_{s+1}\leq X'_{1,j}\boldsymbol{\beta}_1\right\} - 1\left\{y_{s+1}\leq X'_{1,j}\boldsymbol{\beta}_1^\dagger(\tau)\right\}\right)\mathbf{K}\left(\frac{\mathbf{X}_j-\mathbf{X}_s}{h}\right)\\ & \times\left(1\left\{y_{l+1}\leq X'_{1,j}\boldsymbol{\beta}_1\right\} - 1\left\{y_{l+1}\leq X'_{1,j}\boldsymbol{\beta}_1^\dagger(\tau)\right\}\right)\mathbf{K}\left(\frac{\mathbf{X}_j-\mathbf{X}_l}{h}\right).\end{aligned}$$

Also, define the symmetric 'kernel function':

$$\overline{\boldsymbol{\Upsilon}}_{jls,P}(\boldsymbol{\beta}_1) = \frac{1}{3}\left(\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1) + \boldsymbol{\Upsilon}_{sjl,P}(\boldsymbol{\beta}_1) + \boldsymbol{\Upsilon}_{ljs,P}(\boldsymbol{\beta}_1)\right),$$

where we used the fact that $\boldsymbol{\Upsilon}_{jls,P}$ is already symmetric in the last two arguments. Thus:

$$\frac{2\sqrt{P}}{P^3h^{2d}}\sum_{j=R}^{T-1}\sum_{s\neq j}\sum_{l\neq j, l\neq s}\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1) = \frac{\sqrt{P}}{h^{2d}}\binom{P}{3}^{-1}\sum_{j=R}^{T-3}\sum_{s>j}\sum_{l>s}\overline{\boldsymbol{\Upsilon}}_{jls,P}(\boldsymbol{\beta}_1).$$

Now, writing $\mathcal{T}^{(1),B2}_{3,R,P}$ as $\mathcal{T}^{(1),B2}_{3,R,P}(\boldsymbol{\beta}_1)$, first observe that:

$$\begin{aligned}&\sup_{\boldsymbol{\beta}_1\in N_{\boldsymbol{\beta}_1,R}}\left|\mathcal{T}^{(1),B2}_{3,R,P}(\boldsymbol{\beta}_1) - \sqrt{P}\binom{P}{3}^{-1}\sum_{j=R}^{T-3}\sum_{s>j}\sum_{l>s}\frac{1}{h^{2d}}\left(\overline{\boldsymbol{\Upsilon}}_{jls,P}(\boldsymbol{\beta}_1) - \mathrm{E}\left(\overline{\boldsymbol{\Upsilon}}_{jls,P}(\boldsymbol{\beta}_1)\right)\right)\right|\\
= \quad & \sup_{\boldsymbol{\beta}_1\in N_{\boldsymbol{\beta}_1,R}}\left|\frac{2\sqrt{P}}{P^3h^{2d}}\sum_{j=R}^{T-1}\sum_{j\neq s}\boldsymbol{\Upsilon}_{jss,P}(\boldsymbol{\beta}_1) + \frac{2\sqrt{P}}{P^3h^{2d}}\sum_{j=R}^{T-1}\sum_{s\neq j}\boldsymbol{\Upsilon}_{jjs,P}(\boldsymbol{\beta}_1) + \frac{2\sqrt{P}}{P^3h^{2d}}\sum_{j=R}^{T-1}\sum_{l\neq j}\boldsymbol{\Upsilon}_{jsj,P}(\boldsymbol{\beta}_1)\right.\\
&\left.+\frac{2\sqrt{P}}{P^3h^{2d}}\sum_{j=R}^{T-1}\boldsymbol{\Upsilon}_{jjj,P}(\boldsymbol{\beta}_1) + \frac{6\sqrt{P}}{h^{2d}}\mathrm{E}\left(\overline{\boldsymbol{\Upsilon}}_{jls,P}(\boldsymbol{\beta}_1)\right)\right| \qquad\qquad \text{(S.21)}\\
\leq \quad & \sup_{\boldsymbol{\beta}_1\in N_{\boldsymbol{\beta}_1,R}}\left|\frac{2\sqrt{P}}{P^3h^{2d}}\sum_{j=R}^{T-1}\sum_{j\neq s}\boldsymbol{\Upsilon}_{jss,P}(\boldsymbol{\beta}_1)\right| + \sup_{\boldsymbol{\beta}_1\in N_{\boldsymbol{\beta}_1,R}}\left|\frac{2\sqrt{P}}{P^3h^{2d}}\sum_{j=R}^{T-1}\sum_{s\neq j}\boldsymbol{\Upsilon}_{jjs,P}(\boldsymbol{\beta}_1)\right|\\
&+\sup_{\boldsymbol{\beta}_1\in N_{\boldsymbol{\beta}_1,R}}\left|\frac{2\sqrt{P}}{P^3h^{2d}}\sum_{j=R}^{T-1}\sum_{l\neq j}\boldsymbol{\Upsilon}_{jsj,P}(\boldsymbol{\beta}_1)\right| + \sup_{\boldsymbol{\beta}_1\in N_{\boldsymbol{\beta}_1,R}}\left|\frac{2\sqrt{P}}{P^3h^{2d}}\sum_{j=R}^{T-1}\boldsymbol{\Upsilon}_{jjj,P}(\boldsymbol{\beta}_1)\right|\\
&+\sup_{\boldsymbol{\beta}_1\in N_{\boldsymbol{\beta}_1,R}}\left|\frac{6\sqrt{P}}{h^{2d}}\mathrm{E}\left(\overline{\boldsymbol{\Upsilon}}_{jls,P}(\boldsymbol{\beta}_1)\right)\right|\end{aligned}$$

[3]Corradi et al. (2020) assume a second order kernel in Lemma 1A. Notice, however, that by replacing "2" with "r" in their proof, it is immediate to see that, provided $Ph^{(2r+d/2)}\to 0$, the non vanishing bias component depends on the dimension of the covariate set, but not on the order of the kernel. Given A.5(i), the condition $Ph^{(2r+d/2)}\to 0$ is indeed satisfied.

We will postpone the treatment of the first four terms on the RHS of the inequality in (S.21) to the end as they can be analysed by similar arguments to the ones used subsequently, and just note that given A.5 with $\pi > 0$ their probability order is of order smaller than $\frac{\sqrt{P}}{R}$. We next proceed by analysing the last term on the RHS of (S.21) focusing on the first element $\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1)$ for simplicity. Firstly note that by iterated expectations:

$$
\begin{aligned}
& \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \mathrm{E}\left(\frac{1}{h^{2d}} \boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1) \right) \right| \\
= & \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \int\int\int \frac{1}{h^{2d}} \frac{\omega_1(\tau; \mathbf{X}_j)}{f_{\mathbf{X}}(\mathbf{X}_j)^2} \left(F_{s+1}(X_{1,j}'\boldsymbol{\beta}_1 | \mathbf{X}_s) - F_{s+1}(X_{1,j}'\boldsymbol{\beta}_1^{\dagger}(\tau) | \mathbf{X}_s) \right) \mathbf{K}\left(\frac{\mathbf{X}_j - \mathbf{X}_s}{h} \right) \right. \\
& \left. \times \left(F_{l+1}(X_{1,j}'\boldsymbol{\beta}_1 | \mathbf{X}_l) - F_{l+1}(X_{1,j}'\boldsymbol{\beta}_1^{\dagger}(\tau) | \mathbf{X}_l) \right) \mathbf{K}\left(\frac{\mathbf{X}_j - \mathbf{X}_l}{h} \right) f_{\mathbf{X}}(\mathbf{X}_l) f_{\mathbf{X}}(\mathbf{X}_s) f_{\mathbf{X}}(\mathbf{X}_j) \mathrm{d}\mathbf{X}_s \mathrm{d}\mathbf{X}_l \mathrm{d}\mathbf{X}_j \right|
\end{aligned}
$$

By standard change of variables and Taylor expansion arguments around $\mathbf{X}_j$, we obtain that:

$$
\begin{aligned}
& \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \int \omega_1(\tau; \mathbf{X}_j) \left(F_{j+1}(X_{1,j}'\boldsymbol{\beta}_1 | \mathbf{X}_j) - F_{j+1}(X_{1,j}'\boldsymbol{\beta}_1^{\dagger}(\tau) | \mathbf{X}_j) \right)^2 f_{\mathbf{X}}(\mathbf{X}_j) \mathrm{d}\mathbf{X}_j \right| + O(h^{2r}) \\
\leq & \; C \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left\| \boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^{\dagger}(\tau) \right\|^2 \int \omega_1(\tau; \mathbf{X}_j) \|X_{1,j}\|^2 f_{\mathbf{X}}(\mathbf{X}_j) \mathrm{d}\mathbf{X}_j + O(h^{2r}) = O(R^{-1}) + O(h^{2r}),
\end{aligned}
$$

where the $O(h^{2r})$ holds uniformly in $\boldsymbol{\beta}_1$ by A.3 and A.4, and the inequality follows from A.2, A.3(i), A.4. Thus, since $Ph^{2r} \to 0$ as $P \to \infty$ and $\pi > 0$ by A.5, we have that:

$$
\begin{aligned}
& \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \mathcal{T}_{3,R,P}^{(1),B2}(\boldsymbol{\beta}_1) - \sqrt{P} \binom{P}{3}^{-1} \sum_{j=R}^{T-1} \sum_{s>j} \sum_{l>s} \frac{1}{h^{2d}} \left(\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1) - \mathrm{E}\left(\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1) \right) \right) \right| \\
= & \; O_p\left(\frac{\sqrt{P}}{R} + \sqrt{P} h^{2r} \right) = O_p\left(\frac{\sqrt{P}}{R} \right),
\end{aligned}
$$

which implies:

$$
\begin{aligned}
& \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \mathcal{T}_{3,R,P}^{(1),B2}(\boldsymbol{\beta}_1) \right| - \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \sqrt{P} \binom{P}{3}^{-1} \sum_{j=R}^{T-3} \sum_{s>j} \sum_{l>s} \frac{1}{h^{2d}} \left(\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1) - \mathrm{E}\left(\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1) \right) \right) \right| \\
= & \; O_p\left(\frac{\sqrt{P}}{R} \right). \qquad \text{(S.22)}
\end{aligned}
$$

We are now ready to analyse the second term on the LHS of (S.22) as a third order U-process indexed by $\boldsymbol{\beta}_1$, where we will follow the notation in Arcones and Yu (1994). That is, set $f_{\boldsymbol{\beta}_1} = \boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1)$ and $\overline{f}_{\boldsymbol{\beta}_1} = \overline{\boldsymbol{\Upsilon}}_{jls,P}(\boldsymbol{\beta}_1)$, and define the class of functions:

$$
\overline{\mathcal{F}} = \left\{ \overline{f}_{\boldsymbol{\beta}_1} : \boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R} \right\}.
$$

We may write:

$$
U_P^3(\overline{f}_{\boldsymbol{\beta}_1}) \equiv \sqrt{P} \binom{P}{3}^{-1} \sum_{j=R}^{T-3} \sum_{s>j} \sum_{l>s} \frac{1}{h^{2d}} \left(\overline{\boldsymbol{\Upsilon}}_{jls,P}(\boldsymbol{\beta}_1) - \mathrm{E}\left(\overline{\boldsymbol{\Upsilon}}_{jls,P}(\boldsymbol{\beta}_1) \right) \right).
$$

Also, denote by $\pi_{k,3}\overline{f}_{\boldsymbol{\beta}_1}$, $k = 1, 2, 3$, the Pr canonical version (i.e., completely centered, see Arcones and Yu (1994, p.60)) of $\overline{f}_{\boldsymbol{\beta}_1}$, and note that by the Hoeffding projection:

$$
U_P^3(\overline{f}_{\boldsymbol{\beta}_1}) = \sum_{k=1}^{3} U_P^k(\pi_{k,3}\overline{f}_{\boldsymbol{\beta}_1})
$$

We start with the first order term $U_P^1(\pi_{1,3}\overline{f}_{\boldsymbol{\beta}_1})$ and establish its pointwise convergence. Focusing on the first element $\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1)$ of the symmetric kernel for illustration, lengthy, but standard calculations show that for each $\boldsymbol{\beta}_1 \in N_{\beta_1,R}$, the lead term of the variance is of order:

$$
\mathrm{E}\left(\mathrm{E}\left(\frac{1}{h^{2d}} \boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1) \left| (y_{j+1}, \mathbf{X}_j') \right. \right)^2 \right) = O\left(\frac{1}{R^2} + h^{4r} \right),
$$

while:

$$\mathrm{E}\left(\mathrm{E}\left(\frac{1}{h^{2d}}\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1)\,|(y_{s+1},\mathbf{X}_s')\right)^2\right)=\mathrm{E}\left(\mathrm{E}\left(\frac{1}{h^{2d}}\boldsymbol{\Upsilon}_{jls,P}(\boldsymbol{\beta}_1)\,|(y_{l+1},\mathbf{X}_l')\right)^2\right)=O\left(h^{-d+1}\left(\frac{1}{R^{\frac{3}{2}}}+\frac{1}{R^{\frac{1}{2}}}h^{2r}\right)\right)$$

Thus, let $\delta_P=\min\{h^{d-1}R^{3/2},R^{1/2}h^{-2r+d-1}\}$. From Theorem 1 of Yoshihara (1976), it follows that the term $\sqrt{P\delta_P}U_P^1(\pi_{1,3}\overline{f}_{\boldsymbol{\beta}_1})$ converges pointwise in $\boldsymbol{\beta}_1$ to a zero mean Gaussian r.v. provided that:

- There are constants M_0 and $p>2$ such that:

$$\mathrm{E}\left(\left|\overline{\boldsymbol{\Upsilon}}_{jls,P}(\boldsymbol{\beta}_1)\right|^p\right)<M_0,$$

- the β-mixing coefficients satisfy, for some $t>1$:

$$\beta(k)=O\left(k^{-tp/(p-2)}\right).$$

  as $k\to\infty$.

The first condition is indeed satisfied by A.2 and A.4, while the second condition holds by A.1 if for instance $\beta(k)=O\left(k^{-(2+\varepsilon+\eta)/\varepsilon}\right)$ for choices such as $\varepsilon=0.1$, $\eta=0.1$ setting $p=4$ and $t=2$. In addition, note that from Yoshihara (1976) it also follows that pointwise in $\boldsymbol{\beta}_1$:

$$\sqrt{P\delta_P}\sum_{k=2}^{3}U_P^k(\pi_{k,3}\overline{f}_{\boldsymbol{\beta}_1}))=o_p(1),$$

We now establish that the previous pointwise results also hold uniformly in $\boldsymbol{\beta}_1\in N_{\boldsymbol{\beta}_1,R}$. Thus, in a first step, we therefore need to show that the weak convergence of the first order term of the Hoeffding projection, i.e. $U_P^1(\pi_{1,3}\overline{f}_{\boldsymbol{\beta}_1}))$, holds in $l^\infty(\mathcal{F})$, the space of bounded functions $\mathcal{F}\mapsto\mathbb{R}$ (see below) equipped with the uniform norm, i.e. for all functionals ν_∞ on $\mathcal{F}$ such that:

$$\|\nu_\infty\|_{\mathcal{F}}\equiv\sup_{f\in\mathcal{F}}|\nu_\infty(f)|<\infty.$$

This however follows from Corollary 2.1 in Arcones and Yu (1994) if we can show that some finite constants a and b, $\mathcal{F}$ satisfies:

$$N(\epsilon,\mathcal{F},L_2(Q))\leq a\left(\frac{\|F\|_{L_2(Q)}}{\epsilon}\right)^b \tag{S.23}$$

for any $\epsilon>0$, where $N(\epsilon,\mathcal{F},L_2(Q))$ denotes the covering number, F the envelope function of $\mathcal{F}$, and $\left(QF^2\right)^{\frac{1}{2}}$ for some probability measure Q with $QF^2<\infty$. This in turn will follow if we show that $\mathcal{F}$ belongs to a VC subgraph class of function. Thus, note that $f_{\boldsymbol{\beta}_1}$, an element from the symmetrized $\overline{f}_{\boldsymbol{\beta}_1}$, is contained in the product of the classes:

$$\begin{aligned}
\mathcal{F}_1&=\left\{\mathbf{K}\left(\frac{\mathbf{X}_j-\mathbf{X}_l}{h}\right)\right\}\\
\mathcal{F}_2&=\left\{\mathbf{K}\left(\frac{\mathbf{X}_j-\mathbf{X}_s}{h}\right)\right\}\\
\mathcal{F}_3&=\left\{\frac{\nabla^{(2)}L(C_1(\boldsymbol{\beta}_1^\dagger(\tau);\mathbf{X}_j)-\tau)1\{\mathbf{X}_j\in\mathcal{X}\}}{f_{\mathbf{X}}(\mathbf{X}_j)^2}\frac{1}{h^{2d}}1\{\|\mathbf{X}_j-\mathbf{X}_l\|\leq 2h\}1\{\|\mathbf{X}_j-\mathbf{X}_s\|\leq 2h\}\right\}\\
\mathcal{F}_4&=\left\{1\left\{y_{l+1}\leq X_{1,j}'\boldsymbol{\beta}_1\right\}-1\left\{y_{l+1}\leq X_{1,j}'\boldsymbol{\beta}_1^\dagger(\tau)\right\},\ \boldsymbol{\beta}_1\in N_{\boldsymbol{\beta}_1,R}\right\}\\
\mathcal{F}_5&=\left\{1\left\{y_{s+1}\leq X_{1,j}'\boldsymbol{\beta}_1\right\}-1\left\{y_{s+1}\leq X_{1,j}'\boldsymbol{\beta}_1^\dagger(\tau)\right\},\ \boldsymbol{\beta}_1\in N_{\boldsymbol{\beta}_1,R}\right\}
\end{aligned}$$

with envelope function $F=\frac{C}{h^{2d}}1\{\|\mathbf{X}_j-\mathbf{X}_l\|\leq 2h\}1\{\|\mathbf{X}_j-\mathbf{X}_s\|\leq 2h\}\|\mathbf{K}\|_\infty^2$, where C is some positive constant that follows from assumption A.2, A.4, and the fact that $\tau\in\mathcal{T}$ is a compact subset of $(0,1)$. Now, let t, γ_1, and γ_2 be real numbers, and let $\boldsymbol{\delta}\in\mathbb{R}^{d_1}$, so that:

$$g(y_{l+1},X_{1,j};\gamma_1,\gamma_2,\boldsymbol{\delta})=t\gamma_1+y_{l+1}\gamma_2+X_{1,j}'\boldsymbol{\delta}$$

and define:

$$\mathcal{G}=\left\{g(y_{l+1},X_{1,j};\gamma_1,\gamma_2,\boldsymbol{\delta}):\ \gamma_1,\gamma_2\in\mathbb{R},\ \boldsymbol{\delta}\in\mathbb{R}^{d_1}\right\}.$$

Note that $\mathcal{G}$ is a d_1+2 dimensional vector space. By Lemma 2.4 in Pakes and Pollard (1989), the class of sets of the form $1\{g \geq r\}$ or $1\{g > r\}$ for some $g \in \mathcal{G}$ and $r \in \mathbb{R}$ is a VC class. Now, let:

$$f_4\left(y_{l+1}, X_{1,j}; \boldsymbol{\beta}_1\right) = 1\left\{y_{l+1} \leq X_{1,j}' \boldsymbol{\beta}_1\right\},$$

and note that for each $\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}$:

$$\begin{aligned}
\text{subgraph}\left(f_4\left(y_{l+1}, X_{1,j}; \boldsymbol{\beta}_1\right)\right) &= \left\{\left(y_{l+1}, X_{1,j}, t\right) : \ 0 < t < f_3\left(y_{l+1}, X_{1,j}; \boldsymbol{\beta}_1\right)\right\} \\
&= 1\{(X_{1,j}'\boldsymbol{\beta}_1 - y_{l+1}) \geq 0\}1\{t > 0\}1\{t \geq 1\}^c \\
&= 1\{g_1 \geq 0\}1\{g_2 > 0\}1\{g_3 \geq 1\}^c
\end{aligned}$$

for $g_i \in \mathcal{G}$, $i = 1,2,3$. The subgraph is therefore the intersection of three sets, two of which belong to a VC class, and one is a complement of a set belonging to a VC class. Deduce from Lemma 2.6.17 in Van der Vaart and Wellner (1996) that subgraph $\left(f_4\left(y_{l+1}, X_{1,j}; \boldsymbol{\beta}_1\right)\right)$ forms a VC subgraph class. Then, deduce from Lemma 2.6.18 in Van der Vaart and Wellner (1996) that $\mathcal{F}_4$ is a VC class whose covering numbers satisfy (S.23) (the same applies to $\mathcal{F}_5$). We therefore have by Lemma A.1 in Ghosal et al. (2000):

$$N\left(\epsilon \frac{C}{h^{2d}} \|1\{\|\mathbf{X}_j - \mathbf{X}_l\| \leq 2h\}1\{\|\mathbf{X}_j - \mathbf{X}_s\| \leq 2h\}\|_{L_2(Q)} \|\mathbf{K}\|_\infty^2, \mathcal{F}_1\mathcal{F}_2\mathcal{F}_3\mathcal{F}_4\mathcal{F}_5, L_2(Q)\right) \leq a\left(\frac{1}{\epsilon}\right)^b.$$

In addition, note that the arguments in Arcones and Yu (1994) together with Lemma 3.1 therein also imply that:

$$\|\sqrt{P\delta_P} U_P^k(\pi_{k,3}\overline{f}_{\boldsymbol{\beta}_1})\|_{\mathcal{F}} \xrightarrow{p} 0$$

for $k = 2,3$. Finally, note that in the case of a nonlinear location scale model, the same uniformity results can be obtained by using the results in Sancetta (2009), in particular Corollary 2.1 therein.

