Quantile Estimation of Regression Models with GARCH-X Errors

Qianqian Zhu, Guodong Li and Zhijie Xiao

Shanghai University of Finance and Economics, University of Hong Kong and Boston College

Supplementary Material

This supplementary material provides Lemmas 1-5 with proofs, and includes technical details for Theorems 1-3 and Corollaries 1-3, as well as additional simulation results for Section 3.1.

S1 Technical details

This section provides the detailed proofs of Theorems 1-3 and Corollaries 1-3. To show Theorems 1-3, we introduce Lemmas 1-5 with proofs. Specifically, Lemma 1 contains some preliminary results. Lemma 2 will be used to handle initial values in GARCH-X models. Lemma 3 verifies the stochastic differentiability condition defined on Page 298 of Pollard (1985), and its proof mainly uses the bracketing method; see also Lee and Noh (2013) and Zhu and Ling (2011). Lemmas 4 and 5 will be used to verify the root-n consistency and the asymptotic normality of $\widehat{\boldsymbol{\theta}}_{\tau n}$ in Theorem 2, and their proofs are based on Lemma 3 and some approximation arguments.

Throughout this section C is a generic positive constant which may take different values at its different occurrences, $\rho \in (0,1)$ is a generic constant which may take different values at its different occurrences, $o_p(1)$ denotes a sequence of random variables converging to zero in probability, and the notation $o_p^*(1)$ corresponds to the bootstrap probability space. We denote by $\|\cdot\|$ the norm of a matrix or column vector, defined as $\|A\| = \sqrt{\text{tr}(AA')} = \sqrt{\sum_{i,j}|a_{ij}|^2}$. For simplicity, denote $\psi_\tau(x) = \tau - I(x < 0)$ and $\varepsilon_{t,\tau} = \varepsilon_t - b_\tau$. In addition, let $\sigma_t = \sigma_t(\boldsymbol{\lambda}_0)$, $\widetilde{\sigma}_t = \sigma_t(\widehat{\boldsymbol{\lambda}}_n^{int})$, $\widehat{\sigma}_t = \widetilde{\sigma}_t(\widehat{\boldsymbol{\lambda}}_n^{int})$, $\widetilde{\ell}_t(\boldsymbol{\theta}) = \rho_\tau[Y_t - \widetilde{q}_t(\boldsymbol{\theta})]$ and $\ell_t(\boldsymbol{\theta}) = \rho_\tau[Y_t - q_t(\boldsymbol{\theta})]$, where $\boldsymbol{\lambda}_0$ and $\widehat{\boldsymbol{\lambda}}_n^{int}$ are the true value and an appropriate estimator of $\boldsymbol{\lambda}$, respectively, $q_t(\boldsymbol{\theta}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b\sigma_t(\boldsymbol{\lambda})$ and $\widetilde{q}_t(\boldsymbol{\theta}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b\widetilde{\sigma}_t(\boldsymbol{\lambda})$ are the conditional quantile functions of Y_t without and with initial values, respectively.

Lemma 1. *Let* $\xi_{\rho,t} = \sum_{j=0}^{\infty}\rho^j(1 + \|\boldsymbol{X}_{t-j-1}\| + \|\boldsymbol{V}_{t-j-1}\|^{1/2} + |u_{t-j}|)$ *and* $\zeta_{\rho,t} = \sum_{j=0}^{\infty}\rho^j(1 + \|\boldsymbol{X}_{t-j-1}\| + |u_{t-j}|^{\iota})$ *be positive random variables depending on a constant* $\rho \in (0,1)$*, where* ι *is a constant satisfying* $\iota \in (0, 2/(4+\delta))$ *for some* $\delta > 0$*. If Assumption 1 holds, then*

(i) $\sup_{\Theta} \sigma_t^2(\boldsymbol{\lambda}) \leqslant C\xi_{\rho,t-1}^2$;

(ii) $\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| \leqslant C\zeta_{\rho,t}$ *and* $\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\| \leqslant C\zeta_{\rho,t-1}^2$;

(iii) $\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial \widetilde{q}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\| \leqslant C\zeta_{\rho,t}$ *and* $\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 \widetilde{q}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\| \leqslant C\zeta_{\rho,t-1}^2$;

(iv) for any $\kappa > 0$, there exists a constant $c > 0$ such that

$$E\left\{\sup_{\Theta}\left[\frac{\sigma_t^2(\boldsymbol{\lambda}_1)}{\sigma_t^2(\boldsymbol{\lambda}_2)}\right]^{\kappa} : \|\boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2\| \leqslant c\right\} < \infty.$$

Lemma 2. *Let $\xi_\rho = \sum_{j=0}^{\infty} \rho^j (1 + \|\boldsymbol{X}_{-j-1}\| + |u_{-j}|)$ be a positive random variable depending on a constant $\rho \in (0, 1)$. If Assumption 1 holds, then*

(i) $\sup_{\Theta} |\widetilde{\sigma}_t(\boldsymbol{\lambda}) - \sigma_t(\boldsymbol{\lambda})| \leqslant C\rho^t \xi_\rho$; *(ii)* $\sup_{\Theta} \dfrac{1}{\sigma_t(\boldsymbol{\lambda})}\left\|\dfrac{\partial \widetilde{q}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \dfrac{\partial q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\| \leqslant C\rho^t \xi_\rho \zeta_{\rho,t}$.

Lemma 3. *If Assumptions 1, 3 and 4 hold and $E(u_t^2) < \infty$, then for $\boldsymbol{u} = o_p(1)$,*

$$\zeta_n(\boldsymbol{u}) = o_p(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2),$$

where $\zeta_n(\boldsymbol{u}) = \sum_{t=1}^{n} \sigma_t^{-1} q_{1t}(\boldsymbol{u}) \{\xi_{1t}(\boldsymbol{u}) - E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}]\}$ with

$$q_{1t}(\boldsymbol{u}) = \boldsymbol{u}' \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} \quad \text{and} \quad \xi_{1t}(\boldsymbol{u}) = \int_0^1 \left[I\left(\varepsilon_t \leqslant b_\tau + \sigma_t^{-1} q_{1t}(\boldsymbol{u}) s\right) - I(\varepsilon_t \leqslant b_\tau)\right] ds.$$

Lemma 4. *Suppose that $\sqrt{n}(\widetilde{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) = O_p(1)$ and $E(u_t^2) < \infty$. Under Assumptions 1, 3 and 4, for $\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0} = o_p(1)$, it holds that*

$$n[\widehat{L}_n(\boldsymbol{\theta}) - \widehat{L}_n(\boldsymbol{\theta}_{\tau 0})] - n[\widetilde{L}_n(\boldsymbol{\theta}) - \widetilde{L}_n(\boldsymbol{\theta}_{\tau 0})] = o_p(\sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}\| + n\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}\|^2),$$

where $\widetilde{L}_n(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^{n} \widetilde{\sigma}_t^{-1} \rho_\tau[Y_t - q_t(\boldsymbol{\theta})]$ and $\widehat{L}_n(\boldsymbol{\theta}) = n^{-1} \sum_{t=1}^{n} \widehat{\sigma}_t^{-1} \rho_\tau[Y_t - \widetilde{q}_t(\boldsymbol{\theta})]$.

Lemma 5. *Suppose that $\sqrt{n}(\widetilde{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0) = O_p(1)$ and $E(u_t^2) < \infty$. Under Assumptions*

1, 3 and 4, for $\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0} = o_p(1)$*, it holds that*

$$\begin{aligned} n[\widetilde{L}_n(\boldsymbol{\theta}) - \widetilde{L}_n(\boldsymbol{\theta}_{\tau 0})] &= -\sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0})'\boldsymbol{T}_n + \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0})' J_n \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}) \\ &\quad + o_p(\sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}\| + n\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}\|^2), \end{aligned}$$

where $\widetilde{L}_n(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^n \widetilde{\sigma}_t^{-1}\rho_\tau[Y_t - q_t(\boldsymbol{\theta})]$*,*

$$\boldsymbol{T}_n = \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}}\psi_\tau(\varepsilon_{t,\tau}) \ \ \textit{and} \ \ J_n = \frac{f_\varepsilon(b_\tau)}{2n}\sum_{t=1}^n \frac{1}{\sigma_t^2}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}'}.$$

Proof of Lemma 1. Denote $\alpha(B) = \sum_{i=1}^q \alpha_i B^i$ and $\beta(B) = 1 - \sum_{i=1}^p \beta_i B^i$, where B is the back-shift operator. By Assumption 1, it holds that

$$\beta^{-1}(B)\alpha(B) = \sum_{i=1}^\infty a_\gamma(i)B^i \quad \text{and} \quad \beta^{-1}(B) = \sum_{i=0}^\infty a_\beta(i)B^i,$$

where $a_\beta(i) = \boldsymbol{e}'G^i\boldsymbol{e}$ and $a_\gamma(i) = \sum_{j=1}^q \alpha_j a_\beta(i-j)$, $\boldsymbol{e} = (1, 0, \ldots, 0)'$ is $p \times 1$ vector and $p \times p$ matrix G is defined as below

$$G = \begin{pmatrix} \beta_1 & \ldots & \beta_p \\ I_{p-1} & & \mathbf{0} \end{pmatrix},$$

with I_m being the $m \times m$ identity matrix and $\mathbf{0}$ being the zero vector with compatible dimensions. By Lemma 3.1 of Berkes, Horváth, and Kokoszka (2003), we have

$$\sup_{\Theta} a_\gamma(i) \leqslant C\rho^i \quad \text{and} \quad \sup_{\Theta} a_\beta(i) \leqslant C\rho^i \tag{S1.1}$$

for a constant $0 < C < \infty$ and a constant $\rho \in (0,1)$. Moreover, for $\boldsymbol{\theta} \in \Theta$ and any constant vector $\boldsymbol{c}$ with all elements being nonnegative, it holds that $\boldsymbol{c}'\boldsymbol{e}\boldsymbol{e}'\boldsymbol{c} \leqslant \boldsymbol{c}'G\boldsymbol{c}/\underline{\omega}$, and hence $a_\beta(i+k) = \boldsymbol{e}'G^{i+k}\boldsymbol{e} \geqslant \boldsymbol{e}'G^i\boldsymbol{e}\underline{\omega}^k = a_\beta(i)\underline{\omega}^k$. This implies that $\sup_\Theta a_\beta(i)/a_\beta(i+k) \leqslant \underline{\omega}^{-k}$. Note that $a_\gamma(i) \geqslant \alpha_1 a_\beta(i-1)$. As a result, for any integer i and $k \leqslant \max(p,q)$, it can be verified that

$$\sup_{\Theta} \frac{a_\beta(i)}{a_\gamma(i+k)} \leqslant \underline{\omega}^{-k} \quad \text{and} \quad \sup_{\Theta} \frac{a_\gamma(i)}{a_\gamma(i+k)} \leqslant \overline{\omega} \sum_{j=1}^{q} \underline{\omega}^{-j-k}. \tag{S1.2}$$

We first prove (i). Since $\sigma_t^2(\boldsymbol{\lambda}) = 1 + \sum_{i=1}^q \alpha_i u_{t-i}^2(\boldsymbol{\phi}) + \sum_{j=1}^p \beta_j \sigma_{t-j}^2(\boldsymbol{\lambda}) + \boldsymbol{\pi}'\boldsymbol{V}_{t-1}$, then we have

$$\begin{aligned} \sigma_t^2(\boldsymbol{\lambda}) &= \beta^{-1}(B)\left(1 + \boldsymbol{\pi}'\boldsymbol{V}_{t-1}\right) + \beta^{-1}(B)\alpha(B)u_t^2(\boldsymbol{\phi}) \\ &= \frac{1}{1 - \sum_{i=1}^p \beta_i} + \sum_{i=0}^{\infty}\sum_{k=1}^{d} \pi_k a_\beta(i) v_{k,t-i-1}^2 + \sum_{i=1}^{\infty} a_\gamma(i) u_{t-i}^2(\boldsymbol{\phi}), \end{aligned} \tag{S1.3}$$

where $u_t(\boldsymbol{\phi}) = u_t - (\boldsymbol{\phi} - \boldsymbol{\phi}_0)'\boldsymbol{X}_{t-1}$. It follows that

$$\begin{aligned}\sup_{\Theta} \sigma_t^2(\boldsymbol{\lambda}) &\leqslant \frac{1}{1-\sum_{i=1}^p \beta_i} + \sum_{i=0}^{\infty}\sum_{k=1}^{d} \pi_k \sup_{\Theta} a_\beta(i)\upsilon_{k,t-i-1}^2 + \sum_{i=1}^{\infty} \sup_{\Theta} a_\gamma(i) u_{t-i}^2(\boldsymbol{\phi}) \\ &\leqslant C\sum_{i=0}^{\infty} \rho^i [1 + \sup_{\Theta} u_{t-i}^2(\boldsymbol{\phi}) + \sum_{k=1}^{d} \pi_k \upsilon_{k,t-i-1}^2] \leqslant C\xi_{\rho,t-1}^2.\end{aligned}$$

Hence, (i) is asserted.