It remains to analyse the convergence rate of the first three terms on the RHS of the inequality in (S.21). Now note that for the first term with jss we have that:

$$\begin{aligned}
&\frac{2\sqrt{P}}{P^3 h^{2d}} \sum_{j=R}^{T-1} \sum_{s \neq j} \boldsymbol{\Upsilon}_{jss,P}(\boldsymbol{\beta}_1) \\
= \ &\frac{1}{\sqrt{P}h^d} \frac{2}{P^2 h^d} \sum_{j=R}^{T-1} \sum_{s \neq j} \frac{\omega_1\left(\tau; \mathbf{X}_j\right)}{f_{\mathbf{X}}(\mathbf{X}_j)^2} \left(1\left\{y_{s+1} \leq X_{1,j}'\boldsymbol{\beta}_1\right\} - 1\left\{y_{s+1} \leq X_{1,j}'\boldsymbol{\beta}_1^\dagger(\tau)\right\}\right)^2 \mathbf{K}\left(\frac{\mathbf{X}_j - \mathbf{X}_s}{h}\right)^2,
\end{aligned}$$

while for the second term after the inequality in (S.21) with jjs (the second term with jsj can be treated identically) is:

$$\begin{aligned}
&\frac{2\sqrt{P}}{P^3 h^{2d}} \sum_{j=R}^{T-1} \sum_{s \neq j} \boldsymbol{\Upsilon}_{jjs,P}(\boldsymbol{\beta}_1) \\
= \ &\frac{1}{\sqrt{P}h^d} \frac{2}{P^2 h^d} \sum_{j=R}^{T-1} \sum_{s \neq j} \frac{\omega_1\left(\tau; \mathbf{X}_j\right)}{f_{\mathbf{X}}(\mathbf{X}_j)^2} \left(1\left\{y_{s+1} \leq X_{1,j}'\boldsymbol{\beta}_1\right\} - 1\left\{y_{s+1} \leq X_{1,j}'\boldsymbol{\beta}_1^\dagger(\tau)\right\}\right) \mathbf{K}\left(\frac{\mathbf{X}_j - \mathbf{X}_s}{h}\right) \\
&\times \left(1\left\{y_{j+1} \leq X_{1,j}'\boldsymbol{\beta}_1\right\} - 1\left\{y_{j+1} \leq X_{1,j}'\boldsymbol{\beta}_1^\dagger(\tau)\right\}\right) \mathbf{K}(0).
\end{aligned}$$

We start with the term with jss, and define $\boldsymbol{\Upsilon}_{js,P}(\boldsymbol{\beta}_1) \equiv \boldsymbol{\Upsilon}_{jss,P}(\boldsymbol{\beta}_1)$ as well as the 'symmetric kernel':

$$\overline{\boldsymbol{\Upsilon}}_{js,P}(\boldsymbol{\beta}_1) = \frac{1}{2}\left(\boldsymbol{\Upsilon}_{js,P}(\boldsymbol{\beta}_1) + \boldsymbol{\Upsilon}_{sj,P}(\boldsymbol{\beta}_1)\right).$$

Now, similar to before this term can be bounded by:

$$\sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left|\frac{4}{\sqrt{P}h^{2d}} \mathrm{E}\left(\overline{\boldsymbol{\Upsilon}}_{js,P}(\boldsymbol{\beta}_1)\right)\right| + \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left|\frac{2}{\sqrt{P}h^d} \binom{P}{2}^{-1} \sum_{j=R}^{T-1} \sum_{s>j} \frac{1}{h^d}\left(\overline{\boldsymbol{\Upsilon}}_{js,P}(\boldsymbol{\beta}_1) - \mathrm{E}\left(\overline{\boldsymbol{\Upsilon}}_{js,P}(\boldsymbol{\beta}_1)\right)\right)\right| \quad \text{(S.24)}$$

We start with the bias term focusing on the element $\boldsymbol{\Upsilon}_{js,P}(\boldsymbol{\beta}_1)$, and note that similar calculations to before yield that:

$$\sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left|\mathrm{E}\left(\frac{1}{\sqrt{P}h^d} \boldsymbol{\Upsilon}_{js,P}(\boldsymbol{\beta}_1)\right)\right|$$

$$
\begin{aligned}
= & \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \left(\int \mathbf{K}(\mathbf{u})^2 \mathrm{d}\mathbf{u} \right) \int \frac{\omega_1(\tau;\mathbf{X}_j)}{\sqrt{P}h^d} \left(F_{j+1}(X'_{1,j}\boldsymbol{\beta}_1|\mathbf{X}_j) + F_{j+1}(X'_{1,j}\boldsymbol{\beta}_1^{\dagger}(\tau)|\mathbf{X}_j) \right. \right. \\
& \left. \left. -2F_{j+1}(\min\{X'_{1,j}\boldsymbol{\beta}_1^{\dagger}(\tau), X'_{1,j}\boldsymbol{\beta}_1\}|\mathbf{X}_j) \right) \mathrm{d}\mathbf{X}_j(1+O(h)) \right|
\end{aligned}
$$

We start with the case $\min\{X'_{1,j}\boldsymbol{\beta}_1^{\dagger}(\tau), X'_{1,j}\boldsymbol{\beta}_1\} = X'_{1,j}\boldsymbol{\beta}_1$. By A.4, mean value expansions around $\boldsymbol{\beta}_1^{\dagger}(\tau)$ yield:

$$
\begin{aligned}
& \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \left(\int \mathbf{K}(\mathbf{u})^2 \mathrm{d}\mathbf{u} \right) \int \frac{\omega_1(\tau;\mathbf{X}_j)}{\sqrt{P}h^d} \left(f_{j+1}(X'_{1,j}\overline{\boldsymbol{\beta}}_1|\mathbf{X}_j)X'_{1,j}\left(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^{\dagger}(\tau)\right) - 2f_{j+1}(X'_{1,j}\overline{\boldsymbol{\beta}}_1|\mathbf{X}_j) \right. \right. \\
& \left. \left. \times X'_{1,j}\left(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^{\dagger}(\tau)\right)\right) \mathrm{d}\mathbf{X}_j \right| \\
\leq & \frac{C}{\sqrt{P}h^d} \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left\| \boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^{\dagger}(\tau) \right\| \int \omega_1(\tau;\mathbf{X}_j) \|X_{1,j}\| \mathrm{d}\mathbf{X}_j = O\left(\frac{1}{\sqrt{PR}h^d} \right),
\end{aligned}
$$

where the inequality follows from A.2 and A.4, while in the case $\min\{X'_{1,j}\boldsymbol{\beta}_1^{\dagger}(\tau), X'_{1,j}\boldsymbol{\beta}_1\} = X'_{1,j}\boldsymbol{\beta}_1^{\dagger}(\tau)$ we obtain by similar arguments:

$$
\begin{aligned}
& \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \left(\int \mathbf{K}(\mathbf{u})^2 \mathrm{d}\mathbf{u} \right) \int \frac{\omega_1(\tau;\mathbf{X}_j)}{\sqrt{P}h^d} \left(f_{j+1}(X'_{1,j}\overline{\boldsymbol{\beta}}_1|\mathbf{X}_j)X'_{1,j}\left(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^{\dagger}(\tau)\right)\right) \mathrm{d}\mathbf{X}_j \right| \\
\leq & \frac{C}{\sqrt{P}h^d} \sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left\| \boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^{\dagger}(\tau) \right\| \int \omega_1(\tau;\mathbf{X}_j) \|X_{1,j}\| \mathrm{d}\mathbf{X}_j = O\left(\frac{1}{\sqrt{PR}h^d} \right).
\end{aligned}
$$

Turning to the second term of (S.24), one can show that for the lead terms of the variance it holds that:

$$
\mathrm{E}\left(\mathrm{E}\left(\frac{1}{h^d} \boldsymbol{\Upsilon}_{js,P}(\boldsymbol{\beta}_1) \,|(y_{s+1}, \mathbf{X}'_s) \right)^2 \right) = \mathrm{E}\left(\mathrm{E}\left(\frac{1}{h^d} \boldsymbol{\Upsilon}_{js,P}(\boldsymbol{\beta}_1) \,|(y_{j+1}, \mathbf{X}'_j) \right)^2 \right) = O\left(R^{-\frac{1}{2}} \right).
$$

Therefore, using arguments similar to before, we obtain that:

$$
\sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \frac{2}{\sqrt{P}h^d} \binom{P}{2}^{-1} \sum_{j=R}^{T-1} \sum_{s>j} \frac{1}{h^d} \left(\boldsymbol{\Upsilon}_{js,P}(\boldsymbol{\beta}_1) - \mathrm{E}\left(\boldsymbol{\Upsilon}_{js,P}(\boldsymbol{\beta}_1) \right) \right) \right| = O_p\left(\frac{1}{Ph^d R^{1/4}} \right)
$$

and so for all $\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}$ given $Ph^{2d} \to \infty$ and $\pi > 0$:

$$
\frac{2\sqrt{P}}{P^3 h^{2d}} \sum_{j=R}^{T-1} \sum_{s \neq j} \boldsymbol{\Upsilon}_{js,P}(\boldsymbol{\beta}_1) = o_p\left(\frac{\sqrt{P}}{R} \right)
$$

Turning to the third term with jsj after the inequality in (S.21):

$$
\sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \frac{2\sqrt{P}}{P^3 h^{2d}} \sum_{j=R}^{T-1} \sum_{l \neq j} \boldsymbol{\Upsilon}_{jsj,P}(\boldsymbol{\beta}_1) \right|,
$$

and using the fact that $\mathbf{K}(0)$ is of bounded variation, note that identical arguments to the ones used above can be applied to show that this expression is of probability order $o_p\left(\sqrt{P}/R\right)$ for all $\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}$. Finally, using A.1, A.2, A.3, A.4 and A.5, we also have that:

$$
\sup_{\boldsymbol{\beta}_1 \in N_{\boldsymbol{\beta}_1,R}} \left| \frac{2\sqrt{P}}{P^3 h^{2d}} \sum_{j=R}^{T-1} \boldsymbol{\Upsilon}_{jjj,P}(\boldsymbol{\beta}_1) \right| = o_p\left(\frac{\sqrt{P}}{R} \right).
$$

Thus, summarizing the previous steps, since $\widehat{\boldsymbol{\beta}}_{1,R}(\tau) \in N_{\boldsymbol{\beta}_1,R}$ with probability going to one as $P \to \infty$, we therefore obtain that:

$$
\begin{aligned}
\mathcal{T}_{3,R,P}^{(1),B2} &= o_p\left(\frac{\sqrt{P}}{R} \right) + O\left(\frac{\sqrt{P}}{R} \right) + O(\sqrt{P}h^{2r}) + O_p\left(\frac{1}{\sqrt{P\delta_P}} \right) \\
&= O_p\left(\frac{\sqrt{P}}{R} \right),
\end{aligned}
$$

where the last equality follows from A.5. Finally, using the Cauchy-Schwarz inequality and arguments from above, we also obtain that $\mathcal{T}_{3,R,P}^{(1),B3}$ from Eq. (S.20) is $O_p(\sqrt{\mathcal{T}_{3,R,P}^{(1),B1}\mathcal{T}_{3,R,P}^{(1),B2}})$.

Similarly, $\mathcal{T}_{3,R,P}^{(1),C}$ from Eq. (S.17) can also be shown to be of order $O_p(\sqrt{\mathcal{T}_{3,R,P}^{(1),A}\mathcal{T}_{3,R,P}^{(1),B}})$ through another application of Cauchy-Schwarz, which establishes that $\mathcal{T}_{3,R,P} = o_p(1)$.

Proof of Step 3: We now examine $\mathcal{T}_{4,R,P}$. As above, we consider only the contribution of model 1 to $\mathcal{T}_{4,R,P}$. It follows that

$$\begin{aligned}
\mathcal{T}_{4,R,P} &\leq \sup_{\mathbf{X}\in\mathcal{X}} 1\{\mathbf{X}\in\mathcal{X}\}\nabla^{(2)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger};\mathbf{X}\right)-\tau\right) \\
&\quad\times\sup_{\mathbf{X}\in\mathcal{X}}\left|\widehat{C}_{1,R,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}\right)-C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}\right)\right| \\
&\quad\times\frac{1}{3!\sqrt{P}}\sum_{j=R}^{T-1}\omega_1(\tau;\mathbf{X}_j)\left(\widehat{C}_{1,R,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)\right)^2 \\
&= o_p\left(\mathcal{T}_{3,R,P}\right).
\end{aligned}$$

(ii)-(a) We consider the scenario where $C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right) = C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_t\right) = \tau$ almost surely. To begin with, note that $\mathcal{T}_{1,P}=0$ with probability one. In addition, given Assumption A.2(ii)-(iii), it follows that $\mathcal{T}_{2,R,P}=0$ almost surely because $C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right) = C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_t\right) = \tau$ a.s..[4] Hence, given Assumption A.2(iv), we obtain:

$$\widehat{S}_{P,R}(\tau) = \mathcal{T}_{3,R,P}(1) + o_p\left(\mathcal{T}_{3,R,P}\right),$$

since as shown in part (i), $\mathcal{T}_{4,R,P} = o_p\left(\mathcal{T}_{3,R,P}\right)$.

Next, observe that $\mathcal{T}_{3,R,P} = \mathcal{T}_{3,R,P}^{(1)} - \mathcal{T}_{3,R,P}^{(2)}$, where $\mathcal{T}_{3,R,P}^{(1)}$ is defined above in Eq. (S.17), and $\mathcal{T}_{3,R,P}^{(2)}$ is the analogous term for model 2. Also, as above we can decompose, $\mathcal{T}_{3,R,P}^{(1)} = \mathcal{T}_{3,R,P}^{(1),A} + \mathcal{T}_{3,R,P}^{(1),B} + \mathcal{T}_{3,R,P}^{(1),C}$ and an identical decomposition holds for model 2. Note that $\mathcal{T}_{3,R,P}^{(1),A} - \mathcal{T}_{3,R,P}^{(2),A} = \mathrm{vec}\left(\mathcal{T}_{3,R,P}^{(1),A}\right) - \mathrm{vec}\left(\mathcal{T}_{3,R,P}^{(2),A}\right)$, so that

$$\begin{aligned}
&\mathrm{Var}\left(\mathcal{T}_{3,R,P}^{(1),A} - \mathcal{T}_{3,R,P}^{(2),A}\right) \\
=\ &\mathrm{Var}\left(\mathrm{vec}\left(\mathcal{T}_{3,R,P}^{(1),A}\right)\right) + \mathrm{Var}\left(\mathrm{vec}\left(\mathcal{T}_{3,R,P}^{(2),A}\right)\right) - 2\mathrm{Cov}\left(\mathrm{vec}\left(\mathcal{T}_{3,R,P}^{(1),A}\right),\mathrm{vec}\left(\mathcal{T}_{3,R,P}^{(2),A}\right)\right).
\end{aligned}$$

From (S.19), it can be seen that:

$$\underline{c}\frac{P}{R^2} \leq \mathrm{Var}\left(\mathcal{T}_{3,R,P}^{(1),A} - \mathcal{T}_{3,R,P}^{(2),A}\right) \leq \overline{c}\frac{P}{R^2},$$

for some constants $0<\underline{c}\leq\overline{c}<\infty$, since for each variance term, $i=1,2$, $R\times\mathrm{vec}\left(\widehat{\boldsymbol{\beta}}_{i,R}(\tau)-\boldsymbol{\beta}_i^{\dagger}(\tau)\right)\left(\widehat{\boldsymbol{\beta}}_{i,R}(\tau)-\boldsymbol{\beta}_i^{\dagger}(\tau)\right)'$ has a non-degenerate limiting distribution with finite variance, and

$$\frac{1}{2P}\sum_{j=R}^{T-1}\left(X_{ij}'\otimes\left(\omega_1(\mathbf{X}_j,\tau)f\left(X_{ij}'\boldsymbol{\beta}_1^{\dagger}(\tau)|\mathbf{X}_j\right)^2X_{ij}'\right)\right)$$

has a well defined, finite probability limit, while the covariance term is of order $O_p(P/R^2)$.

Thus, we turn to $\mathcal{T}_{3,R,P}^{(1),B} - \mathcal{T}_{3,R,P}^{(2),B}$ directly. Formally, we can expand $\mathcal{T}_{3,R,P}^{(1),B}$ and $\mathcal{T}_{3,R,P}^{(2),B}$ as in the non-overlapping case (i) in Eq. (S.20):

$$\mathcal{T}_{3,R,P}^{(1),B} - \mathcal{T}_{3,R,P}^{(2),B} = \left(\mathcal{T}_{3,R,P}^{(1),B1} - \mathcal{T}_{3,R,P}^{(2),B1}\right) + \left(\mathcal{T}_{3,R,P}^{(1),B2} - \mathcal{T}_{3,R,P}^{(2),B2}\right) + \left(\mathcal{T}_{3,R,P3}^{(1),B3} - \mathcal{T}_{3,R,P}^{(2),B}\right),$$

Starting with the first term on the RHS, firstly note that now:

$$\left(\mathcal{T}_{3,R,P}^{(1),B1} - \mathcal{T}_{3,R,P}^{(2),B1}\right) = O_p\left(\frac{\sqrt{P}}{R}\right)$$

[4]The same also happens in the two-sided case, when both models are misspecified, but the coverage probability just happens to equal to the difference of the nominal levels $\tau_U - \tau_L$.

since $F_{j+1}\left(X_{1,j}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\,|\mathbf{X}_j\right)$ and $F_{j+1}\left(X_{1,j}'\widehat{\boldsymbol{\beta}}_{2,R}(\tau)\,|\mathbf{X}_j\right)$ can be expanded around $\boldsymbol{\beta}_1^{\dagger}(\tau)$ and $\boldsymbol{\beta}_2^{\dagger}(\tau)$, respectively, but the remaining expression is zero almost surely. On the contrary, the same arguments from CASE I yield that:

$$\left(\mathcal{T}_{3,R,P}^{(1),B2}-\mathcal{T}_{3,R,P}^{(2),B2}\right)=O_p\left(\frac{\sqrt{P}}{R}\right),$$

and:

$$\left(\mathcal{T}_{3,R,P}^{(1),B3}-\mathcal{T}_{3,R,P}^{(2),B3}\right)=o_p\left(\frac{\sqrt{P}}{R}\right).$$

Thus, $\mathcal{T}_{3,R,P}^{(1),B}-\mathcal{T}_{3,R,P}^{(2),B}=O_p\left(\frac{\sqrt{P}}{R}\right)$. Finally, by the Cauchy-Schwarz inequality, it follows that $\mathcal{T}_{3,R,P}^{(1),C}-\mathcal{T}_{3,R,P}^{(2),C}$ is at most of probability order $\frac{\sqrt{P}}{R}$. The statement for CASE II with $C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)=C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_t\right)=\tau$ a.s. then follows.

(ii)-(b): We next outline the differences when (conditional) coverage is the same almost surely, but both models are misspecified. For simplicity, we focus again on the one sided case with $\Pr\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)=\tau\right)<1$, which arises when $q_\tau(\boldsymbol{\psi}_1^{\dagger}(\tau);X_{1,t})=q_\tau(\boldsymbol{\psi}_2^{\dagger}(\tau);X_{1,t})$ almost surely, but the models are actually misspecified and so conditional coverage is not equal to the nominal level with probability one.

Thus, when models are misspecified and overlapping such that $q_\tau(\boldsymbol{\psi}_1^{\dagger}(\tau);X_{1,t})=q_\tau(\boldsymbol{\psi}_2^{\dagger}(\tau);X_{1,t})$ almost surely, observe that the arguments remain the same as in part (ii)-(a), but for the terms defined in (S.8):

$$\mathcal{T}_{2,R,P}=\mathcal{T}_{2,R,P}(A)+\mathcal{T}_{2,R,P}(B).$$

Now, as for $\mathcal{T}_{2,R,P}(B)$, when $\mathbf{X}_t$ only contains irrelevant predictors that do not feature into the DGP, observe that:

$$\mathcal{T}_{2,R,P}(B)=0+o_p(1)$$

since:

$$\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)=\nabla^{(1)}L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)=\nabla^{(1)}L\left(C-\tau\right)=0$$

almost surely, for some constant $0\le C\le 1$. More generally, when $q_\tau(\boldsymbol{\psi}_1^{\dagger}(\tau);X_{1,t})=q_\tau(\boldsymbol{\psi}_2^{\dagger}(\tau);X_{1,t})$ almost surely, but the conditional coverage errors are not constant with probability one, we obtain that:

$$\begin{aligned}
&\mathcal{T}_{2,R,P}(B)\\
=&\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}1\{\mathbf{X}_j\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_j\right)-\tau\right)\left(C_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_j\right)-C_2\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_j\right)\right)\\
=&\left[\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left\{1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)C_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_t\right)\right.\right.\\
&\left.-\mathrm{E}\left(1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)C_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_t\right)\right)\right\}\\
&-\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left\{1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)C_2\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_t\right)\right.\\
&\left.\left.-\mathrm{E}\left(1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)C_2\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_t\right)\right)\right\}\right]\\
&+\left[\sqrt{P}\mathrm{E}\left(\frac{1}{P}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)C_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_t\right)\right)\right.\\
&\left.-\sqrt{P}\mathrm{E}\left(\frac{1}{P}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)C_2\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_t\right)\right)\right]\\
=&\mathcal{T}_{2,R,P}(B_1)+\mathcal{T}_{2,R,P}(B_2),
\end{aligned}$$

where we used the fact that $\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)=\nabla^{(1)}L\left(C_2\left(\boldsymbol{\psi}_2^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)$ with probability one. Defining $N_{\boldsymbol{\psi},R}$ as in part (i), and noting that $\boldsymbol{\psi}_1^{\dagger}(\tau)=\boldsymbol{\psi}_2^{\dagger}(\tau)$ element-wise as well as $\widehat{\boldsymbol{\psi}}_{1,R}(\tau),\widehat{\boldsymbol{\psi}}_{2,R}(\tau)\in N_{\boldsymbol{\psi},R}$

with probability approaching one as $P \to \infty$, similar arguments to the proof of part (i) using again Theorem 2.2 in Andrews and Pollard (1994) together with Assumptions A.1 to A.4 yield that the empirical process:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\Big\{1\{\mathbf{X}_t \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)C_j\left(\boldsymbol{\psi}_j;\mathbf{X}_t\right)$$
$$-\mathrm{E}\left(1\{\mathbf{X}_t \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)C_j\left(\boldsymbol{\psi}_j;\mathbf{X}_t\right)\right)\Big\}, \quad j=1,2$$

is stochastically equicontinuous in $\boldsymbol{\psi}_1, \boldsymbol{\psi}_2 \in N_{\boldsymbol{\psi},R}$. As a result, since $\|\widehat{\boldsymbol{\psi}}_{1,R}(\tau)-\widehat{\boldsymbol{\psi}}_{2,R}(\tau)\| = o_p(1)$, we have that:

$$\mathcal{T}_{2,R,P}(B_1) = o_p(1).$$

As for $\mathcal{T}_{2,R,P}(B_2)$, note that by the same mean value expansion as in part (i):

$$\mathcal{T}_{2,R,P}(B_2) = \sqrt{P}\mathrm{E}\left(\frac{1}{P}\sum_{t=R}^{T-1}1\{\mathbf{X}_t \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)f_{t+1}(X_{1,t}'\boldsymbol{\beta}_1^\dagger(\tau)|\mathbf{X}_t)\right.$$
$$\left.\times\left(\left(X_{1,t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)-X_{1,t}'\boldsymbol{\beta}_1^\dagger(\tau)\right)-\left(X_{2,t}'\widehat{\boldsymbol{\beta}}_{2,R}(\tau)-X_{2,t}'\boldsymbol{\beta}_2^\dagger(\tau)\right)\right)\right)(1+o(1)),$$

where we used the fact that $X_{1,t}'\boldsymbol{\beta}_1^\dagger(\tau) = X_{2,t}'\boldsymbol{\beta}_2^\dagger(\tau)$ almost surely. Focusing on the part involving model 1 and recalling the definition of $\Lambda_1(\tau)$ in Equation (S.16), note that:

$$\sqrt{P}\mathrm{E}\left(\frac{1}{P}\sum_{t=R}^{T-1}1\{\mathbf{X}_t \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)f_{t+1}(X_{1,t}'\boldsymbol{\beta}_1^\dagger(\tau)|\mathbf{X}_t)\left(X_{1,t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)-X_{1,t}'\boldsymbol{\beta}_1^\dagger(\tau)\right)\right)$$
$$=\mathrm{E}\left(\left(\frac{1}{P}\sum_{t=R}^{T-1}1\{\mathbf{X}_t \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)f_{t+1}(X_{1,t}'\boldsymbol{\beta}_1^\dagger(\tau)|\mathbf{X}_t)X_{1,t}'-\Lambda_1(\tau)\right)\times\sqrt{P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau)-\boldsymbol{\beta}_1^\dagger(\tau)\right)\right)$$
$$+\Lambda_1(\tau)\times\mathrm{E}\left(\sqrt{P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau)-\boldsymbol{\beta}_1^\dagger(\tau)\right)\right)$$

As for the first term note that by Cauchy-Schwarz's inequality:

$$\left\|\mathrm{E}\left(\left(\frac{1}{P}\sum_{t=R}^{T-1}1\{\mathbf{X}_t \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)f_{t+1}(X_{1,t}'\boldsymbol{\beta}_1^\dagger(\tau)|\mathbf{X}_t)X_{1,t}'-\Lambda_1(\tau)\right)\sqrt{P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau)-\boldsymbol{\beta}_1^\dagger(\tau)\right)\right)\right\|$$
$$\leq\mathrm{E}\left(\left\|\frac{1}{P}\sum_{t=R}^{T-1}\left(1\{\mathbf{X}_t \in \mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_t\right)-\tau\right)f_{t+1}(X_{1,t}'\boldsymbol{\beta}_1^\dagger(\tau)|\mathbf{X}_t)X_{1,t}'-\Lambda_1(\tau)\right)\right\|^2\right)^{\frac{1}{2}}$$
$$\times\mathrm{E}\left(P\left\|\widehat{\boldsymbol{\beta}}_{1,R}(\tau)-\boldsymbol{\beta}_1^\dagger(\tau)\right\|^2\right)^{\frac{1}{2}}$$
$$=O\left(\frac{1}{\sqrt{P}}\right)O(1) = o(1).$$

where the last line follows from A.1, A.2, A.3, and A.5 as well as Lemma Q.1(iii). For the first term on the other hand we have that:

$$\Lambda_1(\tau)\times\mathrm{E}\left(\sqrt{P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau)-\boldsymbol{\beta}_1^\dagger(\tau)\right)\right) = o(1),$$

which follows as an implication of Lemma Q.3(iii) since $\sqrt{R}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau)-\boldsymbol{\beta}_1^\dagger(\tau)\right)$ converges weakly to a standard normal random vector with mean zero and since $P/R \to \pi$ with $0<\pi<\infty$ by A.5.[5]

(iii): This follows from the arguments of part (i) noting that under the alternative hypothesis:

$$\frac{1}{\sqrt{P}}\sum_{j=R}^{T-1}\mathrm{E}\left(1\{\mathbf{X}_j \in \mathcal{X}\}\left(L\left(C_1\left(\boldsymbol{\psi}_1^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)-L\left(C_2\left(\boldsymbol{\psi}_2^\dagger(\tau);\mathbf{X}_j\right)-\tau\right)\right)\right)$$

[5]In fact, in the i.i.d. case, the rate of $\mathrm{E}\left(\sqrt{P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau)-\boldsymbol{\beta}_1^\dagger(\tau)\right)\right)$ is given by $O(\sqrt{P}\ln(R)/R^{-\frac{3}{4}})$, see Bahadur (1966) or more recently Lee et al. (2018) and Franguridi et al. (2022).

diverges to plus or minus infinity. ■

Proof of Theorem 2:

(i) We start with CASE I. Moreover, we will discuss the linear quantile regression model case only. From Theorem 1 (i), recall that under H_0:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(L\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right)-\tau\right)-L\left(\widehat{C}_{2,P}\left(\widehat{\boldsymbol{\beta}}_{2,R}(\tau);\mathbf{X}_t\right)-\tau\right)\right) \tag{S.25}$$

$$\begin{aligned}
=\ & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}L\left(C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)-\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}L\left(C_2\left(\boldsymbol{\beta}_2^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)\\
&+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)\left(1\{y_{t+1}\le q_\tau(\boldsymbol{\beta}_1^{\dagger}(\tau);X_{1,t})\}-F_{t+1}(q_\tau(\boldsymbol{\beta}_1^{\dagger}(\tau);X_{1,t})|\mathbf{X}_{1,t})\right)\\
&-\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_2\left(\boldsymbol{\beta}_2^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)\left(1\{y_{t+1}\le q_\tau(\boldsymbol{\beta}_2^{\dagger}(\tau);X_{2,t})\}-F_{t+1}(q_\tau(\boldsymbol{\beta}_2^{\dagger}(\tau);X_{2,t})|\mathbf{X}_{2,t})\right)\\
&+\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}\varphi(\boldsymbol{\beta}_1^{\dagger}(\tau);y_{t+1},X_{1,t})-\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}\varphi(\boldsymbol{\beta}_2^{\dagger}(\tau);y_{t+1},X_{2,t})+o_p(1)\\
=\ & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(A_{1,t}(\tau)-A_{2,t}(\tau))+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(B_{1,t}(\tau)-B_{2,t}(\tau))+\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}(D_{1,t}(\tau)-D_{1,t}(\tau))+o_p(1),
\end{aligned}$$

where $A_{j,t}(\tau)$, $B_{j,t}(\tau)$, and $D_{j,t}(\tau)$, $j=1,2$, have been defined in the proof of Theorem 1. In what follows, let $\mathrm{E}\left(\cdot\mid\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right)$ and $\mathrm{Var}\left(\cdot\mid\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right)$ denote the expectation and variance, respectively, conditional on the original sample $\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T$. Thus, by noting that, conditional on $\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T$, $\widehat{S}^*_{P,R}(\tau)$ is distributed as:

$$N\left(\mathrm{E}\left(\widehat{S}^*_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right),\mathrm{Var}\left(\widehat{S}^*_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right)\right)$$

for all samples but a subset with probability measure approaching zero, it suffices to show that

$$\mathrm{p}\lim_{\mathrm{R,P}\to\infty}\mathrm{E}\left(\widehat{S}^*_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right)=0$$

and:

$$\begin{aligned}
& \mathrm{p}\lim_{\mathrm{R,P}\to\infty}\mathrm{Var}\left(\widehat{S}^*_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right)\\
=\ & \mathrm{p}\lim_{\mathrm{P,R}\to\infty}\mathrm{Var}\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(A_{1,t}(\tau)-A_{2,t}(\tau))+\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}(B_{1,t}(\tau)-B_{2,t}(\tau))+\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}(D_{1,t}(\tau)-D_{1,t}(\tau))\right)\\
=\ & \Omega(\tau)
\end{aligned}$$

with $\Omega(\tau)$ defined as in Eq.(S.4). Now, by the definition of ε_t and η_t:

$$\mathrm{E}\left(\widehat{S}^*_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right)=0$$

for almost all samples. On the other hand, setting $l_P=l_R=l$ for notational simplicity, we obtain:

$$\begin{aligned}
& \mathrm{Var}\left(\widehat{S}^*_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_i'\}_{t=1}^T\right)=\mathrm{E}^*\left(\widehat{S}^{*2}_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_i'\}_{t=1}^T\right)\\
=\ & \frac{1}{P}\sum_{t=R}^{T-l-1}\frac{1}{l}\left(\sum_{i=t}^{t+l}\left(\widehat{A}_{1,P,R,i}(\tau)-\widehat{A}_{2,P,R,i}(\tau)\right)\right)^2+\frac{1}{P}\sum_{t=R}^{T-l-1}\frac{1}{l}\left(\sum_{i=t}^{t+l}\left(\widehat{B}_{1,P,R,i}(\tau)-\widehat{B}_{2,P,R,i}(\tau)\right)\right)^2\\
&+\frac{2}{P}\sum_{t=R}^{T-l-1}\frac{1}{l}\left(\sum_{i=t}^{t+l}\left(\widehat{A}_{1,P,R,i}(\tau)-\widehat{A}_{2,P,R,i}(\tau)\right)\right)\left(\sum_{i=t}^{t+l}\left(\widehat{B}_{1,P,R,i}(\tau)-\widehat{B}_{2,P,R,i}(\tau)\right)\right)\\
&+\frac{P}{R^2}\sum_{t=1}^{R-l-1}\frac{1}{l}\left(\sum_{i=t}^{t+l}\left(\widehat{D}_{1,P,R,i}(\tau)-\widehat{D}_{2,P,R,i}(\tau)\right)\right)^2
\end{aligned}$$

Simple arithmetic shows that:

$$\frac{1}{P}\sum_{t=R}^{T-l-1}\frac{1}{l}\left(\sum_{i=t}^{t+l}\left(\widehat{A}_{1,P,R,i}(\tau)-\widehat{A}_{2,P,R,i}(\tau)\right)\right)^2$$

$$
\begin{aligned}
= & \frac{1}{P}\sum_{t=R}^{T-l-1}\left(\widehat{A}_{1,P,R,t}(\tau)-\widehat{A}_{2,P,R,t}(\tau)\right)^2\left(1+O_p\left(\frac{l}{P}\right)\right)\\
& +\frac{1}{P}\sum_{t=R}^{T-l-1}\sum_{j=1}^{l}\left(1-\frac{j}{l}\right)\left(\widehat{A}_{1,P,R,t}(\tau)-\widehat{A}_{2,P,R,t}(\tau)\right)\left(\widehat{A}_{1,P,R,t+j}(\tau)-\widehat{A}_{2,P,R,t+j}(\tau)\right)\left(1+O_p\left(\frac{l}{P}\right)\right)
\end{aligned}
$$

where $O_p\left(\frac{l}{P}\right)$ term is due to the fact the the first and last l observations have a smaller contribution (as in the block bootstrap). Thus, letting $\varpi_{j,l}=\left(1-\frac{j}{l}\right)$:

$$
\begin{aligned}
& \operatorname{Var}\left(\widehat{S}^*_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_i'\}_{t=1}^T\right)\\
& =\left(\frac{1}{P}\sum_{t=R}^{T-l-1}\left(\left(\widehat{A}_{1,P,R,t}(\tau)-\widehat{A}_{2,P,R,t}(\tau)\right)^2+2\sum_{j=1}^{l}\varpi_{j,l}\left(\widehat{A}_{1,P,R,t}(\tau)-\widehat{A}_{2,P,R,t}(\tau)\right)\left(\widehat{A}_{1,P,R,t+j}(\tau)-\widehat{A}_{2,P,R,t+j}(\tau)\right)\right)\right.\\
& +\frac{1}{P}\sum_{t=R}^{T-l-1}\left(\left(\widehat{B}_{1,P,R,t}(\tau)-\widehat{B}_{2,P,R,t}(\tau)\right)^2+2\sum_{j=1}^{l}\varpi_{j,l}\left(\widehat{B}_{1,P,R,t}(\tau)-\widehat{B}_{2,P,R,t}(\tau)\right)\left(\widehat{B}_{1,P,R,t+j}(\tau)-\widehat{B}_{2,P,R,t+j}(\tau)\right)\right)\\
& +\frac{2}{P}\sum_{t=R}^{T-l-1}\left(\left(\widehat{A}_{1,P,R,t}(\tau)-\widehat{A}_{2,P,R,t}(\tau)\right)\left(\widehat{B}_{1,P,R,t}(\tau)-\widehat{B}_{2,P,R,t}(\tau)\right)\right.\\
& \left.+2\sum_{j=1}^{l}\varpi_{j,l}\left(\widehat{A}_{1,P,R,t}(\tau)-\widehat{A}_{2,P,R,t}(\tau)\right)\left(\widehat{B}_{1,P,R,t+j}(\tau)-\widehat{B}_{2,P,R,t+j}(\tau)\right)\right)\\
& \left.+\frac{P}{R^2}\sum_{t=1}^{R-l-1}\left(\left(\widehat{D}_{1,P,R,t}(\tau)-\widehat{D}_{2,P,R,t}(\tau)\right)^2+2\sum_{j=1}^{l}\varpi_{j,l}\left(\widehat{D}_{1,P,R,t}(\tau)-\widehat{D}_{2,P,R,t}(\tau)\right)\left(\widehat{D}_{1,P,R,t+j}(\tau)-\widehat{D}_{2,P,R,t+j}(\tau)\right)\right)\right)\\
& \times\left(1+O_p\left(\frac{l}{P}\right)\right)\\
& =\left(\widehat{V}_{11,P,R}+\widehat{V}_{22,P,R}+\widehat{V}_{33,P,R}+\widehat{V}_{12,P,R}\right)\left(1+O_p\left(\frac{l}{P}\right)\right),
\end{aligned}
$$

where, for $j=1,2$, $\widehat{A}_{j,P,R,i}(\tau)$, $\widehat{B}_{j,P,R,i}(\tau)$, and $\widehat{D}_{j,P,R,i}(\tau)$ have been defined in (29), (30), and (31) of the main text, respectively. Let $\widetilde{V}_{11,P,R}(\tau)$, $\widetilde{V}_{22,P,R}(\tau)$, $\widetilde{V}_{12,P,R}(\tau)$, $\widetilde{V}_{33,P,R}(\tau)$ be defined as $\widehat{V}_{11,P,R}(\tau)$, $\widehat{V}_{22,P,R}(\tau)$, $\widehat{V}_{12,P,R}(\tau)$, $\widehat{V}_{33,P,R}(\tau)$ with $\widehat{A}_{j,P,R,t}(\tau)$, $\widehat{B}_{j,P,R,t}(\tau)$, $\widehat{D}_{j,P,R,t}(\tau)$ replaced by $A_{j,t}(\tau)$, $B_{j,t}(\tau)$, $D_{j,t}(\tau)$ for $j=1,2$. By similar arguments as in the proof of Theorem 1(i), for all $l,k=\{(1,1),(2,2),(1,2),(3,3)\}$, pointwise in τ we have that $\widetilde{V}_{lk,P,R}(\tau)=\widetilde{V}_{lk}(\tau)+o_p(1)$, and by Theorem 1 (a) in Andrews (1991):

$$
\left(\widetilde{V}_{11,P,R}(\tau)+\widetilde{V}_{22,P,R}(\tau)+\widetilde{V}_{12,P,R}(\tau)+\widetilde{V}_{33,P,R}(\tau)\right)=\Omega(\tau)+o_p(1).
$$

Thus:

$$
\begin{aligned}
& \operatorname{Var}\left(\widehat{S}^*_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right)\\
= & \left(\widetilde{V}_{11,P,R}(\tau)+\widetilde{V}_{22,P,R}(\tau)+\widetilde{V}_{12,P,R}(\tau)+\widetilde{V}_{33,P,R}(\tau)\right)+o_p(1)=\Omega(\tau)+o_p(1).
\end{aligned}
$$

This establishes the statement in (i) for CASE I.

(ii) We start with the case where both models are correctly specified, i.e. $C_1\left(\boldsymbol{\beta}_1^\dagger(\tau);\mathbf{X}_t\right)=C_2\left(\boldsymbol{\beta}_2^\dagger(\tau);\mathbf{X}_t\right)=\tau$ with probability one. As in proof of Theorem 1 (ii), recall that when $C_1\left(\boldsymbol{\beta}_1^\dagger(\tau);\mathbf{X}_t\right)=C_2\left(\boldsymbol{\beta}_2^\dagger(\tau);\mathbf{X}_t\right)=\tau$ a.s., it holds that $L(0)=\nabla^{(1)}L(0)=0$. Moreover, by Assumption A.2(iii) note that:

$$
\frac{1}{P}\sum_{t=R}^{T-l-1}\left(\widehat{A}_{1,P,R,t}(\tau)-\widehat{A}_{2,P,R,t}(\tau)\right)^2=\frac{1}{P}\sum_{t=R}^{T-l-1}\left(\widehat{A}_{1,P,R,t}(\tau)^2+\widehat{A}_{2,P,R,t}(\tau)^2-2\widehat{A}_{1,P,R,t}(\tau)\widehat{A}_{2,P,R,t}(\tau)\right).
$$

We will only focus on the term involving $\widehat{A}_{1,P,R,t}(\tau)^2$, the other term will follow from identical arguments and the cross-term can be handled via the Cauchy-Schwarz by analogous arguments. By a second order Taylor expansion around $C_1\left(\boldsymbol{\beta}_1^\dagger(\tau);\mathbf{X}_t\right)$:

$$
\frac{1}{P}\sum_{t=R}^{T-l-1}\widehat{A}_{1,P,R,t}(\tau)^2
$$

$$= \frac{1}{2P}\sum_{t=R}^{T-l-1} 1\{\mathbf{X}_t \in \mathcal{X}\}\left(\nabla^{(2)} L(0)\right)^2 \left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right) - C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)\right)^4 + o_p\left(\frac{1}{R^2}\right)$$
$$= O_p\left(\frac{1}{R^2}\right)(1+o_p(1)),$$

where the last line follow from the arguments in Theorem 1(i), noting that this time the term involving $\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right) - C_1\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right)\right)$ will give rise to a fifth order U-process instead. Also, a similar expansion yields that:

$$\frac{1}{P}\sum_{t=R}^{T-l-1}\sum_{j=1}^{l}\varpi_{j,l}\widehat{A}_{1,P,R,t}(\tau)\widehat{A}_{1,P,R,t+j}(\tau)$$
$$= \frac{1}{4!P}\sum_{t=R}^{T-l-1}\sum_{j=1}^{l}\Bigg(1\{\mathbf{X}_t \in \mathcal{X}\}\left(\nabla^{(2)} L(0)\right)\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right) - C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)\right)^2$$
$$\times 1\{\mathbf{X}_{t+j} \in \mathcal{X}\}\left(\nabla^{(2)} L(0)\right)\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_{t+j}\right) - C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_{t+j}\right)\right)^2\Bigg) + o_p\left(\frac{l}{R^2}\right)$$
$$= O_p\left(\frac{l}{R^2}\right)(1+o_p(1))$$

Thus, pointwise in τ:

$$\frac{1}{P}\sum_{t=R}^{T-l-1}\left(\sum_{i=t}^{t+l}\left(\widehat{A}_{1,P,R,i}(\tau) - \widehat{A}_{2,P,R,i}(\tau)\right)\right)^2$$
$$= O_p\left(\frac{l}{R^2}\right)(1+o_p(1))$$

Now, recall that:

$$\widehat{B}_{1,P,R,t}(\tau) = 1\{\mathbf{X}_t \in \mathcal{X}\}\nabla^{(1)} L\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right) - \tau\right)\left(1\left\{y_{t+1} \le X_{1t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\right\} - \widehat{F}_{t+1}\left(X_{1t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\,|\mathbf{X}_t\right)\right)$$

Via another Taylor expansion around $C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)$, we have again that:

$$\frac{1}{P}\sum_{t=R}^{T-l-1}\widehat{B}_{1,P,R,t}(\tau)^2$$
$$= \frac{2}{P}\sum_{t=R}^{T-l-1}\Bigg(1\{\mathbf{X}_t \in \mathcal{X}\}\nabla^{(2)} L(0)^2\left(1\left\{y_{t+1} \le X_{1,t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\right\} - \widehat{F}_{t+1,P}\left(X_{1,t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\,|\mathbf{X}_t\right)\right)^2$$
$$\times\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right) - C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)\right)^2\Bigg) + o_p\left(\frac{1}{R}\right)$$
$$\le \frac{2C}{P}\sum_{t=R}^{T-l-1}\left(1\{\mathbf{X}_t \in \mathcal{X}\}\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right) - C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)\right)^2\right)$$
$$= O_p\left(\frac{1}{R}\right)(1+o_p(1)),$$

for some constant $C > 0$, where the inequality follows from A.2 and the boundedness of the indicator as well as the nonparametric estimator of $F_{t+1}\left(X_{1,t}'\boldsymbol{\beta}_1^{\dagger}(\tau)\,|\mathbf{X}_t\right)$, while the last line follows from the arguments in Theorem 1. In addition, applying the same Taylor expansion from before also yields that:

$$\frac{1}{P}\sum_{t=R}^{T-l-1}\sum_{j=1}^{l}\varpi_{j,l}\widehat{B}_{1,P,R,t}(\tau)\,\widehat{B}_{1,P,R,t+j}(\tau)$$
$$= O_p\left(\frac{l}{R}\right)(1+o_p(1))$$

and thus:

$$\frac{1}{P}\sum_{t=R}^{T-l-1}\left(\sum_{i=t}^{t+l}\left(\widehat{B}_{1,P,R,i}(\tau)-\widehat{B}_{2,P,R,i}(\tau)\right)\right)^2=O_p\left(\frac{l}{R}\right)(1+o_p(1))$$

pointwise in τ. Finally, note that:

$$\begin{aligned}&\frac{1}{R}\sum_{t=1}^{R-l-1}\widehat{D}_{1,P,R,t}(\tau)^2\\ =\;&\widehat{\Lambda}_{1,P,R}(\tau)\widehat{H}_{1,R}^{-1}(\tau)\frac{1}{R}\sum_{t=1}^{R-l-1}\left(X_{1,t}X_{1,t}'\left(1\{y_{t+1}\leq X_{1,t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\}-\tau\right)^2\right)\widehat{H}_{1,R}^{-1}(\tau)\widehat{\Lambda}_{1,P,R}(\tau)',\end{aligned}$$

where $\widehat{\Lambda}_{1,P,R}(\tau)$ was defined as:

$$\widehat{\Lambda}_{1,P,R}(\tau)=\frac{1}{P}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right)-\tau\right)\widehat{f}_{t+1,P}\left(X_{1,t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\,|\mathbf{X}_t\right)X_{1,t}'$$

Now, $\widehat{H}_{1,R}(\tau)$ and $1/R\sum_{t=1}^{R-l-1}\left(X_{1,t}X_{1,t}'\left(1\{y_{t+1}\leq X_{1,t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\}-\tau\right)^2\right)$ converge in probability to strictly positive definite matrices. On the other hand, another Taylor expansion around $C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)$ and arguments as before therefore yield that:

$$\begin{aligned}\widehat{\Lambda}_{1,P,R}(\tau)&=\nabla^{(2)}L(0)\left(\frac{1}{P}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}\widehat{f}_{t+1,P}\left(X_{1,t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\,|\mathbf{X}_t\right)X_{1,t}'\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right)-\tau\right)\right)+o_p\left(\frac{1}{\sqrt{P}}\right)\\&=\nabla^{(2)}L(0)\left(\frac{1}{P}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}f_{t+1,P}\left(X_{1,t}'\boldsymbol{\beta}_1^{\dagger}(\tau)\,|\mathbf{X}_t\right)X_{1,t}'\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right)-\tau\right)\right)+o_p\left(\frac{1}{\sqrt{P}}\right)\\&=O_p\left(\frac{1}{\sqrt{P}}\right)=O_p\left(\frac{1}{\sqrt{R}}\right),\end{aligned}$$

where the last line follows since by A.2(iii), A.3(i) and Theorem 1, $\sqrt{P}\widehat{\Lambda}_{1,P,R}(\tau)$ converges weakly, and by A.5, P and R grow at the same rate. Hence:

$$\frac{1}{R}\sum_{t=1}^{R-l-1}\widehat{D}_{1,P,R,t}(\tau)^2=O_p\left(\frac{1}{R}\right).$$

A similar argument shows that:

$$\frac{1}{R}\sum_{t=1}^{R-l-1}\sum_{j=1}^{l}\varpi_{j,l}\widehat{D}_{1,P,R,t}(\tau)\,\widehat{D}_{1,P,R,t+j}(\tau)=O_p\left(\frac{1}{R}\right)$$

since by A.1 and A.3:

$$\begin{aligned}&\frac{1}{R}\sum_{t=1}^{R-l-1}\sum_{j=1}^{l}\varpi_{j,l}\left(X_{1,t}X_{1,t}'\left(1\{y_{t+1}\leq X_{1,t}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\}-\tau\right)\right.\\&\left.\times X_{1,t+j}X_{1,t+j}'\left(1\{y_{t+j+1}\leq X_{1,t+j}'\widehat{\boldsymbol{\beta}}_{1,R}(\tau)\}-\tau\right)\right)\\=\;&O_p(1).\end{aligned}$$

Putting together the three components, it follows that:

$$\mathrm{Var}\left(\widehat{S}_{P,R}^{*}(\tau)|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)=O_p\left(\frac{l}{R^2}\right)+O_p\left(\frac{l}{R}\right)+O_p\left(\frac{1}{R}\right)=O_p\left(\frac{l}{R}\right).$$

Thus, $\widehat{S}_{P,R}^{*}(\tau)$ is of probability order $\sqrt{\frac{l}{R}}$, while $\widehat{S}_{P,R}(\tau)$ is of probability order $\frac{1}{\sqrt{R}}$ and so, for $l\to\infty$, $\widehat{S}_{P,R}^{*}(\tau)$ is of a larger probability order than $\widehat{S}_{P,R}(\tau)$, and the statement in (ii) follows.

Next, we move to the one-sided case where $C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)=C_2\left(\boldsymbol{\beta}_2^{\dagger}(\tau);\mathbf{X}_t\right)$ almost surely, but:

$$\Pr\left(C_j\left(\boldsymbol{\beta}_j^{\dagger}(\tau);\mathbf{X}_t\right)=\tau\right)<1,$$

for $j=1,2$. In this case, $A_{1,t}(\tau)-A_{2,t}(\tau)=0$ a.s.. Thus, given the definition of $\widehat{A}_{j,P,R,i}(\tau)$, omitting the recentering term for notational simplicity (it follows by similar arguments) and using A.5, we have that:

$$\begin{aligned}
&\frac{1}{P}\sum_{t=R}^{T-l-1}\frac{1}{l}\left(\sum_{i=t}^{t+l}\left(\widehat{A}_{1,P,R,i}(\tau)-\widehat{A}_{2,P,R,i}(\tau)\right)\right)^2\\
=\;&\frac{1}{P}\sum_{t=R}^{T-l-1}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)^2\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right)-\widehat{C}_{2,P}\left(\widehat{\boldsymbol{\beta}}_{2,R}(\tau);\mathbf{X}_t\right)\right)^2(1+o_p(1))\\
&+\frac{1}{P}\sum_{t=R}^{T-l-1}\sum_{j=1}^{l}\left(1-\frac{j}{l}\right)1\{\mathbf{X}_t\in\mathcal{X}\}1\{\mathbf{X}_{t+j}\in\mathcal{X}\}\nabla^{(1)}L\left(C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right)^2\nabla^{(1)}L\left(C_1\left(\boldsymbol{\beta}_1^{\dagger}(\tau);\mathbf{X}_{t+j}\right)-\tau\right)^2\\
&\times\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_t\right)-\widehat{C}_{2,P}\left(\widehat{\boldsymbol{\beta}}_{2,R}(\tau);\mathbf{X}_t\right)\right)\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\beta}}_{1,R}(\tau);\mathbf{X}_{t+j}\right)-\widehat{C}_{2,P}\left(\widehat{\boldsymbol{\beta}}_{2,R}(\tau);\mathbf{X}_{t+j}\right)\right)\\
&\times(1+o_p(1))\\
=\;&O_p\left(\sqrt{\frac{l}{R}}\right).
\end{aligned}$$

Similarly, since $X_{1,t}'\boldsymbol{\beta}_1^{\dagger}(\tau)=X_{1,t}'\boldsymbol{\beta}_2^{\dagger}(\tau)$ almost surely, we have by similar arguments that:

$$\frac{1}{P}\sum_{t=R}^{T-l-1}\frac{1}{l}\left(\sum_{i=t}^{t+l}\left(\widehat{B}_{1,P,R,i}(\tau)-\widehat{B}_{2,P,R,i}(\tau)\right)\right)^2=O_p\left(\sqrt{\frac{l}{R}}\right).$$

On the other hand, observe that for:

$$\widehat{\Lambda}_{j,P,R}(\tau)=\frac{1}{P}\sum_{t=R}^{T-1}1\{\mathbf{X}_t\in\mathcal{X}\}\left(\nabla^{(1)}L\left(\widehat{C}_{j,P,R}\left((0,\tau];\mathbf{X}_t\right)-\tau\right)\widehat{f}_{t+1,P}\left(X_{j,t}'\widehat{\boldsymbol{\beta}}_{j,R}(\tau)|\mathbf{X}_t\right)X_{j,t}'\right),$$

with $j=1,2$, it holds that $\widehat{\Lambda}_{j,P,R}(\tau)\xrightarrow{p}\Lambda_j(\tau)\neq\mathbf{0}$ in general by A.3 and A.4. Consequently, $\widehat{D}_{j,P,R,t}(\tau)$ is no longer $o_p(1)$ for all t. Thus, for the bootstrap statistic, it holds that:

$$\widehat{S}_{P,R}^{*}(\tau)=\frac{\sqrt{P}}{R}\sum_{t=1}^{R-l-1}\eta_t\sum_{i=t}^{l}\left(\widehat{D}_{1,P,R,i}(\tau)-\widehat{D}_{2,P,R,i}(\tau)\right)+o_p^{*}(1),$$

in probability.[6] Now, while the bootstrap mean is zero for almost all samples, the variance is given by:

$$\begin{aligned}
&\mathrm{Var}\left(\frac{\sqrt{P}}{R}\sum_{t=1}^{R-l-1}\eta_t\sum_{i=t}^{t+l}\left(\widehat{D}_{1,P,R,i}(\tau)-\widehat{D}_{2,P,R,i}(\tau)\right)|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)\\
=\;&\pi\frac{1}{R}\sum_{t=1}^{R-l-1}\left(\sum_{i=t}^{t+l}\left(\widehat{D}_{1,P,R,i}(\tau)-\widehat{D}_{2,P,R,i}(\tau)\right)\right)^2\\
=\;&\pi\Bigg(\Lambda_1(\tau)H_1(\tau)^{-1}\mathrm{Avar}\left(\frac{1}{\sqrt{R}}\sum_{t=1}^{R}X_{1,t}1\left\{y_{t+1}\leq X_{1,t}'\boldsymbol{\beta}_1^{\dagger}(\tau)\right\}\right)H_1(\tau)^{-1}\Lambda_1'(\tau)\\
&+\Lambda_2(\tau)H_2(\tau)^{-1}\mathrm{Avar}\left(\frac{1}{\sqrt{R}}\sum_{t=1}^{R}X_{2,t}1\left\{y_{t+1}\leq X_{2,t}'\boldsymbol{\beta}_2^{\dagger}(\tau)\right\}\right)H_2(\tau)^{-1}\Lambda_2'(\tau)\\
&-2\Lambda_1(\tau)H_1(\tau)^{-1}\mathrm{Acov}\left(\frac{1}{\sqrt{R}}\sum_{t=1}^{R}X_{1,t}1\left\{y_{t+1}\leq X_{1,t}'\boldsymbol{\beta}_1^{\dagger}(\tau)\right\},\frac{1}{\sqrt{R}}\sum_{t=1}^{R}X_{2,t}1\left\{y_{t+1}\leq X_{2,t}'\boldsymbol{\beta}_2^{\dagger}(\tau)\right\}\right)H_2(\tau)^{-1}\Lambda_2'(\tau)\Bigg)
\end{aligned}$$

[6]For any bootstrap statistic $\widehat{T}_{P,R}^{*}$, we write $\widehat{T}_{P,R}^{*}=o_p^{*}(1)$, in probability, if for any $\Delta>0$ it holds that $\Pr\left(\widehat{T}_{P,R}^{*}\mid\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)=o_p(1)$, where $\Pr\left(\cdot\mid\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)$ denotes the probability conditional on the original sample.

$$
\begin{aligned}
&\quad +o_p(1)\\
\equiv &\quad \pi\widetilde{\Omega}(\tau)+o_p(1),
\end{aligned}
$$

where $\widetilde{\Omega}(\tau)$ is positive definite by A.3 and A.4. Thus:

$$\widehat{S}^*_{P,R}(\tau)\overset{d^*}{\to}N\left(0,\pi\widetilde{\Omega}(\tau)\right)$$

in probability.[7] Hence, $\widehat{S}_{P,R}(\tau)=o_p(1)$, while $\widehat{S}^*_{P,R}(\tau)$ has a non degenerate limiting standard normal distribution. The statement then follows.