We then prove (ii). Since $q_t(\boldsymbol{\theta}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b\sigma_t(\boldsymbol{\lambda})$, it holds that

$$\frac{\partial q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(\sigma_t(\boldsymbol{\lambda}), \frac{b}{2\sigma_t(\boldsymbol{\lambda})}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}'}, \boldsymbol{X}_{t-1}' + \frac{b}{2\sigma_t(\boldsymbol{\lambda})}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}'}\right)'. \tag{S1.4}$$

Moreover, it can be verified that

$$\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}} = \beta^{-1}(B)\boldsymbol{z}_t(\boldsymbol{\lambda}) \quad \text{and} \quad \frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}} = -2\beta^{-1}(B)\alpha(B)u_t(\boldsymbol{\phi})\boldsymbol{X}_{t-1}, \tag{S1.5}$$

where $\boldsymbol{z}_t(\boldsymbol{\lambda}) = (u_{t-1}^2(\boldsymbol{\phi}), \ldots, u_{t-q}^2(\boldsymbol{\phi}), \sigma_{t-1}^2(\boldsymbol{\lambda}), \ldots, \sigma_{t-p}^2(\boldsymbol{\lambda}), \upsilon_{1,t-1}^2, \ldots, \upsilon_{d,t-1}^2)'$. From (S1.3), it holds that $\sigma_t^2(\boldsymbol{\lambda}) \geqslant a_\gamma(i)u_{t-i}^2(\boldsymbol{\phi})$ and $\sigma_t^2(\boldsymbol{\lambda}) \geqslant \pi_k a_\beta(i)\upsilon_{k,t-i-1}^2$. This together with (S1.1), (S1.2) and $\sigma_t^2(\boldsymbol{\lambda}) \geqslant 1$, implies that

$$\begin{aligned}
& \sup_{\Theta} \left\| \frac{1}{\sigma_t^2(\boldsymbol{\lambda})} \frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}} \right\| = \sup_{\Theta} \left\| \sum_{i=0}^{\infty} \frac{a_\beta(i) \boldsymbol{z}_{t-i}(\boldsymbol{\lambda})}{\sigma_t^2(\boldsymbol{\lambda})} \right\| \\
\leqslant \quad & \sup_{\Theta} \sum_{i=0}^{\infty} \left[\sum_{k=1}^{d} \frac{a_\beta(i) \upsilon_{k,t-i-1}^2}{\sigma_t^2(\boldsymbol{\lambda})} + \sum_{k=1}^{q} \frac{a_\beta(i) u_{t-i-k}^2(\boldsymbol{\phi})}{\sigma_t^2(\boldsymbol{\lambda})} + \sum_{k=1}^{p} \frac{a_\beta(i) \sigma_{t-i-k}^2(\boldsymbol{\lambda})}{\sigma_t^2(\boldsymbol{\lambda})} \right] \\
\leqslant \quad & \sum_{k=1}^{d} \frac{1}{\pi_k} + \sum_{k=1}^{q} \sum_{i=0}^{\infty} \left[\sup_{\Theta} \frac{a_\beta(i)}{a_\gamma(i+k)} \right]^{1-\iota/2} \sup_{\Theta} a_\beta^{\iota/2}(i) |u_{t-i-k}(\boldsymbol{\phi})|^{\iota} \\
& + \sum_{k=1}^{p} \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \left[\sup_{\Theta} \frac{a_\beta(i) a_\gamma(j)}{a_\gamma(i+j+k)} \right]^{1-\iota/2} \sup_{\Theta} [a_\beta(i) a_\gamma(j)]^{\iota/2} |u_{t-i-j-k}(\boldsymbol{\phi})|^{\iota} \\
\leqslant \quad & C \sum_{i=0}^{\infty} \rho^i [1 + \|\boldsymbol{X}_{t-i-1}\|^{\iota} + |u_{t-i}|^{\iota}], \qquad \text{(S1.6)}
\end{aligned}$$

and

$$\begin{aligned}
\sup_{\Theta} \left\| \frac{1}{\sigma_t^2(\boldsymbol{\lambda})} \frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}} \right\| \quad & \leqslant \quad 2 \sup_{\Theta} \left\| \sum_{i=1}^{\infty} \frac{a_\gamma(i) u_{t-i}(\boldsymbol{\phi}) \boldsymbol{X}_{t-i-1}}{\sigma_t(\boldsymbol{\lambda})} \right\| \\
& \leqslant \quad 2 \sum_{i=1}^{\infty} \sup_{\Theta} \sqrt{a_\gamma(i)} \|\boldsymbol{X}_{t-i-1}\| \leqslant C \sum_{i=1}^{\infty} \rho^i \|\boldsymbol{X}_{t-i-1}\|. \qquad \text{(S1.7)}
\end{aligned}$$

In view of (S1.4)-(S1.7), we have

$$\begin{aligned}
\sup_{\Theta} \left\| \frac{1}{\sigma_t(\boldsymbol{\lambda})} \frac{\partial q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \quad & \leqslant \quad 1 + \frac{\bar{b}}{2} \sup_{\Theta} \left\| \frac{1}{\sigma_t^2(\boldsymbol{\lambda})} \frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}} \right\| + \|\boldsymbol{X}_{t-1}\| + \frac{\bar{b}}{2} \sup_{\Theta} \left\| \frac{1}{\sigma_t^2(\boldsymbol{\lambda})} \frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}} \right\| \\
& \leqslant \quad C \sum_{i=0}^{\infty} \rho^i [1 + \|\boldsymbol{X}_{t-i-1}\| + |u_{t-i}|^{\iota}] \leqslant C \zeta_{\rho,t}.
\end{aligned}$$

We next consider the second derivatives. It holds that

$$\begin{aligned}
\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial b^2} &= 0, \;\; \frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial b \partial \boldsymbol{\gamma}'} = \frac{1}{2\sigma_t(\boldsymbol{\lambda})}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}'}, \;\; \frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial b \partial \boldsymbol{\phi}'} = \frac{1}{2\sigma_t(\boldsymbol{\lambda})}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}'}, \\
\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} &= -\frac{b}{4\sigma_t^3(\boldsymbol{\lambda})}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}'} + \frac{b}{2\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}, \\
\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\phi}'} &= -\frac{b}{4\sigma_t^3(\boldsymbol{\lambda})}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}'} + \frac{b}{2\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\phi}'}, \\
\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'} &= -\frac{b}{4\sigma_t^3(\boldsymbol{\lambda})}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}'} + \frac{b}{2\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}.
\end{aligned}$$

Moreover, it can be verified that

$$\frac{\partial^2 \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = \beta^{-1}(B)\frac{\partial \boldsymbol{z}_t(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}'}, \;\; \frac{\partial^2 \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\phi}'} = \beta^{-1}(B)\frac{\partial \boldsymbol{z}_t(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}'}, \;\; \frac{\partial^2 \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'} = 2\beta^{-1}(B)\alpha(B)\boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}',$$

where

$$\begin{aligned}
\frac{\partial \boldsymbol{z}_t(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}} &= \left(\mathbf{0}_{(p+q+d)\times q}, \frac{\partial \sigma_{t-1}^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}}, \cdots, \frac{\partial \sigma_{t-p}^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}}, \mathbf{0}_{(p+q+d)\times d}\right) \quad \text{and} \\
\frac{\partial \boldsymbol{z}_t(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}} &= \left(-2u_{t-1}(\boldsymbol{\phi})\boldsymbol{X}_{t-2}, \cdots, -2u_{t-q}(\boldsymbol{\phi})\boldsymbol{X}_{t-q-1}, \frac{\partial \sigma_{t-1}^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}}, \cdots, \frac{\partial \sigma_{t-p}^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}}, \mathbf{0}_{d\times d}\right).
\end{aligned}$$

Similar to the proof in (S1.6) and (S1.7), it can be verified that

$$\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial^2 \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'}\right\| \leqslant \sum_{i=0}^{\infty}\sup_{\Theta} a_\beta(i)\sum_{k=1}^{p}\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial \sigma_{t-i-k}^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}}\right\| \leqslant C\zeta_{\rho,t-1},$$

$$
\begin{aligned}
\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial^2\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\phi}'}\right\| &\leqslant 2\sum_{i=0}^{\infty}\sup_{\Theta}a_\beta(i)\sum_{k=1}^{q}\sup_{\Theta}\frac{|u_{t-i-k}(\boldsymbol{\phi})|\|\boldsymbol{X}_{t-i-k-1}\|}{\sigma_t(\boldsymbol{\lambda})} \\
&\quad+\sum_{i=0}^{\infty}\sup_{\Theta}a_\beta(i)\sum_{k=1}^{p}\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial\sigma_{t-i-k}^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\phi}}\right\| \\
&\leqslant 2\sum_{k=1}^{q}\sum_{i=0}^{\infty}\left[\sup_{\Theta}\frac{a_\beta(i)}{a_\gamma(i+k)}\right]^{1/2}\sup_{\Theta}a_\beta^{1/2}(i)\|\boldsymbol{X}_{t-i-k-1}\| \\
&\quad+2\sum_{k=1}^{p}\sum_{i=0}^{\infty}\sum_{j=1}^{\infty}\left[\sup_{\Theta}\frac{a_\beta(i)a_\gamma(j)}{a_\gamma(i+j+k)}\right]^{1/2}\sup_{\Theta}[a_\beta(i)a_\gamma(j)]^{1/2}\|\boldsymbol{X}_{t-i-j-k-1}\| \\
&\leqslant C\sum_{i=1}^{\infty}\rho^i\|\boldsymbol{X}_{t-i-1}\|,
\end{aligned}
$$

and

$$
\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial^2\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'}\right\| \leqslant 2\sum_{i=1}^{\infty}\sup_{\Theta}a_\gamma(i)\|\boldsymbol{X}_{t-i-1}\|^2 \leqslant C\sum_{i=1}^{\infty}\rho^i\|\boldsymbol{X}_{t-i-1}\|^2.
$$