(iii) Under both hypotheses, we have that $\mathrm{E}\left(\widehat{S}^*_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right)=0$ for almost all samples and that $\mathrm{p\lim_{R,P\to\infty}}\,\mathrm{Var}\left(\widehat{S}^*_{P,R}(\tau)|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right)$ mimics the asymptotic variance under H_0 in CASE I while, on the other hand, the term $\lim_{R,P\to\infty}\mathrm{E}\left(\widehat{S}_{P,R}(\tau)\right)$ will diverge to plus or minus infinity at rate $\sqrt{P}$ under the alternative. ■

Proof of Theorem 3:
We focus on the one-sided interval case first, and comment at the end on the two-sided case.
Similar to the decomposition in the proof of Theorem 1, a second order Taylor expansion around $\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)$ and $\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)$ yields:

$$
\begin{aligned}
&\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}1\{\mathbf{X}_t\in\mathcal{X}\}\left(L\left(\widehat{\mathcal{E}}_{1,P,R}\left((0,\tau];\mathbf{X}_t\right)\right)-L\left(\widehat{\mathcal{E}}_{2,P,R}\left((0,\tau];\mathbf{X}_t\right)\right)\right)\\
=&\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}1\{\mathbf{X}_t\in\mathcal{X}\}\left(L\left(\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)-L\left(\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)+\frac{\delta_1((0,\tau];\mathbf{X}_t)}{\sqrt{P}}-\frac{\delta_2((0,\tau];\mathbf{X}_t)}{\sqrt{P}}\right)\\
&+\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)\left(\widehat{\mathcal{E}}_{1,P,R}\left((0,\tau];\mathbf{X}_t\right)-\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)\\
&-\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(1)}L\left(\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)\left(\widehat{\mathcal{E}}_{2,P,R}\left((0,\tau];\mathbf{X}_t\right)-\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)\\
&+\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(2)}L\left(\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)\left(\widehat{\mathcal{E}}_{1,P,R}\left((0,\tau];\mathbf{X}_t\right)-\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)^2\\
&-\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(2)}L\left(\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)\left(\widehat{\mathcal{E}}_{2,P,R}\left((0,\tau];\mathbf{X}_t\right)-\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)^2\\
&+\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(3)}L\left(\overline{\mathcal{E}}_1\left((0,\tau];\mathbf{X}_t\right)\right)\left(\widehat{\mathcal{E}}_{1,P,R}\left((0,\tau];\mathbf{X}_t\right)-\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)^3\\
&-\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}1\{\mathbf{X}_t\in\mathcal{X}\}\nabla^{(3)}L\left(\overline{\mathcal{E}}_2\left((0,\tau];\mathbf{X}_t\right)\right)\left(\widehat{\mathcal{E}}_{2,P,R}\left((0,\tau];\mathbf{X}_t\right)-\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)^3\\
=&\mathcal{T}^P_{1,P}+\mathcal{T}^P_{2,R,P}+\mathcal{T}^P_{3,R,P}+\mathcal{T}^P_{4,R,P},
\end{aligned}
$$

where $\nabla^{(1)}L(\cdot)$, $\nabla^{(2)}L(\cdot)$, and $\nabla^{(3)}L(\cdot)$ denote again the first, second, and third order derivative of $L(\cdot)$, while $\overline{\mathcal{E}}_1\left((0,\tau];\mathbf{X}_t\right)$ and $\overline{\mathcal{E}}_2\left((0,\tau];\mathbf{X}_t\right)$ denote intermediate values. Now while the arguments for $\mathcal{T}^P_{2,R,P}$, $\mathcal{T}^P_{3,R,P}$, and $\mathcal{T}^P_{4,R,P}$ are identical to the proof of Theorem 1, for $\mathcal{T}^P_{1,P}$, we obtain:

$$
\begin{aligned}
\mathcal{T}^P_{1,P}=&\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}\Bigg(1\{\mathbf{X}_t\in\mathcal{X}\}\left(L\left(\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)+\frac{\delta_1((0,\tau];\mathbf{X}_t)}{\sqrt{P}}\right)\\
&-\mathrm{E}\left(1\{\mathbf{X}_t\in\mathcal{X}\}\left(L\left(\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)+\frac{\delta_1((0,\tau];\mathbf{X}_t)}{\sqrt{P}}\right)\right)
\end{aligned}
$$

[7]For any bootstrap statistic $\widehat{T}^*_{P,R}$, we write $\widehat{T}^*_{P,R}\overset{d^*}{\to}D$, in probability, if conditional on the sample with probability that converges to one, $\widehat{T}^*_{P,R}$ weakly converges to the distribution D under $\Pr\left(\cdot\mid\{y_{t+1},\mathbf{X}_t'\}_{t=1}^T\right)$.

$$
\begin{aligned}
&-1\{\mathbf{X}_t \in \mathcal{X}\}\left(L\left(\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)+\frac{\delta_2((0,\tau];\mathbf{X}_t)}{\sqrt{P}}\right)\\
&-\mathrm{E}\left(1\{\mathbf{X}_t \in \mathcal{X}\}\left(L\left(\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)+\frac{\delta_2((0,\tau];\mathbf{X}_t)}{\sqrt{P}}\right)\right)\Bigg)\\
&+\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}\left(\mathrm{E}\left(1\{\mathbf{X}_t \in \mathcal{X}\}\left(L\left(\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)+\frac{\delta_1((0,\tau];\mathbf{X}_t)}{\sqrt{P}}-L\left(\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)-\frac{\delta_2((0,\tau];\mathbf{X}_t)}{\sqrt{P}}\right)\right)\right)\\
=\ &\frac{1}{\sqrt{P}}\sum_{t=R+1}^{T}\Bigg(1\{\mathbf{X}_t \in \mathcal{X}\}\left(L\left(\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)+\frac{\delta_1((0,\tau];\mathbf{X}_t)}{\sqrt{P}}\right)\\
&-\mathrm{E}\left(1\{\mathbf{X}_t \in \mathcal{X}\}\left(L\left(\mathcal{E}_1\left((0,\tau];\mathbf{X}_t\right)\right)+\frac{\delta_1((0,\tau];\mathbf{X}_t)}{\sqrt{P}}\right)\right)\\
&-1\{\mathbf{X}_t \in \mathcal{X}\}\left(L\left(\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)+\frac{\delta_2((0,\tau];\mathbf{X}_t)}{\sqrt{P}}\right)\\
&-\mathrm{E}\left(1\{\mathbf{X}_t \in \mathcal{X}\}\left(L\left(\mathcal{E}_2\left((0,\tau];\mathbf{X}_t\right)\right)+\frac{\delta_2((0,\tau];\mathbf{X}_t)}{\sqrt{P}}\right)\right)\Bigg)\\
&+\zeta\left(\tau\right).
\end{aligned}
$$

Now, since $\zeta\left(\tau\right)\neq 0$, and by A.1, A.2, A.4 as well as a CLT for strong mixing observations pointwise in τ:

$$
\mathcal{T}_{1,P}^{P}\xrightarrow{d}N\left(\zeta\left(\tau\right),V_{\mathcal{T}_{1,T}}\right),
$$

where $V_{\mathcal{T}_{1,T}}$ is the same as under H_0. Hence, under $H_{A,P}$ the statistic will converge to a Gaussian distribution with mean $\zeta\left(\tau\right)$ and the same variance as under H_0 in CASE I. Finally, for a single interval of the form $[\tau_L,\tau_U]$, the same arguments as above apply, but with the drifting sequence $\frac{\zeta(\tau)}{\sqrt{P}}$ replaced by:

$$
\left(\frac{\zeta\left(\tau_U\right)}{\sqrt{P}}-\frac{\zeta\left(\tau_L\right)}{\sqrt{P}}\right).
$$

■

Proof of Theorem 4:
(i) In line with the change in notation for Section 4 of the paper, recall that $\mathcal{P}$ denotes the set of probability measures, $\boldsymbol{P}$ (= Pr), defined on the support of $\mathbf{X}_t$ such that Assumptions A.1, A.3, and A.4 hold. Also, denote

$$
\mathcal{P}_0^{I-RC}=\left\{\boldsymbol{P}\in\mathcal{P}:\ H_{0,P}^{RC}\text{ and CASE I-RC holds}\right\},
$$

and recall that:

$$
\varsigma_k=\lim_{P\to\infty}\underbrace{\sqrt{P}\mathrm{E}_{\boldsymbol{P}}\left(\left(L\left(\mathcal{E}_1\left(\left[\tau_{i,L},\tau_{i,U}\right];\mathbf{X}_t^k\right)\right)-L\left(\mathcal{E}_k\left(\left[\tau_{i,L},\tau_{i,U}\right];\mathbf{X}_t^k\right)\right)\right)1\left\{\mathbf{X}_t^k\in\mathcal{X}\right\}\right)}_{=\mu_{k,P}}.
$$

Given Assumptions A.1-A.5, by the same argument used in the proof of Theorem 1(i),

$$
\begin{pmatrix}
\widehat{S}_{P,R,1}-\varsigma_1\\
\vdots\\
\widehat{S}_{P,R,k}-\varsigma_k\\
\vdots\\
\widehat{S}_{P,R,M(J-1)}-\varsigma_{M(J-1)}
\end{pmatrix}\xrightarrow{d}N\left(0,\mathbf{V}\right)
$$

where $\mathbf{V}$ is a $M(J-1)\times M(J-1)$ positive semidefinite matrix as defined in (32)-(33). Now, $\widehat{S}_{P,R}^{\max}$ satisfies the Assumptions 1-6 of Andrews and Soares (2010) by Lemma 1 therein. Thus, since $\widehat{S}_{P,R,k}=\widehat{S}_{P,R,k}-\varsigma_k+\varsigma_k$ for every $k=(j-2)M+i$, $j=2,...,J$, $i=1,...,M$, the statement then follows from the result in Eq.(4.2) of Andrews and Soares (2010), noting that the assumptions of Theorem 4 satisfy conditions (A.2) and (A.3) of Andrews and Soares (2010) for dependent data (see Section A.2 therein), and that A.1, A.3, and A.4 hold for a given $\boldsymbol{P}\in\mathcal{P}_0^{I-RC}$.

(ii) We treat subcases **(a)** and **(b)** together. By the same argument used in the proof of Theorem 1(ii), we know that, in the case of correct specification, for $k = 1, ..., M(J-1)$ and every $\boldsymbol{P} \in \mathcal{P}_0^{IIa-RC}$

$$\lim_{R,P\to\infty} \boldsymbol{P}\left(|\widehat{S}_{P,R,k}| \leq \Delta\frac{\sqrt{P}}{R}\right) = 1$$

for any $\Delta > 0$, so that, recalling that $M(J-1)$ is a finite number:

$$\lim_{R,P\to\infty} \boldsymbol{P}\left(\widehat{S}_{P,R}^{\max} \leq \Delta'\frac{P}{R^2}\right) = 1$$

for any $\Delta' > 0$, where

$$\widehat{S}_{P,R}^{\max} = \sum_{k=1}^{M(J-1)} \widehat{S}_{P,R,k}^2 1\left\{\widehat{S}_{P,R,k} > 0\right\}.$$

Similarly, in the case of misspecification with conditional coverage not equal to the nominal level almost surely, we know from Theorem 1(ii) that $\widehat{S}_{P,R,k} = o_p(1)$ for $k = 1, ..., M(J-1)$ and every $\boldsymbol{P} \in \mathcal{P}_0^{IIb-RC}$. As a result, it also holds that:

$$\widehat{S}_{P,R}^{\max} = o_p(1)$$

since $M(J-1)$ is a finite number.

(iii) Under H_A^{RC}, for at least one k, $\frac{\mu_{k,P}}{\sqrt{P}} \to \mu_k > 0$, with μ_k bounded away above zero. Thus, for at least one k, and a given $\boldsymbol{P} \in \mathcal{P}_A^{RC}$

$$\lim_{P,R\to\infty} \boldsymbol{P}\left(\frac{1}{\sqrt{P}}\widehat{S}_{P,R,k} > \epsilon\right) = 1$$

for any $\epsilon > 0$. As a result, it follows that with probability converging to one:

$$\widehat{S}_{P,R}^{\max} = \widehat{S}_{P,R,k}^2 + \sum_{l\neq k}^{M(J-1)} \widehat{S}_{P,R,l}^2 1\left\{\widehat{S}_{P,R,l} > 0\right\}.$$

Finally, since

$$\widehat{S}_{P,R,k}^2 = \left(\widehat{S}_{P,R,k} - \mu_{k,P}\right)^2 + \mu_{k,P}^2 + 2\left(\widehat{S}_{P,R,k} - \mu_{k,P}\right)\mu_{k,P},$$

the statement follows. ■

Proof of Theorem 5:
(i) We use again the same notation as in the proof of Theorem 4 and Section 4. By the same argument used in the proof of Theorem 2(i)-(ii), for each $k = (j-2)M + i$, $i = 1, \ldots, M$, $j = 2, \ldots J$, $\widehat{S}_{P,R,k}^*$ has, conditionally on the sample and for all samples but a set of probability measure approaching zero, the same limiting distribution:

$$\widehat{S}_{P,R,k}^{\mu_{k,P}} = \widehat{S}_{P,R,k} - \sqrt{P}\mathrm{E}_{\boldsymbol{P}}\left(\left(L\left(\mathcal{E}_1\left([\tau_{i,L}, \tau_{i,U}]; \mathbf{X}_t^j\right)\right) - L\left(\mathcal{E}_j\left([\tau_{i,L}, \tau_{i,U}]; \mathbf{X}_t^j\right)\right)\right) 1\left\{\mathbf{X}_t^j \in \mathcal{X}\right\}\right).$$

Since $M(J-1)$ is finite, as an immediate consequence of the Cramer Wold device,

$$\begin{pmatrix} \widehat{S}_{P,R,1}^* \\ \vdots \\ \widehat{S}_{P,R,M(J-1)}^* \end{pmatrix} \xrightarrow{d^*} N(0, \mathbf{V}) \tag{S.26}$$

in probability, with $\mathbf{V}$ defined in (32)-(33). Given Assumption A.1-A.5, and the definition of the lag truncation parameter π_s in the main text, for all $k = 1, \ldots, M(J-1)$, $\widehat{v}_{kk,P,R} - v_{kk} = o_p(1)$, with $v_{kk} = \mathrm{Avar}\left(\widehat{S}_{P,R,k}\right)$. Thus:

$$\begin{aligned} &\widehat{S}_{P,R}^{*\max} \\ &= \sum_{k=1}^{M(J-1)} \left(\max\left\{0, \widehat{S}_{P,R,k}^* 1\left\{\widehat{S}_{P,R,k} \geq -\sqrt{v_{kk}}\kappa_P\right\}\right\}\right)^2 \end{aligned}$$

$$
\begin{aligned}
&+\sum_{k=1}^{M(J-1)}\left(\max\left\{0,\widehat{S}^{*}_{P,R,k}1\left\{\widehat{S}_{P,R,k}\geq-\sqrt{\widehat{v}_{kk,P,R}}\kappa_P\right\}\right\}\right)^2-\left(\max\left\{0,\widehat{S}^{*}_{P,R,k}1\left\{\widehat{S}_{P,R,k}\geq-\sqrt{v_{kk}}\kappa_P\right\}\right\}\right)^2\\
=&\sum_{k=1}^{M(J-1)}\left(\max\left\{0,\widehat{S}^{*}_{P,R,k}1\left\{\widehat{S}_{P,R,k}\geq-\sqrt{v_{kk}}\kappa_P\right\}\right\}\right)^2(1+o_p(1))\,.
\end{aligned}
$$

To establish the $o_p(1)$ term from the last line note that for each k and some $\epsilon_1>0$ and a given $\boldsymbol{P}\in\mathcal{P}_0^{I-RC}$:

$$
\begin{aligned}
&\boldsymbol{P}\left(\left|\left(\max\left\{0,\widehat{S}^{*}_{P,R,k}1\left\{\widehat{S}_{P,R,k}\geq-\sqrt{\widehat{v}_{kk,P,R}}\kappa_P\right\}\right\}\right)^2\right.\right.\\
&\left.\left.-\left(\max\left\{0,\widehat{S}^{*}_{P,R,k}1\left\{\widehat{S}_{P,R,k}\geq-\sqrt{v_{kk}}\kappa_P\right\}\right\}\right)^2\right|>\epsilon_1\,\middle|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)\\
=&\ \boldsymbol{P}\left(\left(\widehat{S}^{*}_{P,R,k}\right)^2>\epsilon_1\,\middle|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)\times\left(1\left\{|\sqrt{v_{kk}}\kappa_P|\leq|\widehat{S}_{P,R,k}|\leq|\sqrt{\widehat{v}_{kk,P,R}}\kappa_P|\right\}\right.\\
&\left.+1\left\{|\sqrt{\widehat{v}_{kk,P,R}}\kappa_P|\leq|\widehat{S}_{P,R,k}|\leq|\sqrt{v_{kk}}\kappa_P|\right\}\right),
\end{aligned}
$$

for almost all samples, where $\boldsymbol{P}\left(\cdot\mid\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)$ denotes the probability conditional on the original sample. Thus, by Markov's inequality, we have that:

$$
\begin{aligned}
&\boldsymbol{P}\left(\left|\left(\max\left\{0,\widehat{S}^{*}_{P,R,k}1\left\{\widehat{S}_{P,R,k}\geq-\sqrt{\widehat{v}_{kk,P,R}}\kappa_P\right\}\right\}\right)^2\right.\right.\\
&\left.\left.-\left(\max\left\{0,\widehat{S}^{*}_{P,R,k}1\left\{\widehat{S}_{P,R,k}\geq-\sqrt{v_{kk}}\kappa_P\right\}\right\}\right)^2\right|>\epsilon_1\,\middle|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)\\
\leq&\ \left(\mathrm{E}_{\boldsymbol{P}}\left(\left(\widehat{S}^{*}_{P,R,k}\right)^2|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)\times\left(1\left\{\sqrt{v_{kk}}\kappa_P\leq|\widehat{S}_{P,R,k}|\leq\sqrt{\widehat{v}_{kk,P,R}}\kappa_P\right\}\right.\right.\\
&\left.\left.+1\left\{\sqrt{\widehat{v}_{kk,P,R}}\kappa_P\leq|\widehat{S}_{P,R,k}|\leq\sqrt{v_{kk}}\kappa_P\right\}\right)\right)/\epsilon_1,
\end{aligned}
$$

We will only examine the first term in what follows. Now let δ_P denote a deterministic sequence such that $\delta_P\to0$ as $P\to0$ and that $\delta_P^{-1}(\sqrt{\widehat{v}_{kk,P,R}}-\sqrt{v_{kk}})=o_p(1)$. Then, for sufficiently large P and some $\epsilon_2>0$:

$$
\begin{aligned}
&\boldsymbol{P}\left(\left(\mathrm{E}_{\boldsymbol{P}}\left(\left(\widehat{S}^{*}_{P,R,k}\right)^2|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)1\left\{\sqrt{v_{kk}}\kappa_P\leq|\widehat{S}_{P,R,k}|\leq\sqrt{\widehat{v}_{kk,P,R}}\kappa_P\right\}\right)>\epsilon_2\right)\\
\leq&\ \boldsymbol{P}\left(\left(\mathrm{E}_{\boldsymbol{P}}\left(\left(\widehat{S}^{*}_{P,R,k}\right)^2|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)1\left\{\sqrt{v_{kk}}\kappa_P\leq|\widehat{S}_{P,R,k}|\leq(\sqrt{v_{kk,P,R}}+\delta_P)\kappa_P\right\}\right)>\epsilon_2\right)\\
\leq&\ C\mathrm{E}_{\boldsymbol{P}}\left(\mathrm{E}_{\boldsymbol{P}}\left(\left(\widehat{S}^{*}_{P,R,k}\right)^2|\{y_{t+1},\mathbf{X}_t'\}_{t=1}^{T}\right)^2\right)^{\frac{1}{2}}\mathrm{E}_{\boldsymbol{P}}\left(1\left\{\sqrt{v_{kk}}\kappa_P\leq|\widehat{S}_{P,R,k}|\leq(\sqrt{v_{kk,P,R}}+\delta_P)\kappa_P\right\}\right)^{\frac{1}{2}}\\
=&\ O(1)o(1),
\end{aligned}
$$

where C denotes some positive constant and the $O(1)$ term follows from similar arguments to the ones used in Theorem 2(i).

Now, let $c^{*}_{B,R,P,1-\alpha}$ be the $(1-\alpha)$ critical value of $\widehat{S}^{*\max}_{P,R}$ based on B bootstrap replications. Also, consider a sequence $\{\boldsymbol{\gamma}_P\}_{P=1}^{\infty}$ with $\boldsymbol{\gamma}_P=\left(\gamma_{1,P},...,\gamma_{(J-1)M,P}\right)$ and each $\boldsymbol{\gamma}_P\in\mathcal{P}_0^{I-RC}$ such that $\sqrt{P}\gamma_P\to\boldsymbol{h}$ and $\kappa_P^{-1}\sqrt{P}\gamma_P\to\boldsymbol{\xi}$ where $\boldsymbol{h},\boldsymbol{\xi}\in\mathbb{R}_{-,\infty}^{(J-1)M}$ with $\mathbb{R}_{-}=\{x\in\mathbb{R}:x\leq0\}$ and $\mathbb{R}_{-,\infty}=\mathbb{R}_{-}\cup\{-\infty\}$. Then, let $c_{R,P,1-\alpha}$ be the $(1-\alpha)$ critical values of $S^{\max}_{R,P}$ under the drifting sequence $\boldsymbol{\gamma}_P$. By Lemma 2(a) in the supplement of Andrews and Soares (2010), $c^{*}_{P,1-\alpha}\leq c_{R,P,1-\alpha}$ almost surely for all P for a sequence such that $c^{*}_{P,1-\alpha}\xrightarrow{p}c^{*}_{1-\alpha}=\lim_{B,R,P\to\infty}c^{*}_{B,R,P,1-\alpha}$ noting that the assumptions of Theorem 5 together with the HAC estimator $\widehat{v}_{kk,P,R}$ satisfy conditions (A.2) and (A.3) of Andrews and Soares (2010) for dependent data, and that A.1, A.3, and A.4 hold for every $\gamma_P\in\mathcal{P}_0^{I-RC}$. Also, under the drifting sequence $\{\boldsymbol{\gamma}_P,P\geq1\}$, $\lim_{P,R\to\infty}c_{R,P,1-\alpha}=c^{\dagger}_{1-\alpha}$ which is the $(1-\alpha)$ critical value of the limiting distribution of $\widehat{S}^{\max}_{P,R}$ in Theorem 4. The first part of Theorem 5-(i) then follows from subsequence arguments analogous to the ones used in the proof of Theorem 1(i) in Andrews and Soares (2010). The results for the second part of Theorem 5-(i) on the other hand follows analogously by Lemma 3(a) and Theorem 1(ii) in the supplement of Andrews and Soares (2010).

(ii)-(a) From Theorem 1(ii)-(a), for each $k = 1, \ldots, (J-1)M$, $\widehat{S}_{P,R,k} = O_p\left(\frac{\sqrt{P}}{R}\right)$. On the other hand, $\widehat{v}_{kk,P,R} = O_p\left(\frac{\sqrt{P \cdot l_P}}{R}\right)$, and $\kappa_P \to \infty$ as $P \to \infty$. Thus, since $l_P \to \infty$ as $P \to \infty$, from Theorem 2(ii) for all k and all $\boldsymbol{P} \in \mathcal{P}_0^{IIa-RC}$:

$$\lim_{P,R,B\to\infty} \boldsymbol{P}\left(\left(\max\left\{0, \widehat{S}_{P,R,k}\right\} - \max\left\{0, \widehat{S}^*_{P,R,k} 1\left\{\widehat{S}_{P,R,k} \geq -\sqrt{\widehat{v}_{kk,P,R}}\kappa_P\right\}\right\}\right) > 0\right) = 0.$$

The result then follows from a standard subsequence argument for sequences $\{\boldsymbol{P}_P\}_{P=1}^{\infty}$ with $\boldsymbol{P}_P \in \mathcal{P}_0^{IIa-RC}$ for all P, which has been omitted for brevity.

(ii)-(b) This follows by the same argument as in (ii)-(a) noting that here $\widehat{S}_{P,R,k} = o_p(1)$ for each $k = 1, \ldots, (J-1)M$ from Theorem 1(ii)-(b), while:

$$\widehat{S}^*_{P,R,k}(\tau) \xrightarrow{d^*} N\left(0, \pi\widetilde{\Omega}_k(\tau)\right)$$

in probability.

(iii) As in (ii), assume again without loss of generality that the null is violated for the first K models with $K + K' = (J-1)M$ and $K > 0$. For $k \in \{1, \ldots, K\}$, $\frac{\mu_{k,P}}{\sqrt{P}} > 0$ and so as $P \to \infty$, $\mu_{k,P} \to \infty$ at rate $\sqrt{P}$. Now, for sufficiently large P, it holds that:

$$\begin{aligned}
\widehat{S}^{\max}_{P,R} &= \sum_{k\in\{1,\ldots,K\}} \left(\max\left\{0, \widehat{S}_{P,R,k}\right\}\right)^2 + \sum_{k\in\{K+1,\ldots,(J-1)M\}} \left(\max\left\{0, \widehat{S}_{P,R,k}\right\}\right)^2 \\
&= \sum_{k\in\{1,\ldots,K\}} \widehat{S}^2_{P,R,k} + \sum_{k\in\{K+1,\ldots,(J-1)M\}} \left(\max\left\{0, \widehat{S}_{P,R,k}\right\}\right)^2.
\end{aligned}$$

Since,

$$\sum_{k\in\{1,\ldots,K\}} \widehat{S}^2_{P,R,k} = \sum_{k\in\{1,\ldots,K\}} \left(\widehat{S}_{P,R,k} - \mu_{k,P}\right)^2 + \sum_{k\in\{1,\ldots,K\}} \mu^2_{k,P} + 2\sum_{k\in\{1,\ldots,K\}} \left(\widehat{S}_{P,R,k} - \mu_{k,P}\right)\mu_{k,P}$$

and diverges to $+\infty$ as $P \to \infty$, while $\widehat{S}^{*\max}_{P,R}$ converges, conditional on the sample, in distribution also under H_A^{RC}. The statement then follows. ■

S.2.3 Proofs of Lemmas

Proof of Lemma Q.1:[8]

(i) In analogy to the notation of Angrist et al. (2006), let

$$Q_R(\tau, \boldsymbol{\beta}) \equiv \frac{1}{R}\sum_{j=1}^{R-1} \left(\rho_\tau(y_{j+1} - X_j'\boldsymbol{\beta}) - \rho_\tau(y_{j+1} - X_j'\boldsymbol{\beta}^\dagger(\tau))\right)$$

and

$$Q_\infty(\tau, \boldsymbol{\beta}) \equiv \mathrm{E}\left(\rho_\tau(y_{j+1} - X_j'\boldsymbol{\beta}) - \rho_\tau(y_{j+1} - X_j'\boldsymbol{\beta}^\dagger(\tau))\right).$$

Then, pointwise in $\boldsymbol{\beta}$ and τ, $Q_R(\tau, \boldsymbol{\beta}) \xrightarrow{p} Q_\infty(\tau, \boldsymbol{\beta})$ by A.1, A.3, and McLeish's law of large numbers for strong mixing processes. Consistency of $\widehat{\boldsymbol{\beta}}_R(\tau)$ for $\boldsymbol{\beta}^\dagger(\tau)$ pointwise in τ then follows by the same arguments as in the proof of Theorem 3 of Angrist et al. (2006).

(ii) To establish stochastic equicontinuity w.r.t. the pseudo-metric $\rho_{\mathcal{B}\times\mathcal{T}}(\cdot,\cdot)$, first note that the function class $\mathcal{F}_1 = \left\{1\{y_{j+1} \leq X_j'\boldsymbol{\beta}\} : \boldsymbol{\beta} \in \mathcal{B}\right\}$ is a VC subgraph class by the arguments in the proof of Theorem 1 (see pp.16-17) and thus belongs to a bounded Donsker class. Moreover, note that the function class $\mathcal{F}_2 = \{\tau \mapsto \tau : \tau \in \mathcal{T}\}$ also belongs to a bounded Donsker class. As a result, the functional class $\mathcal{F}_1 - \mathcal{F}_2$ is the difference of Donsker classes with envelope 2 whose covering numbers satisfy (S.23). Thus, letting $\mathcal{F}_3 = \{X_j\}$, by Assumption A.1 and Lemma A.1 in Ghosal et al. (2000):

$$N\left(\epsilon\|X_j\|_{L_2(Q)}\|, \mathcal{F}_1\mathcal{F}_2\mathcal{F}_3, L_2(Q)\right) \leq a\left(\frac{1}{\epsilon}\right)^b,$$

[8] For notational simplicity, we drop again the model subscript j.

with square integrable envelope $2\max_{k\in\{1,\ldots,d\}}|X_{jk}|$. Stochastic equicontinuity follows thus again from Corollary 2.1 in Arcones and Yu (1994).

(iii) By Equation (4) and A.3 (see e.g. Gregory et al., 2018), it holds that:

$$\begin{aligned} &\left\| \frac{1}{\sqrt{R}} \sum_{j=1}^{R-1} X_j \left(1\left\{ y_{j+1} \le X_j' \widehat{\boldsymbol{\beta}}_R(\tau) \right\} - \tau \right) \right\| \\ \le\ & d \max_{1\le j\le R} \|X_j\| \frac{1}{\sqrt{R}} \sum_{j=1}^{R-1} 1\left\{ y_{j+1} = X_j'\widehat{\boldsymbol{\beta}}_R(\tau) \right\} = o_p(1), \end{aligned}$$

where the last equality follows again from the fact that $\max_{1\le j\le R}\|X_j\| = o_p(R^{\frac{1}{2}})$. Moreover, by parts (i) and (ii), conditional on the sample, we have that:

$$\begin{aligned} &\frac{1}{\sqrt{R}} \sum_{j=1}^{R-1} \left(X_j \left(1\left\{ y_{j+1} \le X_j'\widehat{\boldsymbol{\beta}}_R(\tau) \right\} - \tau \right) - \mathrm{E}_T\left(X_j \left(1\left\{ y_{j+1} \le X_j'\widehat{\boldsymbol{\beta}}_R(\tau) \right\} - \tau \right) \right) \right) \\ =\ & \frac{1}{\sqrt{R}} \sum_{j=1}^{R-1} \left(X_j \left(1\left\{ y_{j+1} \le X_j'\boldsymbol{\beta}^\dagger(\tau) \right\} - \tau \right) - \underbrace{\mathrm{E}_T\left(X_j \left(1\left\{ y_{j+1} \le X_j'\boldsymbol{\beta}^\dagger(\tau) \right\} - \tau \right) \right)}_{=0} \right) + o_p(1). \end{aligned}$$

Thus:

$$\frac{1}{\sqrt{R}} \sum_{j=1}^{R-1} \mathrm{E}_T\left(X_j \left(1\left\{ y_{j+1} \le X_j'\widehat{\boldsymbol{\beta}}_R(\tau) \right\} - \tau \right) \right) = \frac{1}{\sqrt{R}} \sum_{j=1}^{R-1} X_j \left(\left(1\left\{ y_{j+1} \le X_j'\boldsymbol{\beta}^\dagger(\tau) \right\} - \tau \right) \right) + o_p(1).$$

Via a mean value expansion of the left hand side around $\boldsymbol{\beta}^\dagger(\tau)$, we have that

$$\frac{1}{\sqrt{R}} \sum_{j=1}^{R-1} X_j \left(1\left\{ y_{j+1} \le X_j'\boldsymbol{\beta}^\dagger(\tau) \right\} - \tau \right) = H(\tau)\sqrt{R}\left(\widehat{\boldsymbol{\beta}}_R(\tau) - \boldsymbol{\beta}^\dagger(\tau) \right) + o_p(1)$$

Using A.3(ii) establishes the Bahadur representation for $\sqrt{R}\left(\widehat{\boldsymbol{\beta}}_R(\tau) - \boldsymbol{\beta}^\dagger(\tau) \right)$. ■

Proof of Lemma L.1:

(i) Recall that:

$$y_{t+1} = m(X_t, \boldsymbol{\theta}_m^\dagger) + \sigma\left(X_t, \boldsymbol{\theta}_\sigma^\dagger \right) \epsilon_{t+1}.$$

Consistency of $\widehat{\boldsymbol{\theta}}_{m,R}$ and $\widehat{\boldsymbol{\theta}}_{\sigma,R}$ follows from Assumption A.6(i). To show consistency of $\widehat{\beta}_R(\tau)$ for $\beta^\dagger(\tau)$, let:

$$\epsilon_{t+1} = \frac{y_{t+1} - m(X_t, \boldsymbol{\theta}_m^\dagger)}{\sigma\left(X_t, \boldsymbol{\theta}_\sigma^\dagger \right)} \quad \text{and} \quad \widehat{\epsilon}_{t+1} = \frac{y_{t+1} - m(X_t, \widehat{\boldsymbol{\theta}}_{m,R})}{\sigma\left(X_t, \widehat{\boldsymbol{\theta}}_{\sigma,R} \right)}$$

Then, after some algebra:

$$\widehat{\epsilon}_{t+1} - \epsilon_{t+1} = \frac{m(X_t, \boldsymbol{\theta}_m^\dagger) - m(X_t, \widehat{\boldsymbol{\theta}}_{m,R})}{\sigma\left(X_t, \widehat{\boldsymbol{\theta}}_{\sigma,R} \right)} + \left(\frac{\sigma\left(X_t, \boldsymbol{\theta}_\sigma^\dagger \right) - \sigma\left(X_t, \widehat{\boldsymbol{\theta}}_{\sigma,R} \right)}{\sigma\left(X_t, \widehat{\boldsymbol{\theta}}_{\sigma,R} \right) \sigma\left(X_t, \boldsymbol{\theta}_2^\dagger \right)} \right) \left(y_{t+1} - m(X_t, \boldsymbol{\theta}_m^\dagger) \right).$$

Next, recall that $\widehat{\beta}_R(\tau)$ is defined as:

$$\widehat{\beta}_R(\tau) = \arg\min_{\beta\in\mathcal{B}} \frac{1}{R} \sum_{t=1}^{R-1} \left(\widehat{\epsilon}_{t+1} - \beta \right) \left(\tau - 1\left\{ \widehat{\epsilon}_{t+1} - \beta \le 0 \right\} \right)$$

From Lemma Q.1(iii), we have that:

$$\left(\widehat{\beta}_R(\tau) - \beta^\dagger(\tau) \right)$$

$$
\begin{aligned}
&= H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}\left(1\left\{\widehat{\epsilon}_{t+1}\leq\beta^{\dagger}(\tau)\right\}-\tau\right)+o_p(1)\\
&= -H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}\left(1\left\{\epsilon_{t+1}\leq\beta^{\dagger}(\tau)\right\}-\tau\right)\\
&\quad -H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}1\left\{\widehat{\epsilon}_{t+1}\leq\beta^{\dagger}(\tau)\right\}-1\left\{\epsilon_{t+1}\leq\beta^{\dagger}(\tau)\right\}+o_p(1)
\end{aligned}
\tag{S.27}
$$

Now, using A.6(iv):

$$
\begin{aligned}
& H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}1\left\{\widehat{\epsilon}_{t+1}\leq\beta^{\dagger}(\tau)\right\}-1\left\{\epsilon_{t+1}\leq\beta^{\dagger}(\tau)\right\}\\
=\; & H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}1\left\{\epsilon_{t+1}\leq\beta^{\dagger}(\tau)-(\widehat{\epsilon}_{t+1}-\epsilon_{t+1})\right\}-1\left\{\epsilon_{t+1}\leq\beta^{\dagger}(\tau)\right\}\\
=\; & H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}\left(\left(1\left\{\epsilon_{t+1}\leq\beta^{\dagger}(\tau)-(\widehat{\epsilon}_{t+1}-\epsilon_{t+1})\right\}-F_{\epsilon}\left(\beta^{\dagger}(\tau)-(\widehat{\epsilon}_{t+1}-\epsilon_{t+1})\right)\right)\right.\\
& \left.-\left(1\left\{\epsilon_{t+1}\leq\beta^{\dagger}(\tau)\right\}-F_{\epsilon}\left(\beta^{\dagger}(\tau)\right)\right)\right)\\
& +H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}\left(F_{\epsilon}\left(\beta^{\dagger}(\tau)-(\widehat{\epsilon}_{t+1}-\epsilon_{t+1})\right)-F_{\epsilon}\left(\beta^{\dagger}(\tau)\right)\right)\\
=\; & H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}\left(F_{\epsilon}\left(\beta^{\dagger}(\tau)-(\widehat{\epsilon}_{t+1}-\epsilon_{t+1})\right)-F_{\epsilon}\left(\beta^{\dagger}(\tau)\right)\right)+o_p(R^{-\frac{1}{2}}),
\end{aligned}
$$

where the last line follows again from Corollary 5.1 in Hall and Heyde (1980) and the fact that:

$$
H(\tau)^{-1}\frac{1}{\sqrt{R}}\sum_{t=1}^{R-1}\left(1\left\{\epsilon_{t+1}\leq\beta\right\}-F_{\epsilon}(\beta)\right)
$$

with $H(\tau)=\mathrm{E}\left(f_{t+1}(q(\boldsymbol{\psi}(\tau);X_t)|X_t)\right)$ is stochastically equicontinuous in $\beta\in\mathcal{B}$ using similar arguments as for $\mathcal{A}_{1,P}$ in the proof of Theorem 1 and the fact that $H(\tau)$ is bounded by A.3. Thus, by mean value expansions around $\boldsymbol{\theta}_m^{\dagger}$ and $\boldsymbol{\theta}_{\sigma}^{\dagger}$:

$$
\begin{aligned}
& H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}f_{\epsilon}(\beta^{\dagger}(\tau))\left(\widehat{\epsilon}_{t+1}-\epsilon_{t+1}\right)\\
=\; & -H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}\frac{f_{\epsilon}(\beta^{\dagger}(\tau))}{\sigma\left(X_t,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\nabla_{\boldsymbol{\theta}_m}m(X_t,\boldsymbol{\theta}_m^{\dagger})\left(\widehat{\boldsymbol{\theta}}_{m,R}-\boldsymbol{\theta}_m^{\dagger}\right)\\
& -H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}\frac{f_{\epsilon}(\beta^{\dagger}(\tau))}{\sigma\left(X_t,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\frac{\nabla_{\boldsymbol{\theta}_{\sigma}}\sigma(X_{t+1},\boldsymbol{\theta}_{\sigma}^{\dagger})}{\sigma\left(X_t,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\epsilon_{t+1}\left(\widehat{\boldsymbol{\theta}}_{\sigma,R}-\boldsymbol{\theta}_{\sigma}^{\dagger}\right)+o_p(1)
\end{aligned}
$$

Consistency then follows from Assumption A.6(i) as $\left(\widehat{\boldsymbol{\theta}}_{m,R}-\boldsymbol{\theta}_m^{\dagger}\right)=o_p(1)$ and $\left(\widehat{\boldsymbol{\theta}}_{\sigma,R}-\boldsymbol{\theta}_{\sigma}^{\dagger}\right)=o_p(1)$, and from the fact that:

$$
H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}\left(1\left\{\epsilon_{t+1}\leq\beta^{\dagger}(\tau)\right\}-\tau\right)=o_p(1)
$$

by Lemma Q.1(iii).

(ii) Using the expansion from part (i), note that by Assumption A.6(ii) uniformly $X\in\mathcal{X}$:

$$
\begin{aligned}
& \sqrt{R}\left(q_{\tau}\left(\widehat{\boldsymbol{\psi}}_R;X\right)-q_{\tau}\left(\boldsymbol{\psi}^{\dagger};X\right)\right)\\
=\; & \nabla_{\boldsymbol{\theta}_m}m(X,\boldsymbol{\theta}_m^{\dagger})\sqrt{R}\left(\widehat{\boldsymbol{\theta}}_{m,R}-\boldsymbol{\theta}_m^{\dagger}\right)+\nabla_{\boldsymbol{\theta}_{\sigma}}\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)q_{\tau}(\epsilon_{t+1})\sqrt{R}\left(\widehat{\boldsymbol{\theta}}_{\sigma,R}-\boldsymbol{\theta}_{\sigma}^{\dagger}\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)\left(H(\tau)^{-1}\frac{1}{\sqrt{R}}\sum_{t=1}^{R-1}\left(1\left\{\epsilon_{t+1}\le q_{\tau}\left(\epsilon_{t+1}\right)\right\}-\tau\right)\right.\\
&-H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}f_{\epsilon}(q_{\tau}\left(\epsilon_{t+1}\right))\frac{\nabla_{\boldsymbol{\theta}_m}m(X,\boldsymbol{\theta}_m^{\dagger})}{\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\sqrt{R}\left(\widehat{\boldsymbol{\theta}}_{m,R}-\boldsymbol{\theta}_m^{\dagger}\right)\\
&\left.-H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}f_{\epsilon}(q_{\tau}\left(\epsilon_{t+1}\right))\epsilon_{t+1}\frac{\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}{\sigma^2\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\sqrt{R}\left(\widehat{\boldsymbol{\theta}}_{\sigma,R}-\boldsymbol{\theta}_{\sigma}^{\dagger}\right)\right)+o_p(1),
\end{aligned}
$$

where the $o_p(1)$ term holds uniformly in $X\in\mathcal{X}$ by A.3(iii). Now, using the expansions from Assumption A.6(ii), we obtain:

$$
\begin{aligned}
&\sqrt{R}\left(q_{\tau}\left(\widehat{\boldsymbol{\psi}}_R;X\right)-q\left(\boldsymbol{\psi}^{\dagger};X\right)\right)\\
=\;&\nabla_{\boldsymbol{\theta}_m}m(X,\boldsymbol{\theta}_m^{\dagger})\left(\mathbf{M}_m^{-1}\frac{1}{\sqrt{R}}\sum_{s=1}^{R-1}\nabla_{\boldsymbol{\theta}_m}\zeta(y_{s+1},X_s,\boldsymbol{\theta}_m^{\dagger},\boldsymbol{\theta}_{\sigma}^{\dagger})\right)\\
&+\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)q_{\tau}(\epsilon_{t+1})\left(\mathbf{M}_{\sigma}^{-1}\frac{1}{\sqrt{R}}\sum_{s=1}^{R-1}\nabla_{\boldsymbol{\theta}_\sigma}\zeta(y_{s+1},X_s,\boldsymbol{\theta}_m^{\dagger},\boldsymbol{\theta}_{\sigma}^{\dagger})\right)\\
&+\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)\left(H(\tau)^{-1}\frac{1}{\sqrt{R}}\sum_{t=1}^{R-1}\left(1\left\{\epsilon_{t+1}\le q_{\tau}\left(\epsilon_{t+1}\right)\right\}-\tau\right)\right.\\
&-H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}f_{\epsilon}(q_{\tau}\left(\epsilon_{t+1}\right))\frac{\nabla_{\boldsymbol{\theta}_m}m(X,\boldsymbol{\theta}_m^{\dagger})}{\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\left(\mathbf{M}_m^{-1}\frac{1}{\sqrt{R}}\sum_{s=1}^{R-1}\nabla_{\boldsymbol{\theta}_m}\zeta(y_{s+1},X_s,\boldsymbol{\theta}_m^{\dagger},\boldsymbol{\theta}_{\sigma}^{\dagger})\right)\\
&\left.-H(\tau)^{-1}\frac{1}{R}\sum_{t=1}^{R-1}f_{\epsilon}(q_{\tau}\left(\epsilon_{t+1}\right))\epsilon_{t+1}\frac{\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}{\sigma^2\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\left(\mathbf{M}_{\sigma}^{-1}\frac{1}{\sqrt{R}}\sum_{s=1}^{R-1}\nabla_{\boldsymbol{\theta}_\sigma}\zeta(y_{s+1},X_s,\boldsymbol{\theta}_m^{\dagger},\boldsymbol{\theta}_{\sigma}^{\dagger})\right)\right)+o_p(1).
\end{aligned}
$$

Given A.6(iii)-(iv), we have by a uniform law of large numbers for strong mixing observations uniformly in $X\in\mathcal{X}$:

$$
\frac{1}{R}\sum_{t=1}^{R-1}f_{\epsilon}(q_{\tau}\left(\epsilon_{t+1}\right))\frac{\nabla_{\boldsymbol{\theta}_m}m\left(X,\boldsymbol{\theta}_m^{\dagger}\right)}{\sigma^2\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\xrightarrow{p}\mathrm{E}\left(f_{\epsilon}(q_{\tau}\left(\epsilon_{t+1}\right))\frac{\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}{\sigma^2\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\right)
$$

and

$$
\frac{1}{R}\sum_{t=1}^{R-1}f_{\epsilon}(q_{\tau}\left(\epsilon_{t+1}\right))\epsilon_{t+1}\frac{\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}{\sigma^2\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\xrightarrow{p}\mathrm{E}\left(f_{\epsilon}(q_{\tau}\left(\epsilon_{t+1}\right))\epsilon_{t+1}\frac{\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}{\sigma^2\left(X,\boldsymbol{\theta}_{\sigma}^{\dagger}\right)}\right).
$$

■

S.3 Bootstrap Two-Sided Case

In this section, we extend the bootstrap statistic from Section 3.3 of the main text to two-sided intervals $[\tau_L,\tau_U]$ and then briefly comment on the two-sided extension of the multiple models and intervals test in Section 4 of the paper.

Now, in the two-sided, single interval case $[\tau_L,\tau_U]$ we have that:

$$
\begin{aligned}
&\widehat{S}^{*}_{P,R}([\tau_L,\tau_U])\\
=\;&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-l_P-1}\varepsilon_t\left(\sum_{s=t}^{t+l_P}\left(\widehat{A}_{1,P,R,s}([\tau_L,\tau_U])-\widehat{A}_{2,P,R,s}([\tau_L,\tau_U])\right)+\left(\widehat{B}_{1,P,R,s}([\tau_L,\tau_U])-\widehat{B}_{2,P,R,s}([\tau_L,\tau_U])\right)\right)\\
&+\frac{\sqrt{P}}{R}\sum_{t=1}^{R-l_R-1}\eta_t\sum_{s=t}^{t+l_R}\left(\widehat{D}_{1,P,R,s}([\tau_L,\tau_U])-\widehat{D}_{2,P,R,s}([\tau_L,\tau_U])\right),
\end{aligned}
\tag{S.28}
$$

where for $j=1,2$:

$$
\widehat{A}_{j,P,R,t}([\tau_L,\tau_U])
$$

$$= 1\{\mathbf{X}_t \in \mathcal{X}\} \left(L\left(\widehat{C}_{j,P,R}\left([\tau_L, \tau_U]; \mathbf{X}_t\right) - (\tau_U - \tau_L)\right) \right.$$
$$\left. -\frac{1}{P}\sum_{s=R}^{T-1} L\left(\widehat{C}_{j,P,R}\left([\tau_L, \tau_U]; \mathbf{X}_s\right) - (\tau_U - \tau_L)\right)\right)$$

$$\widehat{B}_{j,P,R,t}([\tau_L, \tau_U])$$
$$= 1\{\mathbf{X}_t \in \mathcal{X}\}\nabla L\left(\widehat{C}_{j,P,R}\left([\tau_L, \tau_U]; \mathbf{X}_t\right) - (\tau_U - \tau_L)\right)\left(1\left\{X'_{j,t}\widehat{\boldsymbol{\beta}}_{j,R}(\tau_L) \leq y_{t+1} \leq X'_{j,t}\widehat{\boldsymbol{\beta}}_{j,R}(\tau_U)\right\}\right.$$
$$\left. - \left(\widehat{F}_{t+1,P}\left(X'_{j,t}\widehat{\boldsymbol{\beta}}_{j,R}(\tau_U)|\mathbf{X}_t\right)\right) - \widehat{F}_{t+1,P}\left(X'_{j,t}\widehat{\boldsymbol{\beta}}_{j,R}(\tau_L)|\mathbf{X}_t\right)\right)$$

and:

$$\begin{aligned}\widehat{D}_{j,P,R,t}([\tau_L, \tau_U]) &= \widehat{\Lambda}_{j,R}(\tau_U)\left(\widehat{H}^{-1}_{j,P,R}(\tau_U)X_{j,t}1\left(\left\{y_{t+1} \leq X'_{j,t}\widehat{\boldsymbol{\beta}}_{j,R}(\tau_U)\right\}\right) - \tau_U\right) \\ &\quad -\widehat{\Lambda}_{j,P,R}(\tau_R)\left(\widehat{H}^{-1}_{j,R}(\tau_R)X_{j,t}1\left(\left\{y_{t+1} \leq X'_{j,t}\widehat{\boldsymbol{\beta}}_{j,R}(\tau_R)\right\}\right) - \tau_R\right),\end{aligned}$$

and the terms $\widehat{\Lambda}_{j,P,R}(\cdot)$, $\widehat{H}^{-1}_{j,R}(\cdot)$, and $\widehat{F}_{t+1,P}(\cdot|\mathbf{X}_t)$ are as defined in the main text.