Using above inequalities and (S1.6)-(S1.7), we can show that

$$
\begin{aligned}
\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\| &\leqslant \frac{1}{2}\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}}\right\|+\frac{1}{2}\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\phi}}\right\| \\
&\quad+\frac{\overline{b}}{4}\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}}\right\|^2+\frac{\overline{b}}{4}\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\phi}}\right\|^2 \\
&\quad+\frac{\overline{b}}{4}\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}}\right\|\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\phi}}\right\| \\
&\quad+\frac{\overline{b}}{2}\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial^2\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right\|+\frac{\overline{b}}{2}\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial^2\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\phi}'}\right\| \\
&\quad+\frac{\overline{b}}{2}\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial^2\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\phi}\partial\boldsymbol{\phi}'}\right\| \leqslant C\zeta_{\rho,t-1}^2.
\end{aligned}
$$

Then (ii) is verified. Note that $\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) \leqslant \sigma_t^2(\boldsymbol{\lambda})$ under the setting for initial values of

$u_t(\boldsymbol{\phi})$ and $\sigma_t^2(\boldsymbol{\lambda})$, then (iii) can be similarly asserted as for (ii). (iv) can be shown by using the same method in proving Lemma A.1(i) of Zheng et al. (2018). Hence, we accomplish the proof of this lemma. □

Proof of Lemma 2. In this paper, we use the setting for initial value of $u_t(\boldsymbol{\phi})$ as follows

$$\widetilde{u}_t(\boldsymbol{\phi}) = u_t(\boldsymbol{\phi}) \quad \text{for} \quad t > 0; \quad \widetilde{u}_t(\boldsymbol{\phi}) = 0 \quad \text{for} \quad t \leqslant 0. \tag{S1.8}$$

We first prove (i). Note that $\widetilde{\sigma}_t(\boldsymbol{\lambda}) - \sigma_t(\boldsymbol{\lambda}) = [\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) - \sigma_t^2(\boldsymbol{\lambda})]/[\widetilde{\sigma}_t(\boldsymbol{\lambda}) + \sigma_t(\boldsymbol{\lambda})]$. Moreover, by (S1.3), it follows that

$$\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) - \sigma_t^2(\boldsymbol{\lambda}) = \sum_{i=1}^{\infty} a_\gamma(i)[\widetilde{u}_{t-i}^2(\boldsymbol{\phi}) - u_{t-i}^2(\boldsymbol{\phi})] = -\sum_{i=t}^{\infty} a_\gamma(i) u_{t-i}^2(\boldsymbol{\phi}).$$

Then by (S1.1), (S1.8) and $\sigma_t^2(\boldsymbol{\lambda}) \geqslant a_\gamma(i)u_{t-i}^2(\boldsymbol{\phi})$ implied by (S1.3), together with the fact that $u_t(\boldsymbol{\phi}) = u_t - (\boldsymbol{\phi} - \boldsymbol{\phi}_0)'\boldsymbol{X}_{t-1}$, we have

$$\sup_{\Theta} |\widetilde{\sigma}_t(\boldsymbol{\lambda}) - \sigma_t(\boldsymbol{\lambda})| \leqslant \sum_{i=t}^{\infty} \sup_{\Theta} \frac{a_\gamma(i)u_{t-i}^2(\boldsymbol{\phi})}{\sigma_t(\boldsymbol{\lambda})} \leqslant \sum_{i=t}^{\infty} \sup_{\Theta} \sqrt{a_\gamma(i)}|u_{t-i}(\boldsymbol{\phi})| \leqslant C\rho^t \xi_\rho.$$

Hence, (i) holds.

We next verify (ii). By (S1.4) and the fact that $\sigma_t^2(\boldsymbol{\lambda}) \geqslant 1$, it can be verified that

$$\begin{aligned}\frac{1}{\sigma_t(\boldsymbol{\lambda})}\left\|\frac{\partial \widetilde{q}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}-\frac{\partial q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\| \;\leqslant\;& |\widetilde{\sigma}_t(\boldsymbol{\lambda})-\sigma_t(\boldsymbol{\lambda})|+\frac{b}{2}\frac{1}{\sigma_t(\boldsymbol{\lambda})}\left\|\frac{1}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}\frac{\partial \widetilde{\sigma}_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}}-\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\gamma}}\right\|\\ &+\frac{b}{2}\left\|\frac{1}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}\frac{\partial \widetilde{\sigma}_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}}-\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\phi}}\right\|. \qquad \text{(S1.9)}\end{aligned}$$

By (S1.3), it holds that

$$\begin{aligned}&\left\|\frac{\widetilde{\boldsymbol{z}}_{t-i}(\boldsymbol{\lambda})}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}-\frac{\boldsymbol{z}_{t-i}(\boldsymbol{\lambda})}{\sigma_t(\boldsymbol{\lambda})}\right\|\\ \leqslant\;& \sum_{k=1}^{q}\left|\frac{\widetilde{u}_{t-i-k}^2(\boldsymbol{\phi})}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}-\frac{u_{t-i-k}^2(\boldsymbol{\phi})}{\sigma_t(\boldsymbol{\lambda})}\right|+\sum_{k=1}^{p}\left|\frac{\widetilde{\sigma}_{t-i-k}^2(\boldsymbol{\lambda})}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}-\frac{\sigma_{t-i-k}^2(\boldsymbol{\lambda})}{\sigma_t(\boldsymbol{\lambda})}\right|+\sum_{k=1}^{d}\upsilon_{k,t-i-1}^2\left|\frac{1}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}-\frac{1}{\sigma_t(\boldsymbol{\lambda})}\right|\\ \leqslant\;& \sum_{k=1}^{q}\left|\frac{\widetilde{u}_{t-i-k}^2(\boldsymbol{\phi})}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}-\frac{u_{t-i-k}^2(\boldsymbol{\phi})}{\sigma_t(\boldsymbol{\lambda})}\right|+\sum_{k=1}^{p}\sum_{j=1}^{\infty}a_{\gamma}(j)\left|\frac{\widetilde{u}_{t-i-j-k}^2(\boldsymbol{\phi})}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}-\frac{u_{t-i-j-k}^2(\boldsymbol{\phi})}{\sigma_t(\boldsymbol{\lambda})}\right|\\ &+|\widetilde{\sigma}_t(\boldsymbol{\lambda})-\sigma_t(\boldsymbol{\lambda})|\sum_{k=1}^{d}\frac{\upsilon_{k,t-i-1}^2}{\widetilde{\sigma}_t(\boldsymbol{\lambda})\sigma_t(\boldsymbol{\lambda})}.\end{aligned}$$

This together with (S1.2), (S1.5), (S1.8) and the fact that $\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) \geqslant 1$, then similar to the proof of (S1.6), we can show that

$$\begin{aligned}
&\sup_{\Theta}\frac{1}{\sigma_t(\boldsymbol{\lambda})}\left\|\frac{1}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}\frac{\partial\widetilde{\sigma}_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}}-\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\gamma}}\right\| \\
\leqslant\ & \sup_{\Theta}|\widetilde{\sigma}_t(\boldsymbol{\lambda})-\sigma_t(\boldsymbol{\lambda})|\sum_{k=1}^{q}\sum_{i=0}^{t-k-1}\sup_{\Theta}a_\beta(i)\frac{u_{t-i-k}^2(\boldsymbol{\phi})}{\sigma_t^2(\boldsymbol{\lambda})}+\sum_{k=1}^{q}\sum_{i=t-k}^{\infty}\sup_{\Theta}a_\beta(i)\frac{u_{t-i-k}^2(\boldsymbol{\phi})}{\sigma_t^2(\boldsymbol{\lambda})} \\
&+\sup_{\Theta}|\widetilde{\sigma}_t(\boldsymbol{\lambda})-\sigma_t(\boldsymbol{\lambda})|\sum_{k=1}^{p}\sum_{j=1}^{\infty}\sum_{i=0}^{t-j-k-1}\sup_{\Theta}a_\beta(i)a_\gamma(j)\frac{u_{t-i-j-k}^2(\boldsymbol{\phi})}{\sigma_t^2(\boldsymbol{\lambda})} \\
&+\sum_{k=1}^{p}\sum_{j=1}^{\infty}\sum_{i=t-j-k}^{\infty}\sup_{\Theta}a_\beta(i)a_\gamma(j)\frac{u_{t-i-j-k}^2(\boldsymbol{\phi})}{\sigma_t^2(\boldsymbol{\lambda})}+\sup_{\Theta}|\widetilde{\sigma}_t(\boldsymbol{\lambda})-\sigma_t(\boldsymbol{\lambda})|\sum_{k=1}^{d}\frac{1}{\pi_k} \\
\leqslant\ & C\rho^t\xi_\rho\left[1+\sum_{i=0}^{t-2}\rho^i|u_{t-i}(\boldsymbol{\phi})|^\iota\right]. \qquad \text{(S1.10)}
\end{aligned}$$

Similarly, by (S1.2), (S1.5), (S1.8) and the fact that $\widetilde{\sigma}_t^2(\boldsymbol{\lambda})\geqslant 1$, we have

$$\begin{aligned}
&\sup_{\Theta}\left\|\frac{1}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}\frac{\partial\widetilde{\sigma}_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\phi}}-\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\phi}}\right\| \\
=\ & 2\sup_{\Theta}\left\|\sum_{i=1}^{\infty}a_\gamma(i)\left[\frac{\widetilde{u}_{t-i}(\boldsymbol{\phi})}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}-\frac{u_{t-i}(\boldsymbol{\phi})}{\sigma_t(\boldsymbol{\lambda})}\right]\boldsymbol{X}_{t-i-1}\right\| \\
\leqslant\ & 2\sum_{i=t}^{\infty}\sup_{\Theta}\frac{a_\gamma(i)|u_{t-i}(\boldsymbol{\phi})|}{\sigma_t(\boldsymbol{\lambda})}\left\|\boldsymbol{X}_{t-i-1}\right\|+2\sup_{\Theta}\frac{|\widetilde{\sigma}_t(\boldsymbol{\lambda})-\sigma_t(\boldsymbol{\lambda})|}{\widetilde{\sigma}_t(\boldsymbol{\lambda})}\sum_{i=1}^{t-1}\sup_{\Theta}\frac{a_\gamma(i)|u_{t-i}(\boldsymbol{\phi})|}{\sigma_t(\boldsymbol{\lambda})}\left\|\boldsymbol{X}_{t-i-1}\right\| \\
\leqslant\ & C\rho^t\xi_\rho\left[1+\sum_{i=0}^{t-2}\rho^i\left\|\boldsymbol{X}_{t-i-1}\right\|\right]. \qquad \text{(S1.11)}
\end{aligned}$$

In view of (S1.9)-(S1.11), together with (i) and $b\leqslant\bar{b}$ for $\boldsymbol{\theta}\in\Theta$, it follows that

$$\sup_{\Theta}\frac{1}{\sigma_t(\boldsymbol{\lambda})}\left\|\frac{\partial\widetilde{q}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}-\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\leqslant C\rho^t\xi_\rho\zeta_{\rho,t}.$$

Therefore, (ii) is asserted. The proof of this lemma is complete. □

Proof of Lemma 3. It can be verified that

$$|\zeta_n(\boldsymbol{u})| \leqslant \sqrt{n}\|\boldsymbol{u}\| \sum_{j=1}^{m+p+q+d+1} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} m_{t,j} \left\{ \xi_{1t}(\boldsymbol{u}) - E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}] \right\} \right|,$$

where $m_{t,j} = \sigma_t^{-1}\partial q_t(\boldsymbol{\theta}_{\tau 0})/\partial\theta_{(j)}$ with $\theta_{(j)}$ being the jth element of $\boldsymbol{\theta}$. For $1 \leqslant j \leqslant m+p+q+d+1$, define $g_t = \max_j\{m_{t,j}, 0\}$ or $g_t = \max_j\{-m_{t,j}, 0\}$. Let $f_t(\boldsymbol{u}) = g_t\xi_{1t}(\boldsymbol{u})$ and

$$D_n(\boldsymbol{u}) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{ f_t(\boldsymbol{u}) - E\left[f_t(\boldsymbol{u})|\mathcal{F}_{t-1} \right] \right\}.$$