Similarly, for the multiple models and intervals case we may compare the relative conditional coverage error of models 1 and 2 over M two-sided intervals, $[\tau_{iL}, \tau_{iU}]$, $i = 1, ..., M$. These intervals could potentially be overlapping in the sense that one could potentially compare models over, say, intervals $[0.1, 0.3], [0.2, 0.4]$ and so on. The null hypothesis is:

$$H_0^{RC}: \max_{j=2,...,J}\max_{i=1,...,M} \mathrm{E}\left(\left(L\left(\mathcal{E}_1\left([\tau_{i,L}, \tau_{i,U}]; \mathbf{X}^j_t\right)\right) - L\left(\mathcal{E}_j\left([\tau_{i,L}, \tau_{i,U}]; \mathbf{X}^j_t\right)\right)\right)1\left\{\mathbf{X}^j_t \in \mathcal{X}\right\}\right) \leq 0$$

versus:

$$H_A^{RC}: \max_{j=2,...,J}\max_{i=1,...,M} \mathrm{E}\left(\left(L\left(\mathcal{E}_1\left([\tau_{i,L}, \tau_{i,U}]; \mathbf{X}^j_t\right)\right) - L\left(\mathcal{E}_j\left([\tau_{i,L}, \tau_{i,U}]; \mathbf{X}^j_t\right)\right)\right)1\left\{\mathbf{X}^j_t \in \mathcal{X}\right\}\right) > 0.$$

In analogy to before, for each competitor model, j, and interval, i, let (suppressing again the P,R dependence):

$$\begin{aligned}&\widehat{A}_{j,t}\left([\tau_{i,U}, \tau_{i,L}], \mathbf{X}^j_t\right) \\ = &\left(L\left(\widehat{\mathcal{E}}_{j,P,R}\left([\tau_{i,U}, \tau_{i,L}], \mathbf{X}^j_t\right)\right) - \frac{1}{P}\sum_{s=R}^{T-1} L\left(\widehat{\mathcal{E}}_{j,P,R}\left([\tau_{i,U}, \tau_{i,L}], \mathbf{X}^j_s\right)\right)\right)1\left\{\mathbf{X}^j_t \in \mathcal{X}^j\right\}\end{aligned}$$

while $\widehat{B}_{j,t}\left([\tau_{i,U}, \tau_{i,L}], \mathbf{X}^j_t\right)$ is the term associated to nonparametric estimation error:

$$\begin{aligned}&\widehat{B}_{j,t}\left([\tau_{i,U}, \tau_{i,L}], \mathbf{X}^j_t\right) \\ = &\nabla L\left(\widehat{\mathcal{E}}_{j,P,R}\left(\mathbf{X}^j_t, [\tau_{i,U}, \tau_{i,L}]\right)\right)\left(1\left\{X'_{j,t}\widehat{\beta}_{j,R}(\tau_{i,L}) \leq y_{t+1} \leq X'_{j,t}\widehat{\beta}_{j,R}(\tau_{i,U})\right\}\right. \\ &\left. -\left(\widehat{F}_{t+1,P}\left(X'_{j,t}\widehat{\beta}_{j,R}(\tau_{i,U})|\mathbf{X}^j_t\right) - \widehat{F}_{t+1,P}\left(X'_{j,t}\widehat{\beta}_{j,R}(\tau_{i,L})|\mathbf{X}^j_t\right)\right)\right)1\left\{\mathbf{X}^j_t \in \mathcal{X}^j\right\}.\end{aligned}$$

In addition, for $t = 1, ..., R-1$, we let $\widehat{D}_{j,t}([\tau_{i,U}, \tau_{i,L}])$ denote the parametric quantile estimation error for model j and interval $[\tau_{i,U}, \tau_{i,L}]$. That is:

$$\begin{aligned}\widehat{D}_{j,t}([\tau_{i,U}, \tau_{i,L}]) &= \widehat{\Lambda}_{j,P,R}(\tau_{i,U})\left(\widehat{H}^{-1}_{j,R}(\tau_{i,U})X_{j,t}\left(1\left\{y_{t+1} \leq X'_{j,t}\widehat{\beta}_{j,R}(\tau_{i,U})\right\} - \tau_{i,U}\right)\right) \\ &\quad -\widehat{\Lambda}_{j,P,R}(\tau_{i,L})\left(\widehat{H}^{-1}_{j,R}(\tau_{i,L})X_{j,t}\left(1\left\{y_{i+1} \leq X'_{j,i}\widehat{\beta}_{j,R}(\tau_{i,L})\right\} - \tau_{i,L}\right)\right).\end{aligned}$$

S.4 Bootstrap Location Scale Model

In this section, we provide the formula of the bootstrap statistic in the case of the location scale model and outline the differences in the proof of Theorem 2 for the case of the nonlinear location scale model. Focusing on model 1 and one-sided intervals $(0, \tau]$, recall that in the location scale case:

$$\widetilde{\Lambda}_{1,1}(\tau) = \mathrm{E}\left(1\{\mathbf{X}_t \in \mathcal{X}\}\nabla L\left(C_1\left(\boldsymbol{\psi}^\dagger_1(\tau); \mathbf{X}_t\right) - \tau\right)f_{t+1}(q_\tau(\boldsymbol{\psi}^\dagger_1(\tau); X_{j,t})|\mathbf{X}_t)\nabla_{\boldsymbol{\theta}_m} m(X_{1,t}, \boldsymbol{\theta}^\dagger_{m_1})\right),$$

$$\widetilde{\Lambda}_{1,2}(\tau) = \mathrm{E}\left(1\{\mathbf{X}_t \in \mathcal{X}\}\nabla L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right) f_{t+1}(q_\tau(\boldsymbol{\psi}_1^{\dagger}(\tau);X_{1,t})|\mathbf{X}_t)\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X_{1,t},\boldsymbol{\theta}_{\sigma_1}^{\dagger}\right) q_\tau(\epsilon_{1,t+1})\right),$$

$$\widetilde{\Lambda}_{1,3}(\tau) = \mathrm{E}\left(1\{\mathbf{X}_t \in \mathcal{X}\}\nabla L\left(C_1\left(\boldsymbol{\psi}_1^{\dagger}(\tau);\mathbf{X}_t\right)-\tau\right) f_{t+1}(q_\tau(\boldsymbol{\psi}_1^{\dagger}(\tau);X_{1,t})|\mathbf{X}_t)\sigma\left(X_{1,t},\boldsymbol{\theta}_{\sigma_1}^{\dagger}\right)\right),$$

as well as:

$$\epsilon_{1,t+1} = \frac{y_{t+1}-m(X_{1,t},\boldsymbol{\theta}_{m_1}^{\dagger})}{\sigma(X_{1,t},\boldsymbol{\theta}_{\sigma_1}^{\dagger})}.$$

Thus:

$$\begin{aligned}
&\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}\varphi(\boldsymbol{\psi}_1^{\dagger}(\tau);y_{t+1},X_{1,t})\\
=\;&\frac{\sqrt{P}}{R}\sum_{t=1}^{R-1}\Bigg(\widetilde{\Lambda}_{1,1}(\tau)\boldsymbol{M}_{1,m}^{-1}\nabla_{\boldsymbol{\theta}_m}\zeta(y_{t+1},X_{1,t},\boldsymbol{\theta}_{m_1}^{\dagger},\boldsymbol{\theta}_{\sigma_1}^{\dagger})+\widetilde{\Lambda}_{1,2}(\tau)\boldsymbol{M}_{1,\sigma}^{-1}\nabla_{\boldsymbol{\theta}_\sigma}\zeta(y_{t+1},X_{1,t},\boldsymbol{\theta}_{m_1}^{\dagger},\boldsymbol{\theta}_{\sigma_1}^{\dagger})\\
&+\widetilde{\Lambda}_{1,3}(\tau)\Bigg(\left(H_1(\tau)^{-1}\left(1\{\epsilon_{1,t+1}\le q_\tau(\epsilon_{1,t+1})\}-\tau\right)\right)-\Bigg(H_1(\tau)^{-1}\mathrm{E}\Bigg(f_{\epsilon_1}(q_\tau(\epsilon_{1,t+1}))\frac{\nabla_{\boldsymbol{\theta}_m}m(X_t,\boldsymbol{\theta}_{m_1}^{\dagger})}{\sigma\left(X_t,\boldsymbol{\theta}_{\sigma_1}^{\dagger}\right)}\Bigg)\\
&\times\boldsymbol{M}_{1,m}^{-1}\nabla_{\boldsymbol{\theta}_m}\zeta(y_{t+1},X_{1,t},\boldsymbol{\theta}_{m_1}^{\dagger},\boldsymbol{\theta}_{\sigma_1}^{\dagger})\Bigg)\\
&-H_1(\tau)^{-1}\mathrm{E}\Bigg(f_{\epsilon_1}(q_\tau(\epsilon_{1,t+1}))\epsilon_{1,t+1}\frac{\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X_{1,t},\boldsymbol{\theta}_{\sigma_1}^{\dagger}\right)}{\sigma\left(X_{1,t},\boldsymbol{\theta}_{\sigma_1}^{\dagger}\right)}\Bigg)\boldsymbol{M}_{1,\sigma}^{-1}\nabla_{\boldsymbol{\theta}_\sigma}\zeta(y_{t+1},X_{1,t},\boldsymbol{\theta}_{m_1}^{\dagger},\boldsymbol{\theta}_{\sigma_1}^{\dagger})\Bigg)\Bigg)
\end{aligned}$$

with $H_1(\tau)$ as defined in A.3 of the main text and:

$$\boldsymbol{M}_{1,m} = \mathrm{E}\left(\nabla_{\boldsymbol{\theta}_m}^{(2)}\zeta(y_{t+1},X_{1,t},\boldsymbol{\theta}_{m_1}^{\dagger},\boldsymbol{\theta}_{\sigma_1}^{\dagger})\right)$$

$$\boldsymbol{M}_{1,\sigma} = \mathrm{E}\left(\nabla_{\boldsymbol{\theta}_\sigma}^{(2)}\zeta(y_{t+1},X_{1,t},\boldsymbol{\theta}_{m_1}^{\dagger},\boldsymbol{\theta}_{\sigma_1}^{\dagger})\right).$$

Now, define:

$$\widehat{\omega}_{1,t}(\tau) = 1\{\mathbf{X}_t \in \mathcal{X}\}\nabla L\left(\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_t\right)-\tau\right)\widehat{f}_{t+1}(q_\tau(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);X_{j,t})|\mathbf{X}_t)$$

where $\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_t\right)$ is again constructed as in the main text, while:

$$\widehat{f}_{t+1}(q_\tau(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);X_{j,t})|\mathbf{X}_t) = \frac{\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau_k);\mathbf{X}_t\right)-\widehat{C}_{1,P}\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau_{k-1});\mathbf{X}_t\right)}{q_\tau(\widehat{\boldsymbol{\psi}}_{1,R}(\tau_k);X_{j,t})-q_\tau(\widehat{\boldsymbol{\psi}}_{1,R}(\tau_{k-1});X_{j,t})}$$

for some $\tau_{k-1} < \tau < \tau_k$. In addition, let:

$$\widehat{\widetilde{\Lambda}}_{1,1,P}(\tau) = \frac{1}{P}\sum_{t=R}^{T-1}\widehat{\omega}_{1,t}(\tau)\nabla_{\boldsymbol{\theta}_m}m(X_{1,t},\boldsymbol{\theta}_{m_1}^{\dagger})$$

$$\widehat{\widetilde{\Lambda}}_{1,2,P}(\tau) = \frac{1}{P}\sum_{t=R}^{T-1}\widehat{\omega}_{1,t}(\tau)\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X_{1,t},\boldsymbol{\theta}_{\sigma_1}^{\dagger}\right)q_\tau(\widehat{\epsilon}_{1,t+1})$$

$$\widehat{\widetilde{\Lambda}}_{1,3,P}(\tau) = \frac{1}{P}\sum_{t=R}^{T-1}\widehat{\omega}_{1,t}(\tau)\sigma\left(X_{1,t},\boldsymbol{\theta}_{\sigma_1}^{\dagger}\right)\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X_{1,t},\boldsymbol{\theta}_{\sigma_1}^{\dagger}\right)$$

where $q_\tau(\widehat{\epsilon}_{1,t+1}) = \widehat{\beta}_{1,R}(\tau)$. Finally, let:

$$\widehat{\boldsymbol{M}}_{1,m,R} = \frac{1}{R}\sum_{t=1}^{R-1}\nabla_{\boldsymbol{\theta}_m}^{(2)}\zeta(y_{t+1},X_{1,t},\boldsymbol{\theta}_{m_1}^{\dagger},\boldsymbol{\theta}_{\sigma_1}^{\dagger})$$

with $\widehat{\boldsymbol{M}}_{1,\sigma,R}$ defined analogously, and:

$$\widehat{\epsilon}_{1,t+1} = \frac{y_{t+1}-m(X_{1,t},\widehat{\boldsymbol{\theta}}_{1,m,R})}{\sigma(X_{1,t},\widehat{\boldsymbol{\theta}}_{1,\sigma,R})},$$

so that:

$$\widehat{V}_{1,m,P}=\frac{1}{P}\sum_{t=R}^{T-l}\widehat{f}_{\epsilon_1}\left(q_\tau\left(\widehat{\epsilon}_{1,t+1}\right)\right)\frac{\nabla_{\boldsymbol{\theta}_m}m\left(X_{1,t},\widehat{\boldsymbol{\theta}}_{1,R,m}\right)}{\sigma\left(X_{1,t},\widehat{\boldsymbol{\theta}}_{1,R,\sigma}\right)}$$

$$\widehat{V}_{1,\sigma,P}=\frac{1}{P}\sum_{t=R}^{T-l}\widehat{f}_{\epsilon_1}\left(q_\tau\left(\widehat{\epsilon}_{1,t+1}\right)\right)\frac{\nabla_{\boldsymbol{\theta}_\sigma}\sigma\left(X_{1,t},\widehat{\boldsymbol{\theta}}_{1,R,\sigma}\right)}{\sigma\left(X_{1,t},\widehat{\boldsymbol{\theta}}_{1,R,\sigma}\right)}\widehat{\epsilon}_{1,t+1}.$$

Here:

$$\widehat{f}_{\epsilon_1}\left(q_\tau\left(\widehat{\epsilon}_{1,t+1}\right)\right)=\frac{1}{Rh_\epsilon}\sum_{i=1}^{R-1}K((q_\tau\left(\widehat{\epsilon}_{1,t+1}\right)-q_\tau\left(\widehat{\epsilon}_{1,i+1}\right))/h_\epsilon)$$

for some bandwidth sequence h_ϵ satisfying $h_\epsilon\to 0$ and $Rh_\epsilon\to\infty$. Then, letting the corresponding quantities for model 2 be defined accordingly, we have that:

$$S_{P,R}^{*LS}(\tau)=S_{P,R}^{*LS,1}(\tau)+S_{P,R}^{*LS,2}(\tau)$$

where, setting again $l_P=l_R=l$ for notational simplicity, we have that:

$$\begin{aligned}
&S_{P,R}^{*LS,1}(\tau)\\
=\;&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-l-1}\varepsilon_t\left(\sum_{i=t}^{t+l}1\{\mathbf{X}_i\in\mathcal{X}\}\left(\left(L\left(\widehat{C}_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_i\right)-\tau\right)-L\left(\widehat{C}_2\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_i\right)-\tau\right)\right)\right.\right.\\
&+\left(\nabla L\left(\widehat{C}_1\left(\widehat{\boldsymbol{\psi}}_{1,R}(\tau);\mathbf{X}_i\right)-\tau\right)\left(1\left\{y_{i+1}\le q_\tau\left(\widehat{\boldsymbol{\psi}}_{1,R};X_{1,i}\right)\right\}-\widehat{F}_{t+1,P}\left(q_\tau\left(\widehat{\boldsymbol{\psi}}_{1,R};X_{1,i}\right)|\mathbf{X}_i\right)\right)\right.\\
&\left.\left.\left.-\nabla L\left(\widehat{C}_2\left(\widehat{\boldsymbol{\psi}}_{2,R}(\tau);\mathbf{X}_i\right)-\tau\right)\left(1\left\{y_{i+1}\le q_\tau\left(\widehat{\boldsymbol{\psi}}_{2,R};X_{2,i}\right)\right\}-\widehat{F}_{t+1,P}\left(q_\tau\left(\widehat{\boldsymbol{\psi}}_{2,R}\left(\tau\right);X_{2,i}\right)|\mathbf{X}_i\right)\right)\right)\right)\right)
\end{aligned}$$

and

$$\begin{aligned}
&S_{P,R}^{*LS,2}(\tau)\\
=\;&\frac{\sqrt{P}}{R}\sum_{t=1}^{R-l-1}\eta_t\left(\sum_{i=t}^{t+l}\left(\widehat{\widehat{\Lambda}}_{1,1,P}(\tau)\widehat{\mathbf{M}}_{1,m,R}^{-1}\nabla_{\theta_m}\zeta\left(y_{i+1},X_{1,i},\widehat{\boldsymbol{\theta}}_{1,R,m},\widehat{\boldsymbol{\theta}}_{1,R,\sigma}\right)\right.\right.\\
&+\widehat{\widehat{\Lambda}}_{1,2,P}(\tau)\widehat{\mathbf{M}}_{1,\sigma,R}^{-1}\zeta\left(y_{i+1},X_{1,i},\widehat{\boldsymbol{\theta}}_{1,R,m},\widehat{\boldsymbol{\theta}}_{1,R,\sigma}\right)+\widehat{\widehat{\Lambda}}_{1,3,P}\left(\widehat{H}_1^{-1}\left(\tau\right)\left(1\left\{\widehat{\epsilon}_{1,i+1}\le q_\tau\left(\widehat{\epsilon}_{1,i+1}\right)\right\}-\tau\right)\right.\\
&-\widehat{H}_1^{-1}\left(\tau\right)\widehat{V}_{1,m,P}\widehat{\mathbf{M}}_{1,m,R}^{-1}\nabla_{\theta_m}\zeta\left(y_{i+1},X_{1,i},\widehat{\boldsymbol{\theta}}_{1,R,m},\widehat{\boldsymbol{\theta}}_{1,R,\sigma}\right)\\
&\left.\left.-\widehat{H}_1^{-1}\left(\tau\right)\widehat{V}_{1,\sigma,P}\widehat{\mathbf{M}}_{1,\sigma,R}^{-1}\nabla_{\theta_\sigma}\zeta\left(y_{i+1},X_{1,i},\widehat{\boldsymbol{\theta}}_{1,R,m},\widehat{\boldsymbol{\theta}}_{1,R,\sigma}\right)\right)\right)\\
&-\left(\widehat{\widehat{\Lambda}}_{2,1,P}(\tau)\widehat{\mathbf{M}}_{2,m,R}^{-1}\nabla_{\theta_m}\zeta\left(y_{i+1},X_{2,i},\widehat{\boldsymbol{\theta}}_{2,R,m},\widehat{\boldsymbol{\theta}}_{2,R,\sigma}\right)\right.\\
&+\widehat{\widehat{\Lambda}}_{2,2,P}(\tau)\widehat{\mathbf{M}}_{2,\sigma,R}^{-1}\zeta\left(y_{i+1},X_{2,i},\widehat{\boldsymbol{\theta}}_{2,R,m},\widehat{\boldsymbol{\theta}}_{2,R,\sigma}\right)+\widehat{\widehat{\Lambda}}_{2,3,P}\left(\widehat{H}_1^{-1}\left(\tau\right)\left(1\left\{\widehat{\epsilon}_{2,i+1}\le q_\tau\left(\widehat{\epsilon}_{2,i+1}\right)\right\}-\tau\right)\right.\\
&-\widehat{H}_2^{-1}\left(\tau\right)\widehat{V}_{2,m,P}\widehat{\boldsymbol{M}}_{2,m,R}^{-1}\nabla_{\theta_m}\zeta\left(y_{i+1},X_{2,i},\widehat{\boldsymbol{\theta}}_{2,R,m},\widehat{\boldsymbol{\theta}}_{2,R,\sigma}\right)\\
&\left.\left.\left.-\widehat{H}_2^{-1}\left(\tau\right)\widehat{V}_{2,\sigma,P}\widehat{\mathbf{M}}_{2,\sigma,R}^{-1}\nabla_{\theta_\sigma}\zeta\left(y_{i+1},X_{2,i},\widehat{\boldsymbol{\theta}}_{2,R,m},\widehat{\boldsymbol{\theta}}_{2,R,\sigma}\right)\right)\right)\right)
\end{aligned}$$

where ε_t and η_t are again i.i.d. random variables independent of the data and drawn from distributions $N(0,1/l)$. Under h_ϵ satisfying $h_\epsilon\to 0$ and $Rh_\epsilon\to\infty$ and the assumptions A.1, A.2, A.4, A.5 and A.6, the results from Theorem 2 then follow using similar arguments to the proof of Theorem 2.