To establish Lemma 3, it suffices to show that, for any $\delta > 0$,

$$\sup_{\|\boldsymbol{u}\| \leqslant \delta} \frac{|D_n(\boldsymbol{u})|}{1 + \sqrt{n}\|\boldsymbol{u}\|} = o_p(1). \tag{S1.12}$$

We follow the method in Lemma 4 of Pollard (1985) to verify (S1.12). Let $\mathfrak{F} = \{f_t(\boldsymbol{u}) : \|\boldsymbol{u}\| \leqslant \delta\}$ be a collection of functions indexed by $\boldsymbol{u}$. First, we verify that $\mathfrak{F}$ satisfies the bracketing condition defined in Pollard (1985), page 304. Let $B_r(\xi)$ be an open neighborhood of ξ with radius $r > 0$, and define a constant C_0 to be selected later. For any $\epsilon > 0$ and $0 < r \leqslant \delta$, there exists a sequence of small cubes $\{B_{\epsilon r/C_0}(\boldsymbol{u}_i)\}_{i=1}^{K(\epsilon)}$ to cover $B_r(\mathbf{0})$, where $K(\epsilon)$ is an integer less than $C\epsilon^{-(m+p+q+d+1)}$, and the constant C is not depending on ϵ and r; see Huber (1967), page 227. Denote $V_i(r) = B_{\epsilon r/C_0}(\boldsymbol{u}_i) \bigcap B_r(\mathbf{0})$, and let $U_1(r) = V_1(r)$ and $U_i(r) = V_i(r) - \bigcup_{j=1}^{i-1} V_j(r)$ for $i \geqslant 2$. Note that $\{U_i(r)\}_{i=1}^{K(\epsilon)}$ is a partition of $B_r(\mathbf{0})$. For each $\boldsymbol{u}_i \in U_i(r)$ with $1 \leqslant i \leqslant K(\epsilon)$, define the following bracketing functions

$$\begin{aligned} f_t^L(\boldsymbol{u}_i) &= g_t \int_0^1 \left[I\left(\varepsilon_t \leqslant b_\tau + \frac{\boldsymbol{u}_i'}{\sigma_t} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} s - \frac{\epsilon r}{C_0} \left\| \frac{1}{\sigma_t} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} \right\| \right) - I(\varepsilon_t \leqslant b_\tau) \right] ds, \\ f_t^U(\boldsymbol{u}_i) &= g_t \int_0^1 \left[I\left(\varepsilon_t \leqslant b_\tau + \frac{\boldsymbol{u}_i'}{\sigma_t} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} s + \frac{\epsilon r}{C_0} \left\| \frac{1}{\sigma_t} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} \right\| \right) - I(\varepsilon_t \leqslant b_\tau) \right] ds. \end{aligned}$$

The indicator function $I(\cdot)$ is non-decreasing and $g_t \geqslant 0$, for any $\boldsymbol{u} \in U_i(r)$, then

$$f_t^L(\boldsymbol{u}_i) \leqslant f_t(\boldsymbol{u}) \leqslant f_t^U(\boldsymbol{u}_i). \tag{S1.13}$$

Furthermore, by the Taylor expansion, it holds that

$$E\left[f_t^U(\boldsymbol{u}_i) - f_i^L(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \leqslant \frac{\epsilon r}{C_0} \cdot 2 \sup_{x \in \mathbb{R}} f_\varepsilon(x) \left\| \frac{1}{\sigma_t} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} \right\|^2. \tag{S1.14}$$

Denote $\Delta_t = 2 \sup_{x \in \mathbb{R}} f_\varepsilon(x) \|\sigma_t^{-1} \partial q_t(\boldsymbol{\theta}_{\tau 0}) / \partial \boldsymbol{\theta}\|^2$. By Assumption 4, we have $\sup_{x \in \mathbb{R}} f_\varepsilon(x) < \infty$. This together with Lemma 1, implies that $E(\Delta_t)$ exists. Let $C_0 = E(\Delta_t)$. Then by the iterated-expectation, it follows that

$$E\left[f_t^U(\boldsymbol{u}_i) - f_t^L(\boldsymbol{u}_i) \right] = E\left\{ E\left[f_t^U(\boldsymbol{\theta}_i) - f_t^L(\boldsymbol{\theta}_i) | \mathcal{F}_{t-1} \right] \right\} \leqslant \epsilon r.$$

This together with (S1.13), implies that the family $\mathfrak{F}$ satisfies the bracketing condition.

Put $r_k = 2^{-k}\delta$. Let $B(k) = B_{r_k}(\mathbf{0})$ and $A(k)$ be the annulus $B(k)\backslash B(k+1)$. From the bracketing condition, for fixed $\epsilon > 0$, there is a partition $U_1(r_k), U_2(r_k), \ldots, U_{K(\epsilon)}(r_k)$ of $B(k)$. First, consider the upper tail case. For $\boldsymbol{u} \in U_i(r_k)$, by (S1.14), it holds that

$$\begin{aligned} D_n(\boldsymbol{u}) &\leqslant \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{f_t^U(\boldsymbol{u}_i) - E\left[f_t^U(\boldsymbol{u}_i)|\mathcal{F}_{t-1}\right]\right\} + \frac{1}{\sqrt{n}}\sum_{t=1}^{n} E\left[f_t^U(\boldsymbol{u}_i) - f_t^L(\boldsymbol{u}_i)|\mathcal{F}_{t-1}\right] \\ &\leqslant D_n^U(\boldsymbol{u}_i) + \sqrt{n}\epsilon r_k \frac{1}{nC_0}\sum_{t=1}^{n}\Delta_t, \end{aligned} \tag{S1.15}$$

where

$$D_n^U(\boldsymbol{u}_i) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{f_t^U(\boldsymbol{u}_i) - E\left[f_t^U(\boldsymbol{u}_i)|\mathcal{F}_{t-1}\right]\right\}.$$

Define the event

$$E_n = \left\{\omega : \frac{1}{nC_0}\sum_{t=1}^{n}\Delta_t(\omega) < 2\right\}.$$

For $\boldsymbol{u} \in A(k)$, $1 + \sqrt{n}\|\boldsymbol{u}\| > \sqrt{n}r_{k+1} = \sqrt{n}r_k/2$. Then by (S1.15) and the

Chebyshev's inequality, we have

$$\begin{aligned}
& \mathrm{P}\left(\sup_{\boldsymbol{u}\in A(k)} \frac{D_n(\boldsymbol{u})}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\epsilon, E_n\right) \\
\leqslant\ & \mathrm{P}\left(\max_{1\leqslant i\leqslant K(\epsilon)} \sup_{\boldsymbol{u}\in U_i(r_k)\cap A(k)} D_n(\boldsymbol{u}) > 3\sqrt{n}\epsilon r_k, E_n\right) \\
\leqslant\ & K(\epsilon) \max_{1\leqslant i\leqslant K(\epsilon)} \mathrm{P}\left(D_n^U(\boldsymbol{u}_i) > \sqrt{n}\epsilon r_k\right) \\
\leqslant\ & K(\epsilon) \max_{1\leqslant i\leqslant K(\epsilon)} \frac{E\{[D_n^U(\boldsymbol{u}_i)]^2\}}{n\epsilon^2 r_k^2}. \qquad \text{(S1.16)}
\end{aligned}$$

Moreover, by the iterated-expectation, the Taylor expansion, Assumption 4 and $\|\boldsymbol{u}_i\| \leqslant r_k$ for $\boldsymbol{u}_i \in U_i(r_k)$, we have

$$\begin{aligned}
& E\left\{[f_t^U(\boldsymbol{u}_i)]^2\right\} = E\left\{E\left\{[f_t^U(\boldsymbol{u}_i)]^2|\mathcal{F}_{t-1}\right\}\right\} \\
\leqslant\ & 2E\left\{g_t^2\left|\int_0^1\left[F_\varepsilon\left(b_\tau + \frac{\boldsymbol{u}_i'}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau0})}{\partial\boldsymbol{\theta}}s + \frac{\epsilon r_k}{C_0}\left\|\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau0})}{\partial\boldsymbol{\theta}}\right\|\right) - F_\varepsilon(b_\tau)\right]ds\right|\right\} \\
\leqslant\ & C\sup_{x\in\mathbb{R}} f_\varepsilon(x) r_k E\left[\left\|\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau0})}{\partial\boldsymbol{\theta}}\right\|^3\right].
\end{aligned}$$

This, together with Lemma 1, $E(\|\boldsymbol{X}_t\|^{4+\delta}) < \infty$ for some $\delta > 0$ by Assumption 3, $\sup_{x\in\mathbb{R}} f_\varepsilon(x) < \infty$ by Assumption 4 and the fact that $f_t^U(\boldsymbol{u}_i) - E[f_t^U(\boldsymbol{u}_i)|\mathcal{F}_{t-1}]$ is a martingale difference sequence, implies that

$$\begin{aligned} E\{[D_n^U(\boldsymbol{u}_i)]^2\} &= \frac{1}{n}\sum_{t=1}^{n} E\{\{f_t^U(\boldsymbol{u}_i) - E[f_t^U(\boldsymbol{u}_i)|\mathcal{F}_{t-1}]\}^2\} \\ &\leqslant \frac{1}{n}\sum_{t=1}^{n} E\{[f_t^U(\boldsymbol{u}_i)]^2\} \\ &\leqslant \frac{Cr_k}{n}\sum_{t=1}^{n} E\left[\left\|\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}}\right\|^3\right] := \Delta(r_k). \end{aligned} \tag{S1.17}$$

Combing (S1.16) and (S1.17), we have

$$\mathrm{P}\left(\sup_{\boldsymbol{u}\in A(k)} \frac{D_n(\boldsymbol{u})}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\epsilon, E_n\right) \leqslant \frac{K(\epsilon)\Delta(r_k)}{n\epsilon^2 r_k^2}.$$

Similar to the proof of the upper tail case, we can obtain the same bound for the lower tail case. Therefore,

$$\mathrm{P}\left(\sup_{\boldsymbol{u}\in A(k)} \frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\epsilon, E_n\right) \leqslant \frac{2K(\epsilon)\Delta(r_k)}{n\epsilon^2 r_k^2}. \tag{S1.18}$$

Note that $\Delta(r_k) \to 0$ as $k \to \infty$, we can choose k_ϵ such that $2K(\epsilon)\Delta(r_k)/(\epsilon^2\delta^2) < \epsilon$ for $k \geqslant k_\epsilon$. Let k_n be the integer such that $n^{-1/2}\delta \leqslant r_{k_n} \leqslant 2n^{-1/2}\delta$, and split $B_\delta(\mathbf{0})$ into two events $B := B(k_n+1)$ and $B^c := B(0) - B(k_n+1)$. Note that $B^c = \bigcup_{k=0}^{k_n} A(k)$. Moreover, by Lemma 1, it follows that $\Delta(r_k)$ is bounded. This together with (S1.18), implies that

$$\begin{aligned}
\mathrm{P}\left(\sup_{\boldsymbol{u}\in B^c}\frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|}>6\epsilon\right) &\leqslant \sum_{k=0}^{k_n}\mathrm{P}\left(\sup_{\boldsymbol{u}\in A(k)}\frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|}>6\epsilon, E_n\right)+\mathrm{P}(E_n^c)\\
&\leqslant \frac{1}{n}\sum_{k=0}^{k_\epsilon-1}\frac{CK(\epsilon)}{\epsilon^2\delta^2}2^{2k}+\frac{\epsilon}{n}\sum_{k=k_\epsilon}^{k_n}2^{2k}+\mathrm{P}(E_n^c)\\
&\leqslant O\left(\frac{1}{n}\right)+4\epsilon+\mathrm{P}(E_n^c). \qquad \text{(S1.19)}
\end{aligned}$$

Furthermore, for $\boldsymbol{u}\in B$, we have $1+\sqrt{n}\|\boldsymbol{u}\|\geqslant 1$ and $r_{k_n+1}\leqslant n^{-1/2}\delta<n^{-1/2}$. Similar to the proof of (S1.16) and (S1.17), we can show that

$$\mathrm{P}\left(\sup_{\boldsymbol{u}\in B}\frac{D_n(\boldsymbol{u})}{1+\sqrt{n}\|\boldsymbol{u}\|}>3\epsilon, E_n\right)\leqslant \mathrm{P}\left(\max_{1\leqslant i\leqslant K(\epsilon)}D_n^U(\boldsymbol{u}_i)>\epsilon, E_n\right)\leqslant\frac{K(\epsilon)\Delta(r_{k_n+1})}{\epsilon^2}.$$

We can obtain the same bound for the lower tail. Therefore, we have

$$\begin{aligned}
\mathrm{P}\left(\sup_{\boldsymbol{u}\in B}\frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|}>3\epsilon\right) &= \mathrm{P}\left(\sup_{\boldsymbol{u}\in B}\frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|}>3\epsilon, E_n\right)+\mathrm{P}(E_n^c)\\
&\leqslant \frac{2K(\epsilon)\Delta(r_{k_n+1})}{\epsilon^2}+\mathrm{P}(E_n^c). \qquad \text{(S1.20)}
\end{aligned}$$

Note that $\Delta(r_{k_n+1})\to 0$ as $n\to\infty$. Moreover, by the ergodic theorem, $\mathrm{P}(E_n)\to 1$ as $n\to\infty$ and thus $\mathrm{P}(E_n^c)\to 0$ as $n\to\infty$. (S1.20) together with (S1.19) asserts (S1.12). The proof of this lemma is complete. □

Proof of Lemma 4. Denote $\boldsymbol{u}=\boldsymbol{\theta}-\boldsymbol{\theta}_{\tau 0}$. Recall that $\widehat{L}_n(\boldsymbol{\theta})=n^{-1}\sum_{t=1}^n\widehat{\sigma}_t^{-1}\widetilde{\ell}_t(\boldsymbol{\theta})$ and $\widetilde{L}_n(\boldsymbol{\theta})=n^{-1}\sum_{t=1}^n\widetilde{\sigma}_t^{-1}\ell_t(\boldsymbol{\theta})$, where $\widehat{\sigma}_t=\widetilde{\sigma}_t(\widehat{\boldsymbol{\lambda}}_n^{int})$, $\widetilde{\sigma}_t=\sigma_t(\widehat{\boldsymbol{\lambda}}_n^{int})$, $\widetilde{\ell}_t(\boldsymbol{\theta})=\rho_\tau[Y_t-\widetilde{q}_t(\boldsymbol{\theta})]$

and $\ell_t(\boldsymbol{\theta}) = \rho_\tau[Y_t - q_t(\boldsymbol{\theta})]$. Then it can be verified that

$$\begin{aligned} & n[\widehat{L}_n(\boldsymbol{\theta}) - \widehat{L}_n(\boldsymbol{\theta}_{\tau 0})] - n[\widetilde{L}_n(\boldsymbol{\theta}) - \widetilde{L}_n(\boldsymbol{\theta}_{\tau 0})] \\ = & \sum_{t=1}^{n} \widehat{\sigma}_t^{-1}[\widetilde{\ell}_t(\boldsymbol{\theta}) - \widetilde{\ell}_t(\boldsymbol{\theta}_{\tau 0})] - \sum_{t=1}^{n} \widetilde{\sigma}_t^{-1}[\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_{\tau 0})] \\ = & \widetilde{R}_{3n}(\boldsymbol{\theta}) + \widetilde{R}_{4n}(\boldsymbol{\theta}), \end{aligned} \tag{S1.21}$$

where

$$\begin{aligned} \widetilde{R}_{3n}(\boldsymbol{\theta}) &= \sum_{t=1}^{n} (\widehat{\sigma}_t^{-1} - \widetilde{\sigma}_t^{-1})[\widetilde{\ell}_t(\boldsymbol{\theta}) - \widetilde{\ell}_t(\boldsymbol{\theta}_{\tau 0})] \quad \text{and} \\ \widetilde{R}_{4n}(\boldsymbol{\theta}) &= \sum_{t=1}^{n} \widetilde{\sigma}_t^{-1}\{[\widetilde{\ell}_t(\boldsymbol{\theta}) - \widetilde{\ell}_t(\boldsymbol{\theta}_{\tau 0})] - [\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_{\tau 0})]\}. \end{aligned}$$

First, we show that

$$\widetilde{R}_{3n}(\boldsymbol{\theta}) = o_p(\sqrt{n}\|\boldsymbol{u}\|). \tag{S1.22}$$

By the Taylor expansion and the Lipschitz continuity of $\rho_\tau(x)$, we have

$$|\widetilde{\ell}_t(\boldsymbol{\theta}) - \widetilde{\ell}_t(\boldsymbol{\theta}_{\tau 0})| \leqslant C|\widetilde{q}_t(\boldsymbol{\theta}) - \widetilde{q}_t(\boldsymbol{\theta}_{\tau 0})| \leqslant C\|\boldsymbol{u}\| \left\| \frac{\partial \widetilde{q}_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} \right\|, \tag{S1.23}$$

where $\boldsymbol{\theta}^*$ is between $\boldsymbol{\theta}_{\tau 0}$ and $\boldsymbol{\theta}$. Recall that $\zeta_{\rho,t} = \sum_{j=0}^{\infty} \rho^j (1 + \|\boldsymbol{X}_{t-j-1}\| + |u_{t-j}|^\iota)$. By Assumptions 1 and 3, it holds that $E(\zeta_{\rho,t}^{4+\delta}) < \infty$. Combing (S1.63) and (S1.23), by Lemma 1, $\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) \leqslant \sigma_t^2(\boldsymbol{\lambda})$ and $\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) \geqslant 1$, we can show that

$$\sup_{\Theta} \frac{|\widetilde{R}_{3n}(\boldsymbol{\theta})|}{\sqrt{n}\|\boldsymbol{u}\|} \leqslant \frac{C\xi_\rho}{\sqrt{n}} \sum_{t=1}^{n} \rho^t \sup_{\Theta} \left[\left\| \frac{1}{\widetilde{\sigma}_t(\boldsymbol{\lambda})} \frac{\partial \widetilde{q}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \frac{\widetilde{\sigma}_t(\boldsymbol{\lambda})}{\widetilde{\sigma}_t} \right] \leqslant \frac{C\xi_\rho}{\sqrt{n}} \sum_{t=1}^{n} \rho^t \zeta_{\rho,t} \sup_{\Theta} \frac{\widetilde{\sigma}_t(\boldsymbol{\lambda})}{\widetilde{\sigma}_t} = o_p(1).$$

Thus, (S1.22) holds. Next, we verify that

$$\widetilde{R}_{4n}(\boldsymbol{\theta}) = o_p(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2). \tag{S1.24}$$

Denote $\nu_t(\boldsymbol{u}) = q_t(\boldsymbol{\theta}) - q_t(\boldsymbol{\theta}_{\tau 0})$ and $\widetilde{\nu}_t(\boldsymbol{u}) = \widetilde{q}_t(\boldsymbol{\theta}) - \widetilde{q}_t(\boldsymbol{\theta}_{\tau 0})$. Define

$$\begin{aligned} \xi_t(\boldsymbol{u}) &= \int_0^1 \left[I\left(\varepsilon_t \leqslant b_\tau + \sigma_t^{-1}\nu_t(\boldsymbol{u})s\right) - I(\varepsilon_t \leqslant b_\tau) \right] ds \quad \text{and} \\ \widetilde{\xi}_t(\boldsymbol{u}) &= \int_0^1 \left[I\left(\varepsilon_t \leqslant b_\tau \widetilde{\sigma}_t \sigma_t^{-1} + \sigma_t^{-1}\widetilde{\nu}_t(\boldsymbol{u})s\right) - I(\varepsilon_t \leqslant b_\tau \widetilde{\sigma}_t \sigma_t^{-1}) \right] ds. \end{aligned}$$

By the Knight equation (S1.66), it can be verified that

$$\begin{aligned} \widetilde{R}_{4n}(\boldsymbol{\theta}) &= \sum_{t=1}^{n} \widetilde{\sigma}_t^{-1} \left\{ \widetilde{\nu}_t(\boldsymbol{u})[-\psi_\tau(\varepsilon_t - b_\tau \widetilde{\sigma}_t \sigma_t^{-1}) + \widetilde{\xi}_t(\boldsymbol{u})] - \nu_t(\boldsymbol{u})\left[-\psi_\tau(\varepsilon_{t,\tau}) + \xi_t(\boldsymbol{u})\right] \right\} \\ &= \widetilde{\Pi}_1(\boldsymbol{u}) + \widetilde{\Pi}_2(\boldsymbol{u}) + \widetilde{\Pi}_3(\boldsymbol{u}) + \widetilde{\Pi}_4(\boldsymbol{u}), \end{aligned} \tag{S1.25}$$

where $\varepsilon_{t,\tau} = \varepsilon_t - b_\tau$,

$$\begin{aligned}
\widetilde{\Pi}_1(\boldsymbol{u}) &= -\sum_{t=1}^{n} \widetilde{\sigma}_t^{-1}[\widetilde{\nu}_t(\boldsymbol{u}) - \nu_t(\boldsymbol{u})]\psi_\tau(\varepsilon_t - b_\tau\widetilde{\sigma}_t\sigma_t^{-1}), \\
\widetilde{\Pi}_2(\boldsymbol{u}) &= -\sum_{t=1}^{n} \widetilde{\sigma}_t^{-1}\nu_t(\boldsymbol{u})[\psi_\tau(\varepsilon_t - b_\tau\widetilde{\sigma}_t\sigma_t^{-1}) - \psi_\tau(\varepsilon_{t,\tau})], \\
\widetilde{\Pi}_3(\boldsymbol{u}) &= \sum_{t=1}^{n} \widetilde{\sigma}_t^{-1}[\widetilde{\nu}_t(\boldsymbol{u}) - \nu_t(\boldsymbol{u})]\widetilde{\xi}_t(\boldsymbol{u}) \quad \text{and} \quad \widetilde{\Pi}_4(\boldsymbol{u}) = \sum_{t=1}^{n} \widetilde{\sigma}_t^{-1}\nu_t(\boldsymbol{u})[\widetilde{\xi}_t(\boldsymbol{u}) - \xi_t(\boldsymbol{u})].
\end{aligned}$$

Moreover, by the Taylor expansion, we have

$$\nu_t(\boldsymbol{u}) = \boldsymbol{u}'\frac{\partial q_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} \quad \text{and} \quad \widetilde{\nu}_t(\boldsymbol{u}) = \boldsymbol{u}'\frac{\partial \widetilde{q}_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}, \tag{S1.26}$$

where $\boldsymbol{\theta}^*$ is between $\boldsymbol{\theta}_{\tau 0}$ and $\boldsymbol{\theta}$. Then it follows that

$$\widetilde{\Pi}_1(\boldsymbol{u}) = -\boldsymbol{u}'\sum_{t=1}^{n}\frac{1}{\widetilde{\sigma}_t}\left[\frac{\partial \widetilde{q}_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}} - \frac{\partial q_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}\right]\psi_\tau(\varepsilon_t - b_\tau\widetilde{\sigma}_t\sigma_t^{-1}).$$