S.5 Recursive Estimation Scheme

For notational simplicity, consider again the case of quantile regression. Also, for brevity, we outline only the test statistic for the pairwise comparison and one-sided interval case. The key difference between the fixed and the recursive estimation scheme is that for the latter, at each time $t \geq R$, we re-estimate the quantile models, using the newly available data. That is, for $t \geq R$, we define:

$$\widehat{\boldsymbol{\beta}}_{j,t}(\tau) = \arg\min_{\boldsymbol{\beta}\in\mathcal{B}} \frac{1}{t}\sum_{s=1}^{t} \rho_\tau\left(y_{s+1} - X'_{j,s}\boldsymbol{\beta}\right),$$

$q_\tau(\widehat{\boldsymbol{\beta}}_{j,t}; X_{j,t}) = X'_{j,t}\widehat{\boldsymbol{\beta}}_{j,t}(\tau)$. For $t \geq R$, the associated conditional coverage is defined as

$$\begin{aligned}&\widehat{C}^r_{j,P,t}(\tau;\mathbf{X}_t)\\ =\;& \frac{1}{Ph^d}\sum_{s=R}^{T-1}\frac{1}{\widehat{f}_X(\mathbf{X}_t)}1\left\{y_{s+1} \leq q_\tau(\widehat{\boldsymbol{\beta}}_{j,t}; X_{j,t})\right\}\mathbf{K}\left(\frac{\mathbf{X}_s - \mathbf{X}_t}{h}\right).\end{aligned}$$

Letting $\widehat{\mathcal{E}}^r_{j,P,t}(\tau;\mathbf{X}_t) = \widehat{C}^r_{j,P,t}(\tau;\mathbf{X}_t) - \tau$, define also

$$\widehat{S}^r_{P,R}(\tau) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1} L\left(\widehat{\mathcal{E}}^r_{1,P,t}(\tau;\mathbf{X}_t)\right) - L\left(\widehat{\mathcal{E}}^r_{2,P,t}(\tau;\mathbf{X}_t)\right), \tag{S.29}$$

with $\mathbf{X}_t = X_{1,t} \cup X_{2,t}$. Heuristically, in order to obtain the "recursive scheme counterpart" of Theorem 1, we need to strengthen the statements Lemma Q.1 so that for all $\tau \in \mathcal{T}$ and $j = 1, ..., J$:

(i*) $\sup_{t\geq R}\left\|\widehat{\boldsymbol{\beta}}_{j,t}(\tau) - \boldsymbol{\beta}^\dagger_j(\tau)\right\| = o_p(1)$

(ii*) For all $t \geq R$, the following linear expansion holds:

$$\begin{aligned}&\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\widehat{\boldsymbol{\beta}}_{j,t}(\tau) - \boldsymbol{\beta}^\dagger_j(\tau)\right)\\ =\;& H_j^{-1}(\tau)\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{s=1}^{t} X'_s\left(1\left\{y_{s+1} \leq X'_s\boldsymbol{\beta}^\dagger_j(\tau)\right\} - \tau\right) + o_p(1).\end{aligned}$$

Using the asymptotic linear representation from Theorem 1 in CASE I, we have that

$$\begin{aligned}&\widehat{S}^r_{P,R}(\tau)\\ =\;& \frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\underbrace{1\{\mathbf{X}_t \in \mathcal{X}\}L\left(C_1(\tau;\mathbf{X}_t) - \tau\right)}_{A_{1,t}(\tau)} - \underbrace{1\{\mathbf{X}_t \in \mathcal{X}\}L\left(C_2(\tau;\mathbf{X}_t) - \tau\right)}_{A_{2,t}(\tau)}\right)\\ &+\left(\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\underbrace{1\{\mathbf{X}_t \in \mathcal{X}\}\nabla L\left(C_1(\tau;\mathbf{X}_t) - \tau\right)\left(1\{y_{t+1} \leq q_\tau(\boldsymbol{\beta}^\dagger_1; X_{1,t})\} - F_{t+1}(q_\tau(\boldsymbol{\beta}^\dagger_1; X_{1,t})|\mathbf{X}_t)\right)}_{B_{1,t}(\tau)}\right.\\ &\left.-\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\underbrace{1\{\mathbf{X}_t \in \mathcal{X}\}\nabla L\left(C_2(\tau;\mathbf{X}_t) - \tau\right)\left(1\{y_{t+1} \leq q_\tau(\boldsymbol{\beta}^\dagger_2; X_{2,t})\} - F_{t+1}(q_\tau(\boldsymbol{\beta}^\dagger_2; X_{2,t})|\mathbf{X}_t)\right)}_{B_{2,t}(\tau)}\right)\\ &+\left(\Lambda_1(\tau)H_1^{-1}(\tau)\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{s=1}^{t}\underbrace{X'_{1,s}\left(1\left\{y_{s+1} \leq X'_{1,s}\boldsymbol{\beta}^\dagger_{1,t}(\tau)\right\} - \tau\right)}_{D_{1,s}(\tau)}\right.\\ &\left.-\Lambda_2(\tau)H_2^{-1}(\tau)\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{s=1}^{t}\underbrace{X'_{2,s}\left(1\left\{y_{s+1} \leq X'_{2,s}\boldsymbol{\beta}^\dagger_{2,t}(\tau)\right\} - \tau\right)}_{D_{2,s}(\tau)}\right) + o_p(1)\end{aligned} \tag{S.30}$$

with $\Lambda_j(\tau)$ and $H_j(\tau)$, $j = 1, 2$, defined as the main text. The asymptotic variances of the first two terms on the RHS of (S.30) are as in the fixed estimation scheme. As for the asymptotic variance of the third term, by

Lemma A5 in West (1996),

$$\begin{aligned} & \text{Avar}\left(\Lambda_1(\tau) H_1^{-1}(\tau)\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{s=1}^{t} D_{1,s}(\tau) - \Lambda_2(\tau) H_2^{-1}(\tau)\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\frac{1}{t}\sum_{s=1}^{t} D_{2,s}(\tau)\right) \\ = \; & 2\Pi\Lambda_1(\tau) H_1^{-1}(\tau)\Sigma_{D_{1,\tau}D_{1,\tau}} H_1^{-1}(\tau)\Lambda_1(\tau) + 2\Pi\Lambda_2(\tau) H_2^{-1}(\tau)\Sigma_{D_{2,\tau}D_{2,\tau}} H_2^{-1}(\tau)\Lambda_2(\tau) \\ & -4\Pi\Lambda_1(\tau) H_1^{-1}(\tau)\Sigma_{D_{1,\tau}D_{2,\tau}} H_2^{-1}(\tau)\Lambda_2(\tau) \end{aligned}$$

with $\Pi = \left(1 - \pi^{-1}\ln(1+\pi)\right)$, $\pi = \lim_{P,R\to\infty}\frac{P}{R}$, and $\Sigma_{D_{1,\tau}D_{1,\tau}} = \sum_{i=-\infty}^{\infty} \mathrm{E}\left(D_{1,t}(\tau)D_{1,t+i}(\tau)'\right)$. Also, using model 1 as an example, by Lemma A6 in West (1996):

$$\begin{aligned} & \text{Acov}\left(\Lambda_1(\tau) H_1^{-1}(\tau)\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\frac{1}{t}\sum_{s=1}^{t} D_{1,s}(\tau)\times A_t(\tau)\right)\right) \\ = \; & \Pi\Lambda_1(\tau) H_1^{-1}(\tau)\Sigma_{D_{1,\tau}A_\tau} \end{aligned}$$

with $A_t(\tau) = (A_{1,t}(\tau) - A_{2,t}(\tau))$ and $\Sigma_{D_{1,\tau},A_\tau(\tau)} = \sum_{i=-\infty}^{\infty} \mathrm{E}\left(D_{1,t}(\tau)A_{t+i}(\tau)\right)$, and

$$\begin{aligned} & \text{Acov}\left(\Lambda_1(\tau) H_1^{-1}(\tau)\frac{1}{\sqrt{P}}\sum_{t=R}^{T-1}\left(\frac{1}{t}\sum_{s=1}^{t} D_{1,s}(\tau)\times B_t(\tau)\right)\right) \\ = \; & \Pi\Lambda_1(\tau) H_1^{-1}(\tau)\Sigma_{D_{1,\tau}B_\tau} \end{aligned}$$

with $B_t(\tau) = (B_{1,t}(\tau) - B_{2,t}(\tau))$ and $\Sigma_{D_{1,\tau}B_\tau} = \sum_{i=-\infty}^{\infty} \mathrm{E}\left(D_{1,t}(\tau)B_{t+i}(\tau)\right)$. Following the proof of Theorem 2 in Corradi and Swanson (2002), we can then modify the wild bootstrap statistic in such a way that it properly mimics quantile recursive estimation error. That is, let:

$$\begin{aligned} & \widehat{S}^{*,r}_{P,R}(\tau) \\ = \; & \frac{1}{\sqrt{P}}\sum_{t=R}^{T-l_P-1}\varepsilon_t\left(\sum_{i=t}^{t+l_P}\left(\left(\widehat{A}_{1,P,R,i}(\tau) - \widehat{A}_{2,P,R,i}(\tau)\right) + \left(\widehat{B}_{1,P,R,i}(\tau) - \widehat{B}_{2,P,R,i}(\tau)\right)\right.\right. \\ & \left.\left. + \left(\Pi\widehat{\Lambda}_{1,P,R}(\tau)\widehat{H}^{-1}_{1,P,R}(\tau)\widehat{\widehat{D}}_{1,P,R,i}(\tau) - \Pi\widehat{\Lambda}_{2,P,R}(\tau)\widehat{H}^{-1}_{2,P,R}(\tau)\widehat{\widehat{D}}_{2,P,R,i}(\tau)\right)\right)\right) \\ & + (2\Pi - \Pi)^{1/2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-l_P-1}\eta_t\left(\sum_{i=t}^{t+l_P}\left(\widehat{\Lambda}_{1,P,R}(\tau)\widehat{H}^{-1}_{1,R}(\tau)\widehat{\widehat{D}}_{1,P,R,i}(\tau) - \widehat{\Lambda}_{2,P,R}(\tau)\widehat{H}^{-1}_{2,R}(\tau)\widehat{\widehat{D}}_{2,P,R,i}(\tau)\right)\right), \end{aligned} \tag{S.31}$$

where, for $j = 1, 2$, the quantities $\widehat{A}_{j,P,R,i}(\tau)$, $\widehat{B}_{j,P,R,i}(\tau)$, $\widehat{H}_{j,R}(\tau)$ and $\widehat{\Lambda}_{j,P,R}(\tau)$ are defined as in the main text, while:

$$\widehat{\widehat{D}}_{j,P,R,t}(\tau) = X'_{j,t}\left(1\left\{y_{t+1} \le X'_{j,t}\widehat{\boldsymbol{\beta}}_{j,R}(\tau)\right\} - \tau\right).$$

Importantly, note that in the construction of $\widehat{A}_{j,P,R,i}(\tau)$, $\widehat{B}_{j,P,R,i}(\tau)$, and $\widehat{\widehat{D}}_{j,P,R,t}(\tau)$ we just need a consistent estimator for $\boldsymbol{\beta}^{\dagger}_j(\tau)$, and hence can just use an estimator based only on the first R observations as in the fixed estimation scheme.

As explained in detail in Corradi and Swanson (2002), the logic underlying $\widehat{S}^{*,r}_{P,R}(\tau)$ is the following: The first three terms on the RHS of (S.31) properly mimic, conditionally on the sample, the limiting behaviour of $\widehat{S}^{r}_{P,R}(\tau)$ but for the fact that the contribution to the variance of quantile recursive estimation error is multiplied by Π rather than by 2Π. This is why a correction term is applied to the statistic, namely:

$$(2\Pi - \Pi)^{1/2}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-l_P-1}\eta_t\left(\sum_{i=t}^{t+l_P}\left(\widehat{\Lambda}_{1,P,R}(\tau)\widehat{H}^{-1}_{1,R}(\tau)\widehat{\widehat{D}}_{1,P,R,i}(\tau) - \widehat{\Lambda}_{2,P,R}(\tau)\widehat{H}^{-1}_{2,R}(\tau)\widehat{\widehat{D}}_{2,P,R,i}(\tau)\right)\right).$$

In addition, note that ε_t does no longer exclusively multiply the expressions involving $\widehat{A}_{j,P,R,i}(\tau)$ and $\widehat{B}_{j,P,R,i}(\tau)$, $j = 1, 2$, but also the term involving $\widehat{\widehat{D}}_{1,P,R,i}(\tau)$ and $\widehat{\widehat{D}}_{2,P,R,i}(\tau)$ to capture the dependence of the parametric estimation error and the remaining components.

S.6 Extension to CAViaR Models

Conditional Autoregressive Value-at-Risk (CAViaR) models (Engle and Manganelli, 2004) are a popular choice to model dynamics in financial data. In this section, we outline how this model type can be accommodated in our set-up and used for our test(s). In particular, we will exploit the CAViaR representation of the linear GARCH model and the two-step quantile regression procedure proposed by Koenker and Xiao (2009). To this end, we focus on a linear GARCH(1,1) for notational simplicity, and drop again the model subscript j. That is, let:

$$y_{t+1} = \sigma_{t+1}\epsilon_{t+1},$$

where ϵ_{t+1} is i.i.d. with mean zero, unknown distribution, and strictly positive density everywhere. The variance parameter on the other hand is given by:

$$\sigma_{t+1} = \beta_0 + \beta_1\sigma_t + \gamma_1|y_t|. \tag{S.32}$$

Then, denoting $X_t = (1, \sigma_t, |y_t|)$, the τ-level quantile of y_t conditional on X_t is given by:

$$q_\tau(\boldsymbol{\psi}^\dagger; X_t) = \sigma_{t+1}F_\epsilon^{-1}(\tau) = (\beta_0 + \beta_1\sigma_t + \gamma_1|y_t|)\, F_\epsilon^{-1}(\tau) = X_t'\boldsymbol{\theta}^\dagger(\tau), \tag{S.33}$$

with $\boldsymbol{\theta}^\dagger(\tau) = \left((\beta_0, \beta_1, \gamma_1)\, F_\epsilon^{-1}(\tau)\right)'$, where $F_\epsilon(\cdot)$ denotes the marginal cumulative distribution function of ϵ_t, and so $F_\epsilon^{-1}(\tau)$ denotes the τ unconditional quantile of ϵ_t. Given (S.32) and (S.33), we can now obtain a CAViaR representation for $q_\tau(\boldsymbol{\psi}^\dagger; X_t)$ as:

$$q_\tau(\boldsymbol{\psi}^\dagger; X_t) = \beta_0^*(\tau) + \beta_1 q_\tau(\boldsymbol{\theta}^\dagger; X_{t-1}) + \gamma_1^*(\tau)|y_t| \tag{S.34}$$

with $\beta_0^*(\tau) = \beta_0 F_\epsilon^{-1}(\tau)$ and $\gamma^*(\tau) = \gamma_0 F_\epsilon^{-1}(\tau)$. Note that (S.34) corresponds to the symmetric absolute value CAViaR representation in Engle and Manganelli (2004, p.369). Also, note that $\beta_0^*(\tau)$ and $\gamma^*(\tau)$ are quantile-level dependent, while β_1 is instead a global parameter, independent of τ. Since $q_\tau(\boldsymbol{\theta}^\dagger; X_t)$ depends on unknown parameters, we cannot directly estimate $\beta_0^*, \beta_1, \gamma^*$ via nonlinear quantile regression. Koenker and Xiao (2009) suggest a two-step estimating procedure and establish the asymptotic properties of both parameters and conditional quantiles, respectively. In what follows, we outline the case in which the estimation steps use only the first $R-1$ observations.

Given (S.33) and a set of regularity conditions (see Koenker and Xiao, 2009), σ_t has an ARCH(∞) representation, and

$$q_\tau(\boldsymbol{\psi}^\dagger; X_t) = \alpha_0(\tau) + \sum_{j=1}^{\infty} \alpha_j(\tau)\,|y_{t-j}| \tag{S.35}$$

where $\alpha_j(\tau) = \alpha_j F_\epsilon^{-1}(\tau)$ and α_0 set equal to 1, and with the α_j, $j = 1, \ldots$, satisfying certain summability conditions. In particular, since $\alpha_j(\tau)$ decays at a geometric rate in j, we can approximate $q_\tau(\boldsymbol{\theta}^\dagger; X_t)$ in Equation (S.35) with m lags (where $m \to \infty$ at a logarithmic rate). Estimation can then proceed by taking a grid of quantile ranks $\tau_1, \tau_2, ..., \tau_K$, and by running K separate quantile autoregressions of order m using the truncated version of Equation (S.35), to get:

$$\left(\underbrace{\widehat{\alpha}_{1,R}(\tau_1), ..., \widehat{\alpha}_{m,R}(\tau_1)}_{\widehat{\pi}_{1,R}}, ..., \underbrace{\widehat{\alpha}_{1,R}(\tau_K), ..., \widehat{\alpha}_{m,R}(\tau_K)}_{\widehat{\pi}_{K,R}}\right).$$

Then, letting $\mathbf{a} = (\alpha_1, ..., \alpha_m, q_1, ..., q_K)$ and $\varphi(\mathbf{a}) = (q_1, a_1q_1, \ldots, a_mq_1, \ldots, q_K, a_1q_K, \ldots, a_mq_K)$, with $q_j = F_\epsilon^{-1}(\tau_j)$, compute:

$$\widehat{\mathbf{a}}_R = \arg\min_a (\widehat{\pi}_R - \varphi(\mathbf{a}))' A_R (\widehat{\pi} - \varphi(\mathbf{a})) \tag{S.36}$$

with A_R is a $K(m+1) \times K(m+1)$ weighting matrix, with K and m denoting the number of quantile levels and the truncation lag, respectively.

Given $\widehat{\mathbf{a}}_R$, we can then obtain an estimate for σ_t, $t = R, ..., T$ as:

$$\widehat{\sigma}_{t,R} = \widehat{a}_{0,R}(\tau) + \sum_{j=1}^{m} \widehat{a}_{j,R}(\tau)\,|y_{t-j}|,\ \ t = R, ..., T$$

so that $\widehat{X}_{t,R} = (1, \widehat{\sigma}_{t,R}, |y_t|)'$. In the last step, we perform a quantile regression of y_{t+1} onto $\widehat{X}_{t,R}$ to obtain an estimator of the conditional quantile parameters $\boldsymbol{\theta}^\dagger(\tau)$ as defined in Eq. (S.33):

$$\widehat{\boldsymbol{\theta}}_R(\tau) = \arg\min_{\boldsymbol{\theta}} \frac{1}{R}\sum_{t=1}^{R-1} \rho_\tau\left(y_{t+1} - \widehat{X}_{t,R}'\boldsymbol{\theta}\right).$$

This in turn provides an estimator of the τ conditional quantile:

$$\widehat{q}_{\tau,R}(\widehat{\boldsymbol{\psi}}_R; \widehat{X}_{t,R}) = \widehat{X}_{t,R}' \widehat{\boldsymbol{\theta}}_R(\tau).$$

For the test and the wild bootstrap statistic, the key difference with respect to "standard" quantile regression is that we need to take into account also the generated regressor $\widehat{X}_{t,R}$, due to the fact that $\widehat{X}_{t,R}$ is an estimator of X_t since $\widehat{\sigma}_{t+1,R}$ is an estimator of σ_t.

From the proof of Theorem 1, in CASE I, the first order term of the contribution of parametric quantile estimation error (in the one sided case) due to one of the models is given by:

$$\begin{aligned}
& \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} 1\{\mathbf{X}_t \in \mathcal{X}\} \nabla L \left(C\left(\boldsymbol{\psi}^{\dagger}; \mathbf{X}_t\right) - \tau \right) f_{t+1}\left(q_\tau \left(\boldsymbol{\psi}^{\dagger}; X_t\right) | \mathbf{X}_t \right) \left(\widehat{X}_{t,R}' \widehat{\boldsymbol{\theta}}_R(\tau) - X_t' \boldsymbol{\theta}^{\dagger}(\tau) \right) + o_p(1) \\
= & \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} 1\{\mathbf{X}_t \in \mathcal{X}\} \nabla L \left(C\left(\boldsymbol{\psi}^{\dagger}; \mathbf{X}_t\right) - \tau \right) f_{t+1}\left(q_\tau \left(\boldsymbol{\psi}^{\dagger}; X_t\right) | \mathbf{X}_t \right) X_t' \left(\widehat{\boldsymbol{\theta}}_R(\tau) - \boldsymbol{\theta}^{\dagger}(\tau) \right) \\
& + \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} 1\{\mathbf{X}_t \in \mathcal{X}\} \nabla L \left(C\left(\boldsymbol{\psi}^{\dagger}; \mathbf{X}_t\right) - \tau \right) f_{t+1}\left(q_\tau \left(\boldsymbol{\psi}^{\dagger}; X_t\right) | \mathbf{X}_t \right) \left(\widehat{X}_{t,R} - X_t \right)' \boldsymbol{\theta}^{\dagger}(\tau) \\
& + O_p(\text{smaller order})
\end{aligned}$$

Now, under the regularity conditions of Koenker and Xiao (2009), we have from their Theorem 3 that:

$$\begin{aligned}
& \sqrt{R} \left(\widehat{\boldsymbol{\theta}}_R(\tau) - \boldsymbol{\theta}^{\dagger}(\tau) \right) \\
= & \frac{1}{f_\epsilon\left(F_\epsilon^{-1}(\tau)\right)} H^{-1}(\tau) \frac{1}{\sqrt{R}} \sum_{s=1}^{R-1} X_t 1\left\{ y_{s+1} \leq X_s' \boldsymbol{\theta}^{\dagger}(\tau) \right\} \\
& + H^{-1}(\tau) \boldsymbol{\Psi} \sqrt{R} \left(\widehat{\boldsymbol{\alpha}}_R - \boldsymbol{\alpha}^{\dagger} \right) + o_p(1)
\end{aligned} \tag{S.37}$$

where the second component captures the contribution of generated regressors to the asymptotic distribution with $\widehat{\boldsymbol{\alpha}}_R$ being a vector of the first m components of $\widehat{\mathbf{a}}_R$ as defined in (S.36) ($\boldsymbol{\alpha}^{\dagger}$ denotes its population counterpart), and $\boldsymbol{\Psi}$ an $m \times m$ matrix (see Theorem 3 of Koenker and Xiao (2009) for details). Also, from the proof of Corollary 1 in Koenker and Xiao (2009),

$$\begin{aligned}
& \frac{1}{\sqrt{P}} \sum_{t=R}^{T-1} 1\{\mathbf{X}_t \in \mathcal{X}\} \nabla L \left(C\left(\boldsymbol{\psi}^{\dagger}; \mathbf{X}_t\right) - \tau \right) f_{t+1}\left(q_\tau \left(\boldsymbol{\psi}^{\dagger}; X_t\right) | \mathbf{X}_t \right) \left(\widehat{X}_{t,R} - X_t \right)' \boldsymbol{\theta}^{\dagger}(\tau) \\
= & \frac{1}{\sqrt{P}\sqrt{R}} \sum_{t=R}^{T-1} 1\{\mathbf{X}_t \in \mathcal{X}\} \nabla L \left(C\left(\boldsymbol{\psi}^{\dagger}; \mathbf{X}_t\right) - \tau \right) f_{t+1}\left(q_\tau \left(\boldsymbol{\psi}^{\dagger}; X_t\right) | \mathbf{X}_t \right) \\
& \times \begin{pmatrix} 0 \\ \left(\ |y_t|, \quad \ldots \quad , |y_{t-m}| \ \right) \sqrt{R} \left(\widehat{\boldsymbol{\alpha}}_R - \boldsymbol{\alpha}^{\dagger} \right) \\ 0 \end{pmatrix}' \boldsymbol{\theta}^{\dagger}(\tau) + o_p(1).
\end{aligned} \tag{S.38}$$

It is immediate from (S.37) and (S.38) that when constructing the wild bootstrap statistic, we would need an extra term capturing the contribution of $(\widehat{\alpha}_R - \alpha_R)$.

Given that estimation of CAViaR models is typically implemented using daily observations, and so the available sample consists of several thousands of observations, it may be more convenient in practice to rely on subsample based critical values. On the other hand, in the GaR applications we only use few hundreds observations and so subsampling is not a viable option.