By Lemma 1(iv), Lemma 2(ii) and Assumption 1, together with the fact that $|\psi_\tau(x)| < 1$, it follows that

$$\begin{aligned}
\sup_{\Theta}\frac{|\widetilde{\Pi}_1(\boldsymbol{u})|}{\sqrt{n}\|\boldsymbol{u}\|} &\leqslant \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sup_{\Theta}\frac{1}{\sigma_t(\boldsymbol{\lambda})}\left\|\frac{\partial \widetilde{q}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\|\sup_{\Theta}\frac{\sigma_t(\boldsymbol{\lambda})}{\sigma_t(\widetilde{\boldsymbol{\lambda}}_n)} \\
&\leqslant \frac{C\xi_\rho}{\sqrt{n}}\sum_{t=1}^{n}\rho^t\zeta_{\rho,t}\sup_{\Theta}\frac{\sigma_t(\boldsymbol{\lambda})}{\sigma_t(\widetilde{\boldsymbol{\lambda}}_n)} = o_p(1).
\end{aligned}$$

Hence, we have

$$\widetilde{\Pi}_1(\boldsymbol{u}) = o_p(\sqrt{n}\|\boldsymbol{u}\|). \tag{S1.27}$$

We next consider $\widetilde{\Pi}_2(\boldsymbol{u})$. Since $\psi_\tau(x) = \tau - I(x < 0)$, by the Taylor expansion, we have

$$E\left[\psi_\tau(\varepsilon_t - b_\tau\widetilde{\sigma}_t\sigma_t^{-1}) - \psi_\tau(\varepsilon_{t,\tau})|\mathcal{F}_{t-1}\right] = f_\varepsilon(b_{\tau 1})b_\tau\sigma_t^{-1}(\sigma_t - \widetilde{\sigma}_t),$$

where $b_{\tau 1}$ is between b_τ and $b_\tau\widetilde{\sigma}_t\sigma_t^{-1}$. As a result, by the iterated-expectation, the Cauchy-Schwarz inequality and the Taylor expansion in (S1.26), together with $\widetilde{\sigma}_t^2 \geqslant 1$, Lemmas 1-2, $E(\zeta_{\rho,t}^{4+\delta}) < \infty$ by Assumption 3, $E(\xi_\rho^2) < \infty$ by Assumption 1 and $E(u_t^2) < \infty$, and $\sup_{x\in\mathbb{R}} f_\varepsilon(x) < \infty$ by Assumption 4, it follows that

$$\begin{aligned}
E\left[\sup_{\Theta}\frac{|\widetilde{\Pi}_2(\boldsymbol{u})|}{\sqrt{n}\|\boldsymbol{u}\|}\right] &\leqslant \frac{1}{\sqrt{n}}\sum_{t=1}^{n} E\left\{\frac{1}{\widetilde{\sigma}_t}\sup_{\Theta}\left\|\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\left|E\left[\psi_\tau(\varepsilon_t - b_\tau\widetilde{\sigma}_t\sigma_t^{-1}) - \psi_\tau(\varepsilon_{t,\tau})|\mathcal{F}_{t-1}\right]\right|\right\} \\
&\leqslant \bar{b}\sup_{x\in\mathbb{R}} f_\varepsilon(x)\frac{1}{\sqrt{n}}\sum_{t=1}^{n} E\left\{|\sigma_t - \widetilde{\sigma}_t|\sup_{\Theta}\frac{\sigma_t(\boldsymbol{\lambda})}{\sigma_t}\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\right\} \\
&\leqslant \frac{C}{\sqrt{n}}\sum_{t=1}^{n}\rho^t\left[E(\xi_\rho^2)\right]^{1/2}\left[E\sup_{\Theta}\frac{\sigma_t^4(\boldsymbol{\lambda})}{\sigma_t^4}\right]^{1/4}\left[E(\zeta_{\rho,t}^4)\right]^{1/4} = o(1).
\end{aligned}$$

Therefore,

$$\widetilde{\Pi}_2(\boldsymbol{u}) = o_p(\sqrt{n}\|\boldsymbol{u}\|). \tag{S1.28}$$

Note that $|\widetilde{\xi}_t(\boldsymbol{u})| < 2$. For $\widetilde{\Pi}_3(\boldsymbol{u})$, similar to the proof of $\widetilde{\Pi}_1(\boldsymbol{u})$, we can show that

$$\widetilde{\Pi}_3(\boldsymbol{u}) = o_p(\sqrt{n}\|\boldsymbol{u}\|). \tag{S1.29}$$

Finally, we consider $\widetilde{\Pi}_4(\boldsymbol{u})$. By the Taylor expansion, we have

$$\begin{aligned}
& E[\widetilde{\xi}_t(\boldsymbol{u}) - \xi_t(\boldsymbol{u})|\mathcal{F}_{t-1}] \\
= & \int_0^1 \left[F_\varepsilon\left(\frac{b_\tau\widetilde{\sigma}_t}{\sigma_t} + \frac{\widetilde{\nu}_t(\boldsymbol{u})}{\sigma_t}s\right) - F_\varepsilon\left(\frac{b_\tau\widetilde{\sigma}_t}{\sigma_t}\right)\right] ds - \int_0^1 \left[F_\varepsilon\left(b_\tau + \frac{\nu_t(\boldsymbol{u})}{\sigma_t}s\right) - F_\varepsilon(b_\tau)\right] ds \\
= & \int_0^1 f_\varepsilon(b_{\tau 2})\sigma_t^{-1}\widetilde{\nu}_t(\boldsymbol{u})sds - \int_0^1 f_\varepsilon(b_{\tau 3})\sigma_t^{-1}\nu_t(\boldsymbol{u})sds \\
= & \frac{1}{2}f_\varepsilon(b_\tau)\sigma_t^{-1}[\widetilde{\nu}_t(\boldsymbol{u}) - \nu_t(\boldsymbol{u})] + \sigma_t^{-1}\widetilde{\nu}_t(\boldsymbol{u})\int_0^1 [f_\varepsilon(b_{\tau 2}) - f_\varepsilon(b_\tau)]sds \\
& -\sigma_t^{-1}\nu_t(\boldsymbol{u})\int_0^1 [f_\varepsilon(b_{\tau 3}) - f_\varepsilon(b_\tau)]sds,
\end{aligned}$$

where $b_{\tau 2} = b_\tau\widetilde{\sigma}_t\sigma_t^{-1} + \sigma_t^{-1}\widetilde{\nu}_t(\boldsymbol{u})s_2$ and $b_{\tau 3} = b_\tau + \sigma_t^{-1}\nu_t(\boldsymbol{u})s_3$ with $0 < s_2, s_3 < s \leqslant 1$. As a result, by the iterated-expectation and the Taylor expansion in (S1.26), it follows that

$$E\left[\sup_{\Theta}\frac{|\widetilde{\Pi}_4(\boldsymbol{u})|}{n\|\boldsymbol{u}\|^2}\right] \leqslant \widetilde{K}_1(\boldsymbol{u}) + \widetilde{K}_2(\boldsymbol{u}) + \widetilde{K}_3(\boldsymbol{u}), \tag{S1.30}$$

where

$$\begin{aligned}
\widetilde{K}_1(\boldsymbol{u}) &= \frac{f_\varepsilon(b_\tau)}{2n}\sum_{t=1}^n E\left[\frac{1}{\widetilde{\sigma}_t\sigma_t}\sup_\Theta\left\|\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\left\|\frac{\partial \widetilde{q}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}-\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right\|\right],\\
\widetilde{K}_2(\boldsymbol{u}) &= \frac{1}{n}\sum_{t=1}^n E\left[\frac{1}{\widetilde{\sigma}_t\sigma_t}\sup_\Theta\left\|\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\left\|\frac{\partial \widetilde{q}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right\|\sup_\Theta\int_0^1|f_\varepsilon(b_{\tau 2})-f_\varepsilon(b_\tau)|sds\right],\\
\widetilde{K}_3(\boldsymbol{u}) &= \sum_{t=1}^n E\left[\frac{1}{\widetilde{\sigma}_t\sigma_t}\sup_\Theta\left\|\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^2\sup_\Theta\int_0^1|f_\varepsilon(b_{\tau 3})-f_\varepsilon(b_\tau)|sds\right].
\end{aligned}$$

First, consider $\widetilde{K}_1(\boldsymbol{\theta})$. By the Hölder inequality and Lemmas 1-2, together with $E(\zeta_{\rho,t}^{4+\delta})<\infty$ by Assumption 3, $E(\xi_\rho^2)<\infty$ by Assumption 1 and $E(u_t^2)<\infty$, and the fact that $\widetilde{\sigma}_t^2\leqslant\sigma_t^2$, we have

$$\begin{aligned}
\widetilde{K}_1(\boldsymbol{u}) &\leqslant \frac{f_\varepsilon(b_\tau)}{2}\frac{1}{n}\sum_{t=1}^n E\left\{\sup_\Theta\frac{\sigma_t^2(\boldsymbol{\lambda})}{\sigma_t\widetilde{\sigma}_t}\sup_\Theta\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\sup_\Theta\frac{1}{\sigma_t(\boldsymbol{\lambda})}\left\|\frac{\partial \widetilde{q}_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}-\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|\right\}\\
&\leqslant \frac{C}{n}\sum_{t=1}^n\rho^t E\left\{\sup_\Theta\frac{\sigma_t^2(\boldsymbol{\lambda})}{\sigma_t^2(\widetilde{\boldsymbol{\lambda}}_n)}\zeta_{\rho,t}^2\xi_\rho\right\}\\
&\leqslant \frac{C}{n}\sum_{t=1}^n\rho^t[E(\xi_\rho^2)]^{\frac{1}{2}}\left[E(\zeta_{\rho,t}^{4+\delta})\right]^{\frac{2}{4+\delta}}\left\{E\sup_\Theta\left[\frac{\sigma_t(\boldsymbol{\lambda})}{\sigma_t}\right]^{\frac{4(4+\delta)}{\delta}}\right\}^{\frac{\delta}{2(4+\delta)}} = o(1).
\end{aligned}$$

Hence,

$$\widetilde{K}_1(\boldsymbol{u}) = o_p(1). \tag{S1.31}$$

Next, we consider $\widetilde{K}_2(\boldsymbol{u})$. By the Taylor expansion, it follows that

$$\int_0^1 |f_\varepsilon(b_{\tau 2}) - f_\varepsilon(b_\tau)| s ds \leqslant \frac{1}{2\sigma_t} \sup_{x\in\mathbb{R}} |\dot{f}_\varepsilon(x)||b_\tau(\widetilde{\sigma}_t - \sigma_t) + \widetilde{\nu}_t(\boldsymbol{u}) s_2|.$$

Then by the Taylor expansion in (S1.26) and the Hölder inequality, together with Lemmas 1-2, $\sup_{x\in\mathbb{R}} |\dot{f}_\varepsilon(x)| < \infty$ by Assumption 4, $|b_\tau| < \bar{b}$ and the fact that $\sigma_t^2 \geqslant 1$, we have

$$\begin{aligned}
\sup_{\|\boldsymbol{u}\|\leqslant\eta} \widetilde{K}_2(\boldsymbol{u}) &\leqslant \frac{C}{n}\sum_{t=1}^n E\left\{\sup_{\Theta}\left[\left\|\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\|\left\|\frac{1}{\widetilde{\sigma}_t}\frac{\partial \widetilde{q}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\|\right]\left[\frac{\bar{b}}{\sigma_t}|\widetilde{\sigma}_t - \sigma_t| + \sup_{\|\boldsymbol{u}\|\leqslant\eta}\frac{|\widetilde{\nu}_t(\boldsymbol{u})|}{\sigma_t}s_2\right]\right\} \\
&\leqslant \frac{C}{n}\sum_{t=1}^n \rho^t \left\{E\left[\sup_{\Theta}\frac{\sigma_t(\boldsymbol{\lambda})}{\sigma_t}\right]^{\frac{4(4+\delta)}{\delta}}\right\}^{\frac{\delta}{2(4+\delta)}} [E(\zeta_{\rho,t}^{4+\delta})]^{\frac{2}{4+\delta}}[E(\xi_\rho^2)]^{\frac{1}{2}} \\
&\quad + \frac{C\eta}{n}\sum_{t=1}^n \left\{E\left[\sup_{\Theta}\frac{\sigma_t(\boldsymbol{\lambda})}{\sigma_t}\right]^{\frac{3(3+\delta)}{\delta}}\right\}^{\frac{\delta}{3+\delta}} [E(\zeta_{\rho,t}^{3+\delta})]^{\frac{3}{3+\delta}}
\end{aligned}$$

tends to 0 as $\eta \to 0$ and n is large enough. Similar to (S1.50) and (S1.51), we can show that

$$\widetilde{K}_2(\boldsymbol{u}) = o_p(1). \tag{S1.32}$$

We finally consider $\widetilde{K}_3(\boldsymbol{u})$. By the Taylor expansion, we have

$$\int_0^1 |f_\varepsilon(b_{\tau 3}) - f_\varepsilon(b_\tau)| s ds \leqslant \sup_{x\in\mathbb{R}} |\dot{f}_\varepsilon(x)| \sigma_t^{-1} |\nu_t(\boldsymbol{u})|.$$

Then by the Hölder inequality and Lemma 1, it follows that

$$\begin{aligned}\sup_{\|\boldsymbol{u}\|\leqslant\eta}\widetilde{K}_3(\boldsymbol{u}) &\leqslant \frac{C}{n}\sum_{t=1}^{n}E\left\{\sup_{\Theta}\frac{1}{\widetilde{\sigma}_t\sigma_t}\left\|\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^2\sup_{\|\boldsymbol{u}\|\leqslant\eta}\left|\frac{\nu_t(\boldsymbol{u})}{\sigma_t}\right|\right\} \\ &\leqslant \frac{C\eta}{n}\sum_{t=1}^{n}E\left[\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^3\sup_{\Theta}\frac{\sigma_t^3(\boldsymbol{\lambda})}{\sigma_t^2\sigma_t(\widetilde{\boldsymbol{\lambda}}_n)}\right] \\ &\leqslant C\eta\left[E(\zeta_{\rho,t}^{3+\delta})\right]^{\frac{3}{\delta+3}}\left\{\sup_{\Theta}\left[\frac{\sigma_t(\boldsymbol{\lambda})}{\sigma_t}\right]^{\frac{3(3+\delta)}{\delta}}\right\}^{\frac{\delta}{\delta+3}}\end{aligned}$$

tends to 0 as $\eta\to 0$ and n is large enough. Similar to (S1.50) and (S1.51), we can show that

$$\widetilde{K}_3(\boldsymbol{u}) = o_p(1). \tag{S1.33}$$

In view of (S1.30)-(S1.33), we have

$$\widetilde{\Pi}_4(\boldsymbol{u}) = o_p(n\|\boldsymbol{u}\|^2). \tag{S1.34}$$

Combing (S1.25), (S1.27)-(S1.29) and (S1.34), we assert (S1.24). By (S1.21), (S1.22) and (S1.24), it follows that

$$n[\widehat{L}_n(\boldsymbol{\theta})-\widehat{L}_n(\boldsymbol{\theta}_{\tau 0})]-n[\widetilde{L}_n(\boldsymbol{\theta})-\widetilde{L}_n(\boldsymbol{\theta}_{\tau 0})] = o_p(\sqrt{n}\|\boldsymbol{u}\|+n\|\boldsymbol{u}\|^2). \tag{S1.35}$$

Hence, the proof of this lemma is complete. □

Proof of Lemma 5. Denote $\boldsymbol{u} = \boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}$. Recall that $\widetilde{L}_n(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^n \widetilde{\sigma}_t^{-1}\ell_t(\boldsymbol{\theta})$ and $L_n(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^n \sigma_t^{-1}\ell_t(\boldsymbol{\theta})$, where $\ell_t(\boldsymbol{\theta}) = \rho_\tau[Y_t - q_t(\boldsymbol{\theta})]$. To show this lemma, we decompose the proof into two steps. In the first step, we will show that

$$n[\widetilde{L}_n(\boldsymbol{\theta}) - \widetilde{L}_n(\boldsymbol{\theta}_{\tau 0})] - n[L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_{\tau 0})] = o_p(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2). \tag{S1.36}$$

Denote $\varepsilon_{t,\tau} = \varepsilon_t - b_\tau$ and $\nu_t(\boldsymbol{u}) = q_t(\boldsymbol{\theta}) - q_t(\boldsymbol{\theta}_{\tau 0})$. Define

$$\xi_t(\boldsymbol{u}) = \int_0^1 \left[I\left(\varepsilon_t \leqslant b_\tau + \sigma_t^{-1}\nu_t(\boldsymbol{u})s\right) - I(\varepsilon_t \leqslant b_\tau)\right] ds. \tag{S1.37}$$

By the Knight equation (S1.66), it can be verified that

$$\begin{aligned}
& n[\widetilde{L}_n(\boldsymbol{\theta}) - \widetilde{L}_n(\boldsymbol{\theta}_{\tau 0})] - n[L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_{\tau 0})] \\
= & \sum_{t=1}^n (\widetilde{\sigma}_t^{-1} - \sigma_t^{-1})\left[\rho_\tau\left(\varepsilon_{t,\tau}\sigma_t - \nu_t(\boldsymbol{u})\right) - \rho_\tau\left(\varepsilon_{t,\tau}\sigma_t\right)\right] \\
= & K_{1n}(\boldsymbol{u}) + K_{2n}(\boldsymbol{u}),
\end{aligned} \tag{S1.38}$$

where

$$K_{1n}(\boldsymbol{u}) = -\sum_{t=1}^n \left(\frac{1}{\widetilde{\sigma}_t} - \frac{1}{\sigma_t}\right)\nu_t(\boldsymbol{u})\psi_\tau(\varepsilon_{t,\tau}) \text{ and } K_{2n}(\boldsymbol{u}) = \sum_{t=1}^n \left(\frac{1}{\widetilde{\sigma}_t} - \frac{1}{\sigma_t}\right)\nu_t(\boldsymbol{u})\xi_t(\boldsymbol{u}).$$

By the Taylor expansion, it holds that

$$\frac{1}{\widetilde{\sigma}_t} - \frac{1}{\sigma_t} = -(\widehat{\boldsymbol{\lambda}}_n^{int} - \boldsymbol{\lambda}_0)' \frac{1}{2\sigma_t^3(\boldsymbol{\lambda}^*)} \frac{\partial \sigma_t^2(\boldsymbol{\lambda}^*)}{\partial \boldsymbol{\lambda}} \text{ and } \nu_t(\boldsymbol{u}) = \boldsymbol{u}' \frac{\partial q_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}, \tag{S1.39}$$

where $\boldsymbol{\lambda}^*$ is between $\widehat{\boldsymbol{\lambda}}_n^{int}$ and $\boldsymbol{\lambda}_0$, and $\boldsymbol{\theta}^*$ is between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{\tau 0}$. By (S1.39), it follows that

$$K_{1n}(\boldsymbol{u}) = \sqrt{n}(\widehat{\boldsymbol{\lambda}}_n^{int} - \boldsymbol{\lambda}_0)' \cdot \frac{1}{n} \sum_{t=1}^{n} Z_{1t}(\boldsymbol{\theta}^*) \cdot \sqrt{n}\boldsymbol{u}, \tag{S1.40}$$

where

$$Z_{1t}(\boldsymbol{\theta}^*) = \frac{1}{2\sigma_t^3(\boldsymbol{\lambda}^*)} \frac{\partial \sigma_t^2(\boldsymbol{\lambda}^*)}{\partial \boldsymbol{\lambda}} \frac{\partial q_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}'} \psi_\tau(\varepsilon_{t,\tau}).$$

By the iterated-expectation and Lemma 1, together with the fact that $E[\psi_\tau(\varepsilon_{t,\tau})] = 0$, we can show that $E\left[Z_{1t}(\boldsymbol{\theta}^*)\right] = 0$. Furthermore, by the Cauchy-Schwarz inequality, Lemma 1 and the fact that $|\psi_\tau(x)| \leqslant 1$, we have

$$E\left(\sup_{\Theta} \|Z_{1t}(\boldsymbol{\theta}^*)\|\right) \leqslant \frac{1}{2}\left[E \sup_{\Theta} \left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})} \frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right\|^2\right]^{1/2} \left[E \sup_{\Theta} \left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})} \frac{\partial q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\|^2\right]^{1/2} < \infty.$$

Then by the Theorem 3.1 in Ling and McAleer (2003), it follows that

$$\sup_{\Theta} \left\| \frac{1}{n} \sum_{t=1}^{n} Z_{1t}(\boldsymbol{\theta}^*) \right\| = o_p(1).$$

This together with (S1.40) and the fact that $\sqrt{n}(\widehat{\boldsymbol{\lambda}}_n^{int} - \boldsymbol{\lambda}_0) = O_p(1)$, implies that

$$K_{1n}(\boldsymbol{u}) = o_p(\sqrt{n}\|\boldsymbol{u}\|). \tag{S1.41}$$

For $K_{2n}(\boldsymbol{u})$, by (S1.39), it can be verified that

$$K_{2n}(\boldsymbol{u}) = -(\check{\boldsymbol{\lambda}}_n - \boldsymbol{\lambda}_0)' Z_{2t}(\boldsymbol{u}), \tag{S1.42}$$

where

$$Z_{2t}(\boldsymbol{u}) = \sum_{t=1}^{n} \frac{1}{2\sigma_t^3(\boldsymbol{\lambda}^*)} \frac{\partial \sigma_t^2(\boldsymbol{\lambda}^*)}{\partial \boldsymbol{\lambda}} \nu_t(\boldsymbol{u}) \xi_t(\boldsymbol{u}).$$

By the Taylor expansion and Assumption 4, it follows that

$$E[\xi_t(\boldsymbol{u})|\mathcal{F}_{t-1}] = \int_0^1 \left[F_\varepsilon\left(b_\tau + \frac{\nu_t(\boldsymbol{u})s}{\sigma_t}\right) - F_\varepsilon(b_\tau) \right] ds = \int_0^1 f_\varepsilon\left(b_\tau + \frac{\nu_t(\boldsymbol{u})s^*}{\sigma_t}\right) \frac{\nu_t(\boldsymbol{u})}{\sigma_t} s ds,$$

where s^* is between 0 and s. Then by the iterated-expectation and the Cauchy-Schwarz inequality, together with (S1.39), Lemma 1 and $\sup_{x \in \mathbb{R}} f_\varepsilon(x) < \infty$ by Assumption 4, we have

$$
\begin{aligned}
& E\left[\sup_{\Theta}\frac{\|Z_{2t}(\boldsymbol{u})\|}{n\|\boldsymbol{u}\|^2}\right] \\
\leqslant\ & \frac{1}{4}\sup_{x\in\mathbb{R}} f_{\varepsilon}(x)\frac{1}{n}\sum_{t=1}^{n} E\left[\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\right\|\sup_{\Theta}\frac{\nu_t^2(\boldsymbol{\theta})}{\sigma_t\sigma_t(\boldsymbol{\lambda}^*)\|\boldsymbol{u}\|^2}\right] \\
\leqslant\ & CE\left[\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\right\|\sup_{\Theta}\frac{\sigma_t(\boldsymbol{\lambda})}{\sigma_t}\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^2\right] \\
\leqslant\ & C\left[E\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial\sigma_t^2(\boldsymbol{\lambda})}{\partial\boldsymbol{\lambda}}\right\|^4\right]^{1/4}\left[E\sup_{\Theta}\frac{\sigma_t^4(\boldsymbol{\lambda})}{\sigma_t^4}\right]^{1/4}\left[E\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\|^4\right]^{1/2} < \infty.
\end{aligned}
$$

This together with (S1.42), the ergodic theorem and the fact that $\sqrt{n}(\widehat{\boldsymbol{\lambda}}_n^{int} - \boldsymbol{\lambda}_0) = O_p(1)$, implies that

$$
K_{2n}(\boldsymbol{u}) = o_p(n\|\boldsymbol{u}\|^2). \tag{S1.43}
$$

By (S1.38), (S1.41) and (S1.43), it follows that (S1.36) holds.