S.7 Additional Monte Carlo Results

S.7.1 Small Sample Size ($T = 120$), otherwise same set-up as Main Results

Table S1: Rejection Rates: Pairwise - Single Quantile Level - Small Sample Size ($T = 120$)

$T = 120$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0655	0.0220	0.0520
	$\tau = 0.2$	0.0460	0.0105	0.0880
	$\tau = 0.3$	0.0435	0.0130	0.1256
$l = 2$	$\tau = 0.1$	0.0530	0.0235	0.0515
	$\tau = 0.2$	0.0480	0.0125	0.1111
	$\tau = 0.3$	0.0415	0.0125	0.1326
$l = 5$	$\tau = 0.1$	0.0430	0.0220	0.0455
	$\tau = 0.2$	0.0415	0.0135	0.0900
	$\tau = 0.3$	0.0400	0.0105	0.1231

Table S2: Rejection Rates: Pairwise - Multiple Quantile Levels - Small Sample Size ($T = 120$)

$T = 120$	DGP1	DGP2	DGP3
$l = 1$	0.0600	0.0185	0.1721
$l = 2$	0.0565	0.0185	0.1896
$l = 5$	0.0475	0.0160	0.1761

Table S3: Rejection Rates: Multiple Models - Multiple Quantile Levels - Small Sample Size ($T = 120$)

$T = 120$	DGP1	DGP2	DGP3
$l = 1$	0.0800	0.0055	0.1676
$l = 2$	0.0655	0.0060	0.1726
$l = 5$	0.0575	0.0065	0.1446

S.7.2 High Time Series Dependence ($\rho = 0.7$)

Table S4: Rejection Rates: Pairwise - Single Quantile Level - High Time Series Dependence ($\rho = 0.7$)

$T = 240$		DGP1	DGP2	DGP3
$l = 5$	$\tau = 0.1$	0.0645	0.0220	0.1251
	$\tau = 0.2$	0.0730	0.0175	0.2051
	$\tau = 0.3$	0.0725	0.0170	0.4932
$l = 10$	$\tau = 0.1$	0.0570	0.0205	0.1071
	$\tau = 0.2$	0.0730	0.0175	0.2081
	$\tau = 0.3$	0.0595	0.0175	0.4527
$l = 20$	$\tau = 0.1$	0.0595	0.0170	0.1081
	$\tau = 0.2$	0.0555	0.0165	0.2071
	$\tau = 0.3$	0.0555	0.0165	0.4187
$T = 480$		DGP1	DGP2	DGP3
$l = 5$	$\tau = 0.1$	0.0900	0.0145	0.2576
	$\tau = 0.2$	0.0990	0.0220	0.6123
	$\tau = 0.3$	0.0900	0.0140	0.9585
$l = 10$	$\tau = 0.1$	0.0635	0.0135	0.2101
	$\tau = 0.2$	0.0795	0.0230	0.6033
	$\tau = 0.3$	0.0805	0.0140	0.9390
$l = 20$	$\tau = 0.1$	0.0710	0.0145	0.1971
	$\tau = 0.2$	0.0730	0.0210	0.5013
	$\tau = 0.3$	0.0625	0.0115	0.9260
$T = 960$		DGP1	DGP2	DGP3
$l = 5$	$\tau = 0.1$	0.1021	0.0085	0.5883
	$\tau = 0.2$	0.1281	0.0160	0.9865
	$\tau = 0.3$	0.1116	0.0155	1.0000
$l = 10$	$\tau = 0.1$	0.0840	0.0085	0.5448
	$\tau = 0.2$	0.1056	0.0120	0.9850
	$\tau = 0.3$	0.0950	0.0145	1.0000
$l = 20$	$\tau = 0.1$	0.0830	0.0095	0.5273
	$\tau = 0.2$	0.0795	0.0135	0.9665
	$\tau = 0.3$	0.0780	0.0120	1.0000

S.7.3 High Time Series Dependence ($\rho = 0.9$)

Table S5: Rejection Rates: Pairwise - Single Quantile Level - High Time Series Dependence ($\rho = 0.9$)

$T = 240$		DGP1	DGP2	DGP3
$l = 5$	$\tau = 0.1$	0.1436	0.0320	0.1676
	$\tau = 0.2$	0.1576	0.0305	0.2391
	$\tau = 0.3$	0.1516	0.0260	0.4122
$l = 10$	$\tau = 0.1$	0.1101	0.0305	0.1471
	$\tau = 0.2$	0.1221	0.0320	0.2216
	$\tau = 0.3$	0.1306	0.0290	0.3007
$l = 20$	$\tau = 0.1$	0.1026	0.0340	0.1541
	$\tau = 0.2$	0.1026	0.0295	0.2131
	$\tau = 0.3$	0.1096	0.0240	0.2166
$T = 480$		DGP1	DGP2	DGP3
$l = 5$	$\tau = 0.1$	0.1736	0.0200	0.2591
	$\tau = 0.2$	0.1911	0.0265	0.4487
	$\tau = 0.3$	0.1826	0.0205	0.8324
$l = 10$	$\tau = 0.1$	0.1331	0.0170	0.1996
	$\tau = 0.2$	0.1441	0.0250	0.4577
	$\tau = 0.3$	0.1256	0.0185	0.8229
$l = 20$	$\tau = 0.1$	0.1146	0.0225	0.1816
	$\tau = 0.2$	0.1176	0.0270	0.3607
	$\tau = 0.3$	0.1136	0.0175	0.6363
$T = 960$		DGP1	DGP2	DGP3
$l = 5$	$\tau = 0.1$	0.1861	0.0150	0.4812
	$\tau = 0.2$	0.2161	0.0155	0.8844
	$\tau = 0.3$	0.2171	0.0220	0.9985
$l = 10$	$\tau = 0.1$	0.1476	0.0165	0.3992
	$\tau = 0.2$	0.1636	0.0160	0.8379
	$\tau = 0.3$	0.1596	0.0210	0.9950
$l = 20$	$\tau = 0.1$	0.1086	0.0165	0.3247
	$\tau = 0.2$	0.1221	0.0150	0.7374
	$\tau = 0.3$	0.1211	0.0230	0.9920

S.7.4 Correlation in $X_{j,t}$ ($\phi = 0.25$)

Table S6: Rejection Rates: Pairwise - Single Quantile Level - Correlation in $X_{j,t}$

$T = 240$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0685	0.0180	0.1036
	$\tau = 0.2$	0.0715	0.0210	0.2351
	$\tau = 0.3$	0.0555	0.0125	0.5163
$l = 2$	$\tau = 0.1$	0.0610	0.0180	0.0840
	$\tau = 0.2$	0.0645	0.0125	0.2731
	$\tau = 0.3$	0.0565	0.0125	0.4802
$l = 5$	$\tau = 0.1$	0.0485	0.0165	0.1016
	$\tau = 0.2$	0.0475	0.0140	0.2361
	$\tau = 0.3$	0.0430	0.0135	0.4877
$T = 480$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0625	0.0090	0.2656
	$\tau = 0.2$	0.0720	0.0160	0.6823
	$\tau = 0.3$	0.0825	0.0145	0.9685
$l = 2$	$\tau = 0.1$	0.0655	0.0115	0.2341
	$\tau = 0.2$	0.0725	0.0150	0.6978
	$\tau = 0.3$	0.0790	0.0165	0.9710
$l = 5$	$\tau = 0.1$	0.0475	0.0095	0.2246
	$\tau = 0.2$	0.0610	0.0155	0.7194
	$\tau = 0.3$	0.0600	0.0160	0.9680
$T = 960$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0770	0.0080	0.6018
	$\tau = 0.2$	0.0980	0.0100	0.9915
	$\tau = 0.3$	0.0835	0.0175	1.0000
$l = 2$	$\tau = 0.1$	0.0670	0.0100	0.5773
	$\tau = 0.2$	0.0860	0.0165	0.9915
	$\tau = 0.3$	0.0780	0.0170	1.0000
$l = 5$	$\tau = 0.1$	0.0560	0.0095	0.6193
	$\tau = 0.2$	0.0770	0.0125	0.9865
	$\tau = 0.3$	0.0745	0.0190	1.0000

S.7.5 No Trimming

Table S7: Rejection Rates: Pairwise - Single Quantile Level - No Trimming

$T = 240$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0515	0.0120	0.1196
	$\tau = 0.2$	0.0670	0.0130	0.2831
	$\tau = 0.3$	0.0480	0.0120	0.5918
$l = 2$	$\tau = 0.1$	0.0515	0.0115	0.1096
	$\tau = 0.2$	0.0570	0.0110	0.3312
	$\tau = 0.3$	0.0480	0.0115	0.5343
$l = 5$	$\tau = 0.1$	0.0445	0.0100	0.1051
	$\tau = 0.2$	0.0465	0.0095	0.2731
	$\tau = 0.3$	0.0415	0.0115	0.5073
$T = 480$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0720	0.0090	0.3327
	$\tau = 0.2$	0.0700	0.0100	0.8014
	$\tau = 0.3$	0.0730	0.0075	0.9815
$l = 2$	$\tau = 0.1$	0.0610	0.0100	0.3052
	$\tau = 0.2$	0.0720	0.0070	0.7679
	$\tau = 0.3$	0.0785	0.0090	0.9705
$l = 5$	$\tau = 0.1$	0.0415	0.0120	0.3192
	$\tau = 0.2$	0.0605	0.0105	0.7734
	$\tau = 0.3$	0.0640	0.0075	0.9715
$T = 960$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0680	0.0080	0.7489
	$\tau = 0.2$	0.0940	0.0115	0.9945
	$\tau = 0.3$	0.0915	0.0085	1.0000
$l = 2$	$\tau = 0.1$	0.0510	0.0080	0.7089
	$\tau = 0.2$	0.0785	0.0100	0.9940
	$\tau = 0.3$	0.0910	0.0115	1.0000
$l = 5$	$\tau = 0.1$	0.0470	0.0100	0.6978
	$\tau = 0.2$	0.0725	0.0130	0.9925
	$\tau = 0.3$	0.0755	0.0130	1.0000

S.7.6 Fourth-Order Kernel

Table S8: Rejection Rates: Pairwise - Single Quantile Level - Fourth Order Kernel

$T = 240$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0595	0.0085	0.1161
	$\tau = 0.2$	0.0845	0.0130	0.3132
	$\tau = 0.3$	0.0565	0.0095	0.5663
$l = 2$	$\tau = 0.1$	0.0625	0.0075	0.0970
	$\tau = 0.2$	0.0645	0.0115	0.3302
	$\tau = 0.3$	0.0620	0.0090	0.5353
$l = 5$	$\tau = 0.1$	0.0570	0.0100	0.0935
	$\tau = 0.2$	0.0585	0.0120	0.2651
	$\tau = 0.3$	0.0495	0.0100	0.4962
$T = 480$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0880	0.0090	0.3112
	$\tau = 0.2$	0.0925	0.0095	0.7889
	$\tau = 0.3$	0.0895	0.0155	0.9790
$l = 2$	$\tau = 0.1$	0.0710	0.0075	0.2901
	$\tau = 0.2$	0.0995	0.0105	0.7499
	$\tau = 0.3$	0.0990	0.0135	0.9765
$l = 5$	$\tau = 0.1$	0.0655	0.0090	0.2971
	$\tau = 0.2$	0.0705	0.0110	0.7934
	$\tau = 0.3$	0.0780	0.0125	0.9765
$T = 960$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0750	0.0090	0.7344
	$\tau = 0.2$	0.1126	0.0115	0.9960
	$\tau = 0.3$	0.0975	0.0165	1.0000
$l = 2$	$\tau = 0.1$	0.0665	0.0120	0.6233
	$\tau = 0.2$	0.0985	0.0150	0.9945
	$\tau = 0.3$	0.0915	0.0165	1.0000
$l = 5$	$\tau = 0.1$	0.0710	0.0135	0.6513
	$\tau = 0.2$	0.0830	0.0180	0.9920
	$\tau = 0.3$	0.0790	0.0180	1.0000

S.7.7 Student's-t Distributed Errors

Table S9: Rejection Rates: Pairwise - Single Quantile Level - Student's-t Errors

$T = 240$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0615	0.0150	0.1031
	$\tau = 0.2$	0.0945	0.0145	0.3297
	$\tau = 0.3$	0.0915	0.0125	0.6813
$l = 2$	$\tau = 0.1$	0.0675	0.0150	0.0780
	$\tau = 0.2$	0.0870	0.0150	0.2966
	$\tau = 0.3$	0.0755	0.0125	0.7044
$l = 5$	$\tau = 0.1$	0.0500	0.0155	0.0880
	$\tau = 0.2$	0.0665	0.0155	0.2996
	$\tau = 0.3$	0.0620	0.0145	0.6323
$T = 480$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0685	0.0155	0.2736
	$\tau = 0.2$	0.0965	0.0145	0.7894
	$\tau = 0.3$	0.1006	0.0165	0.9950
$l = 2$	$\tau = 0.1$	0.0725	0.0135	0.2726
	$\tau = 0.2$	0.0895	0.0155	0.7799
	$\tau = 0.3$	0.0810	0.0185	0.9915
$l = 5$	$\tau = 0.1$	0.0610	0.0135	0.2716
	$\tau = 0.2$	0.0620	0.0160	0.7659
	$\tau = 0.3$	0.0715	0.0200	0.9920
$T = 960$		DGP1	DGP2	DGP3
$l = 1$	$\tau = 0.1$	0.0990	0.0125	0.6413
	$\tau = 0.2$	0.1451	0.0055	0.9970
	$\tau = 0.3$	0.1436	0.0160	1.0000
$l = 2$	$\tau = 0.1$	0.1001	0.0105	0.5703
	$\tau = 0.2$	0.1311	0.0100	0.9970
	$\tau = 0.3$	0.1331	0.0160	1.0000
$l = 5$	$\tau = 0.1$	0.0805	0.0095	0.5713
	$\tau = 0.2$	0.1076	0.0075	0.9950
	$\tau = 0.3$	0.1011	0.0165	1.0000

S.8 Additional Monte Carlo Set-Up

In this section we slightly modify the set-up of the main text so that the DGP for y_{t+1} only depends on the variable $X_{1,t}$ and its square:

$$y_{t+1} = \beta_1 X_{1,t} + \beta_2 X_{1,t}^2 + e_{t+1} \tag{S.39}$$

We also have two other variables, $X_{2,t}$ and $X_{3,t}$. We generate forecasts using two different linear quantile models, the first of which uses the set of regressors $[X_{1,t}, X_{2,t}]'$ and the second uses $[X_{1,t}, X_{3,t}]'$. The models are overlapping as they both use an irrelevant regressor and only $X_{1,t}$ from both models actually features in the DGP. They are also mis-specified as the linear quantile model does not capture the non-linearity in the DGP in (20). This corresponds to the overlapping case discussed in Theorem 1(ii) subcase (b). We will use values of the parameters in the DGP given by $(\beta_1, \beta_2) = (1, 1)$.

The remainder of the set-up is the basically same as that in the main paper. The predictors are generated as $X_{j,t} = \rho X_{j,t-1} + v_{j,t}$ for $j = 1, 2, 3$ and we set $\rho = 0.5$. The errors $v_{j,t}$ follow independent normal distributions with variance equal to $1 - \rho^2$. The error term e_{t+1} in (20) is drawn from a standard normal distribution. The sample sizes, bootstrap parameters, quantile levels etc. are all exactly as in the main paper.

Table S10 presents the rejection rates for the pairwise single quantile version of the test as in Table 1 of the main paper. This confirms that the rejection rate approaches zero with the sample size, as predicted by the results in Theorem 2 of the main paper, for this subcase of Case II where models are overlapping but have equally incorrect coverage.

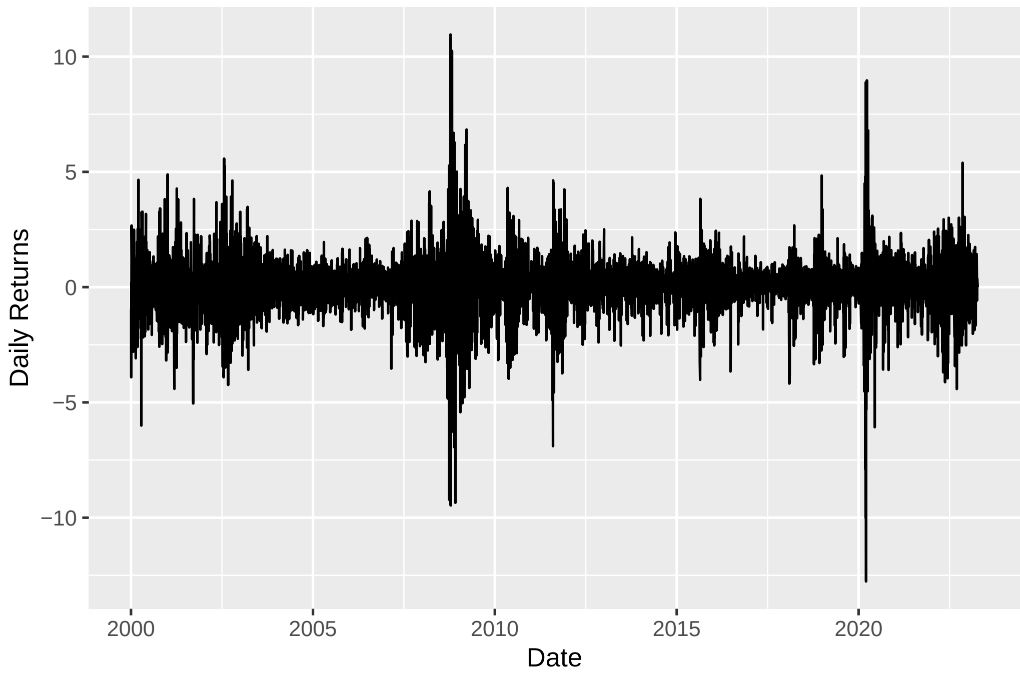
Table S10: Rejection Rates: Pairwise - Single Quantile Level

$T = 240$		
$l = 1$	$\tau = 0.1$	0.0215
	$\tau = 0.2$	0.0170
	$\tau = 0.3$	0.0120
$l = 2$	$\tau = 0.1$	0.0195
	$\tau = 0.2$	0.0165
	$\tau = 0.3$	0.0110
$l = 5$	$\tau = 0.1$	0.0170
	$\tau = 0.2$	0.0185
	$\tau = 0.3$	0.0115
$T = 480$		
$l = 1$	$\tau = 0.1$	0.0165
	$\tau = 0.2$	0.0175
	$\tau = 0.3$	0.0105
$l = 2$	$\tau = 0.1$	0.0150
	$\tau = 0.2$	0.0185
	$\tau = 0.3$	0.0095
$l = 5$	$\tau = 0.1$	0.0135
	$\tau = 0.2$	0.0150
	$\tau = 0.3$	0.0095
$T = 960$		
$l = 1$	$\tau = 0.1$	0.0145
	$\tau = 0.2$	0.0155
	$\tau = 0.3$	0.0090
$l = 2$	$\tau = 0.1$	0.0175
	$\tau = 0.2$	0.0135
	$\tau = 0.3$	0.0085
$l = 5$	$\tau = 0.1$	0.0165
	$\tau = 0.2$	0.0150
	$\tau = 0.3$	0.0085

S.9 Additional Empirical Illustration

In this section we perform a brief additional empirical illustration of our methods in the financial context of VaR prediction. This allows us to apply our test in a high-frequency data environment and also to use the test with location-scale type models, the theory for which is developed earlier in this supplementary material. We focus on the daily returns on the S&P500 index, which was the base series used for the SV variable used in the GaR application in the main text (see that section for details of the data source). We use daily observations of the series from 3rd January 2000 to 11th April 2023 which gives a total of $T = 5855$ observations. The series is displayed below in Figure 1 which clearly shows the pronounced volatility around the Great Recession and the shorter period of volatility around the beginning of the Covid-19 period.

Figure 1: S&P500 Daily Returns



We perform backtests of one day ahead VaR predictions from two classic volatility models, the GARCH(1,1) model (Bollerslev, 1986) and the exponential GARCH(1,1) model (eGARCH, Nelson, 1991). These models are very widely used in practice, as well as in similar empirical illustrations of VaR backtesting methods such as in Escanciano and Olmo (2010). We start making VaR forecasts in 2020 which gives an out-of-sample window of $P = 824$, with a large in-sample window of $R = 5031$ used for estimation. Since R is substantially larger than P in our application, we ignore the presence of parameter estimation error (PEE) from the parametric GARCH models which greatly simplifies the bootstrap implementation. Specifically, we construct the bootstrap statistic using only $\widehat{A}_{j,R,P,t}(\tau)$ and $\widehat{B}_{j,R,P,t}(\tau)$, $j = 1, 2$, to capture the contribution of the population coverage error and the estimation error of conditional coverage, respectively. Indeed, we believe that PEE from the parametric quantile models is a much more pressing issue in the smaller samples encountered in GaR applications, which we illustrate in the main text. Since we ignore PEE, the out-of-sample estimation scheme (fixed, rolling, recursive) is also irrelevant. We therefore use recursive estimation to obtain the VaR predictions which is simple to implement using the `rugarch` package in `R`.

We will implement the test using quantile levels of $\tau \in \{0.01, 0.025, 0.05\}$ which are commonly used in the VaR literature. As in the main paper, we will assess the results of the single-quantile test for these different quantiles and then look at the multiple quantile test. We will consider lag truncation parameters $l \in \{10, 20, 30\}$, and otherwise the remainder of the set-up (bootstrap draws, trimming fraction, kernel and bandwidth rules) is the same as described in the main text.

The results of the pairwise test for a single quantile level are given in Table S11 below, reported for the three different quantile levels. The GARCH(1,1) model is set to be Model 1 and the eGARCH(1,1) model is Model 2. The results suggest that, in fact, the standard GARCH(1,1) model has lower coverage error loss than the eGARCH(1,1) model across all of the three individual quantile levels considered, as evidenced by the negative values of the test statistics. In terms of the significance, however, there is no evidence to reject the null even at the 10% significance level, with the lowest p-value just above 0.3 for the $\tau = 0.025$ quantile level which has the test statistic furthest from zero. The p-values are very stable across all values of the lag truncation parameter l. This indicates that there is no statistical evidence that the GARCH(1,1) significantly improves over the eGARCH(1,1) model. This makes sense when looking at the VaR forecasts themselves (see Figure 2). With the exception of the periods of high volatility, the models produce very similar predictions, like in the overlapping case, and clearly the small periods of deviation are not enough to drive a rejection of the null hypothesis. Since the test statistics are all negative at every quantile, the multiple quantile version of the test with the GARCH(1,1) model as the benchmark has a statistic of 0 as it is given by:

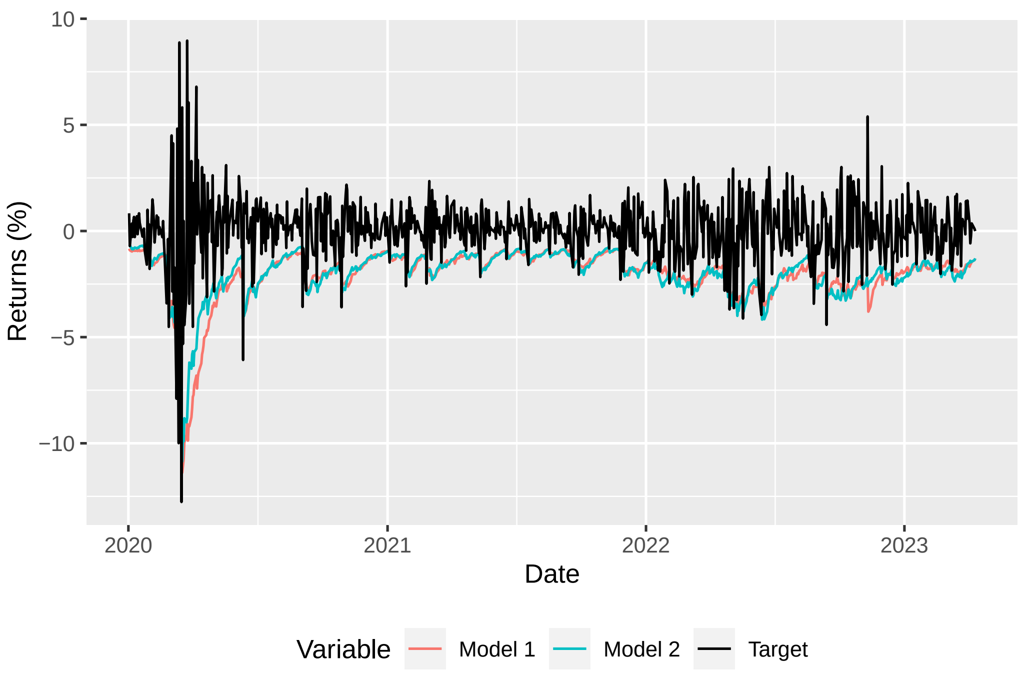
$$\widehat{S}_{P,R}^{\max} = \sum_{k=1}^{M} \left(0, \max\left\{0, \widehat{S}_{P,R,k}\right\}\right)^2,$$

and so we are obviously unable to reject the null that the GARCH(1,1) model has equal or superior coverage error loss than the eGARCH(1,1) across quantiles.

Table S11: GARCH(1,1) vs. eGARCH(1,1) - Pairwise Comparison - Single Quantile Level

		Stat	p-value
$l = 10$	$\tau = 0.01$	-0.0074	0.6703
	$\tau = 0.025$	-0.0155	0.3482
	$\tau = 0.05$	-0.0110	0.4762
$l = 20$	$\tau = 0.01$	-	0.6793
	$\tau = 0.025$	-	0.3382
	$\tau = 0.05$	-	0.4512
$l = 30$	$\tau = 0.01$	-	0.6993
	$\tau = 0.025$	-	0.3102
	$\tau = 0.05$	-	0.4292

Figure 2: 5% VaR Forecasts - GARCH(1,1) versus eGARCH(1,1) - S&P500 Daily Returns



References

Andrews, D. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 3*, 817–858.

Andrews, D. (1994). Empirical process methods in econometrics. In R. Engle and D. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 37, pp. 228–2294. Elsevier.

Andrews, D. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory 11*, 560–596.

Andrews, D. and D. Pollard (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review 62*(1), 119–132.

Andrews, D. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*, 119–157.

Angrist, J., V. Chernozhukov, and I. Fernandez-Val (2006). Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure. *Econometrica 74*(2), 539–563.

Arcones, M. and B. Yu (1994). Central Limit Theorems for Empirical and U-Processes of Stationary Mixing Sequences. *Journal of Theoretical Probability 7*(1), 47–71.

Bahadur, R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics 37*(3), 577–580.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

Corradi, V., W. Distaso, and M. Fernandes (2020). Testing for jump spillovers without testing for jumps. *Journal of the American Statistical Association 115*, 1214–1226.

Corradi, V. and N. Swanson (2002). A consistent test for nonlinear out of sample predictive accuracy. *Journal of Econometrics 110*, 353–381.

Engle, R. and S. Manganelli (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics 22*(4), 367–381.

Escanciano, J. C. and J. Olmo (2010). Backtesting parametric value-at-risk with estimation risk. *Journal of Business and Economic Statistics 28*(1), 36–51.

Franguridi, G., B. Gafarov, and K. Wüthrich (2022). Bias correction for quantile regression estimators. arXiv working paper arXiv:2011.03073v5.

Ghosal, S., A. Sen, and A. Van der Vaart (2000). Testing monotonicity of regression. *Annals of Statistics 28*(4), 1054–1082.

Gregory, K., S. Lahiri, and D. Nordman (2018). A smooth block bootstrap for quantile regression with time series. *Annals of Statistics 46*(3), 1138–1166.

Hall, P. and C. Heyde (1980). *Martingale Limit Theory and Its Application.* Academic Press Inc.

Koenker, R. and Z. Xiao (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association 109*(488), 1696–1712.

Lee, T.-W., A. Ullah, and H. Wang (2018). The second-order bias of quantile estimators. *Economic Letters 173*, 143–147.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica 59*(2), 347–370.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

Sancetta, A. (2009). Maximal Inequalities for U-Processes of Strongly Mixing Random Variables. *Journal of Mathematical Statistics 29*, 155–167.

Van der Vaart, A. and J. Wellner (1996). *Weak Convergence and Empirical Processes* (First ed.). Springer Series in Statistics. Springer Verlag.

West, K. D. (1996). Asymptotic Inference about Predictive Ability. *Econometrica 64*(5), 1067–1084.

Yoshihara, K. (1976). Limiting Behavior of U-statistics for Stationary, Absolute Regular Processes. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 35*(3), 237–252.