In the second step, we will establish that

$$
n[L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_{\tau 0})] = -\sqrt{n}\boldsymbol{u}'\boldsymbol{T}_n + \sqrt{n}\boldsymbol{u}'J_n\sqrt{n}\boldsymbol{u} + o_p(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2). \tag{S1.44}
$$

By the Knight equation (S1.66), we have

$$
n[L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}_{\tau 0})] = R_{1n}(\boldsymbol{u}) + R_{2n}(\boldsymbol{u}), \tag{S1.45}
$$

where

$$R_{1n}(\boldsymbol{u}) = -\sum_{t=1}^{n}\frac{1}{\sigma_t}\nu_t(\boldsymbol{u})\psi_\tau(\varepsilon_{t,\tau}) \text{ and } R_{2n}(\boldsymbol{u}) = \sum_{t=1}^{n}\frac{1}{\sigma_t}\nu_t(\boldsymbol{u})\xi_t(\boldsymbol{u}).$$

By the Taylor expansion, we have $\nu_t(\boldsymbol{u}) = q_{1t}(\boldsymbol{u}) + q_{2t}(\boldsymbol{u})$, where

$$q_{1t}(\boldsymbol{u}) = \boldsymbol{u}'\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} \text{ and } q_{2t}(\boldsymbol{u}) = \frac{\boldsymbol{u}'}{2}\frac{\partial^2 q_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\boldsymbol{u}$$

with $\boldsymbol{\theta}^*$ between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_{\tau 0}$. Then it can be verified that

$$R_{1n}(\boldsymbol{u}) = -\sqrt{n}\boldsymbol{u}'\boldsymbol{T}_n - \sqrt{n}\boldsymbol{u}'K_{3n}(\boldsymbol{\theta}^*)\sqrt{n}\boldsymbol{u}, \tag{S1.46}$$

where

$$\boldsymbol{T}_n = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}}\psi_\tau(\varepsilon_{t,\tau}) \text{ and } K_{3n}(\boldsymbol{\theta}^*) = \frac{1}{2n}\sum_{t=1}^{n}\frac{1}{\sigma_t}\frac{\partial^2 q_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\psi_\tau(\varepsilon_{t,\tau}).$$

By the iterated-expectation and the fact that $E[\psi_\tau(\varepsilon_{t,\tau})] = 0$, it follows that

$$E\left[\frac{1}{\sigma_t}\frac{\partial^2 q_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\psi_\tau(\varepsilon_{t,\tau})\right] = 0.$$

Moreover, by Lemma 1 and the fact that $|\psi_\tau(\varepsilon_{t,\tau})| \leqslant 1$, we have

$$E\left[\sup_{\Theta}\left\|\frac{1}{\sigma_t}\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\psi_\tau(\varepsilon_{t,\tau})\right\|\right] \leqslant \left[E\sup_{\Theta}\frac{\sigma_t^2(\boldsymbol{\lambda})}{\sigma_t^2}\right]^{1/2}\left[E\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|^2\right]^{1/2} < \infty.$$

Then by the Theorem 3.1 in Ling and McAleer (2003), it follows that

$$\sup_{\Theta} \|K_{3n}(\boldsymbol{\theta}^*)\| = o_p(1).$$

This together with (S1.46), implies that

$$R_{1n}(\boldsymbol{u}) = -\sqrt{n}\boldsymbol{u}'\boldsymbol{T}_n + o_p(n\|\boldsymbol{u}\|^2). \tag{S1.47}$$

By simple calculation, we have $\xi_t(\boldsymbol{u}) = \xi_{1t}(\boldsymbol{u}) + \xi_{2t}(\boldsymbol{u})$, where

$$\begin{aligned}
\xi_{1t}(\boldsymbol{u}) &= \int_0^1 [I(\varepsilon_t \leqslant b_\tau + \sigma_t^{-1} q_{1t}(\boldsymbol{u})s) - I(\varepsilon_t \leqslant b_\tau)]ds \text{ and} \\
\xi_{2t}(\boldsymbol{u}) &= \int_0^1 [I(\varepsilon_t \leqslant b_\tau + \sigma_t^{-1} \nu_t(\boldsymbol{u})s) - I(\varepsilon_t \leqslant b_\tau + \sigma_t^{-1} q_{1t}(\boldsymbol{u})s)]ds.
\end{aligned}$$

Then for $R_{2n}(\boldsymbol{u})$, it can be verified that

$$R_{2n}(\boldsymbol{u}) = K_{4n}(\boldsymbol{u}) + K_{5n}(\boldsymbol{u}) + K_{6n}(\boldsymbol{u}) + K_{7n}(\boldsymbol{u}), \tag{S1.48}$$

where

$$\begin{aligned}
K_{4n}(\boldsymbol{u}) &= \boldsymbol{u}' \sum_{t=1}^{n} \frac{1}{\sigma_t} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}], \\
K_{5n}(\boldsymbol{u}) &= \boldsymbol{u}' \sum_{t=1}^{n} \frac{1}{\sigma_t} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} \{\xi_{1t}(\boldsymbol{u}) - E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}]\}, \\
K_{6n}(\boldsymbol{u}) &= \boldsymbol{u}' \sum_{t=1}^{n} \frac{1}{\sigma_t} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} \xi_{2t}(\boldsymbol{u}) \text{ and } K_{7n}(\boldsymbol{u}) = \frac{\boldsymbol{u}'}{2} \sum_{t=1}^{n} \frac{1}{\sigma_t} \frac{\partial^2 q_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \xi_t(\boldsymbol{u}) \boldsymbol{u}.
\end{aligned}$$

First, consider $K_{4n}(\boldsymbol{u})$. By the Taylor expansion, it follows that

$$\begin{aligned}
E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}] &= \int_0^1 [F_\varepsilon(b_\tau + \sigma_t^{-1} q_{1t}(\boldsymbol{u}) s) - F_\varepsilon(b_\tau)] ds \\
&= \frac{1}{2} f_\varepsilon(b_\tau) \sigma_t^{-1} q_{1t}(\boldsymbol{u}) + \sigma_t^{-1} q_{1t}(\boldsymbol{u}) \int_0^1 [f_\varepsilon(b_\tau + \sigma_t^{-1} q_{1t}(\boldsymbol{u}) s^*) - f_\varepsilon(b_\tau)] s ds,
\end{aligned}$$

where s^* is between 0 and s. Therefore, we have

$$K_{4n}(\boldsymbol{u}) = \sqrt{n}\boldsymbol{u}' J_n \sqrt{n}\boldsymbol{u} + \sqrt{n}\boldsymbol{u}' \Pi_{1n}(\boldsymbol{u}) \sqrt{n}\boldsymbol{u}, \tag{S1.49}$$

where

$$\begin{aligned}
J_n &= \frac{f_\varepsilon(b_\tau)}{2n} \sum_{t=1}^{n} \frac{1}{\sigma_t^2} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}'} \text{ and} \\
\Pi_{1n}(\boldsymbol{u}) &= \frac{1}{n} \sum_{t=1}^{n} \frac{1}{\sigma_t^2} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}} \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}'} \int_0^1 [f_\varepsilon(b_\tau + \sigma_t^{-1} q_{1t}(\boldsymbol{u}) s^*) - f_\varepsilon(b_\tau)] s ds.
\end{aligned}$$

By the iterated-expectation, the Taylor expansion, Lemma 1 and $\sup_{x \in \mathbb{R}} |\dot{f}(x)| < \infty$

by Assumption 4, for any $\eta > 0$, it holds that

$$\begin{aligned} E\left(\sup_{\|\boldsymbol{u}\|\leqslant\eta}\|\Pi_{1n}(\boldsymbol{u})\|\right) &\leqslant \frac{1}{n}\sum_{t=1}^{n} E\left\{\sup_{\|\boldsymbol{u}\|\leqslant\eta}\left\|\frac{1}{\sigma_t^2}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial\boldsymbol{\theta}}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial\boldsymbol{\theta}'}\sup_{x\in\mathbb{R}}|\dot{f}(x)|\frac{\boldsymbol{u}'}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial\boldsymbol{\theta}}\right\|\right\} \\ &\leqslant \eta\sup_{x\in\mathbb{R}}|\dot{f}(x)|E\left(\left\|\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial\boldsymbol{\theta}}\right\|^3\right) \end{aligned}$$

tends to 0 as $\eta \to 0$. Therefore, for any ϵ, $\delta > 0$, there exists $\eta_0 = \eta_0(\epsilon) > 0$ such that

$$\mathrm{P}\left(\sup_{\|\boldsymbol{u}\|\leqslant\eta_0}\|\Pi_{1n}(\boldsymbol{u})\| > \delta\right) < \frac{\epsilon}{2} \tag{S1.50}$$

for all $n \geqslant 1$. Since $\boldsymbol{u} = o_p(1)$, it follows that

$$\mathrm{P}\left(\|\boldsymbol{u}\| > \eta_0\right) < \frac{\epsilon}{2} \tag{S1.51}$$

as n is large enough. From (S1.50) and (S1.51), we have

$$\begin{aligned} \mathrm{P}\left(\|\Pi_{1n}(\boldsymbol{u})\| > \delta\right) &\leqslant \mathrm{P}\left(\|\Pi_{1n}(\boldsymbol{u})\| > \delta, \|\boldsymbol{u}\| \leqslant \eta_0\right) + \mathrm{P}\left(\|\boldsymbol{u}\| > \eta_0\right) \\ &\leqslant \mathrm{P}\left(\sup_{\|\boldsymbol{u}\|\leqslant\eta_0}\|\Pi_{1n}(\boldsymbol{u})\| > \delta\right) + \frac{\epsilon}{2} < \epsilon \end{aligned}$$

as n is large enough. Therefore, $\Pi_{1n}(\boldsymbol{u}) = o_p(1)$. This together with (S1.49), implies that

$$K_{4n}(\boldsymbol{u}) = \sqrt{n}\boldsymbol{u}' J_n \sqrt{n}\boldsymbol{u} + o_p(n\|\boldsymbol{u}\|^2). \tag{S1.52}$$

For $K_{5n}(\boldsymbol{u})$, by Lemma 3, it holds that

$$K_{5n}(\boldsymbol{u}) = o_p(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2). \tag{S1.53}$$

Next, we consider $K_{6n}(\boldsymbol{u})$. By the Taylor expansion, it follows that

$$\begin{aligned} E[\xi_{2t}(\boldsymbol{u})|\mathcal{F}_{t-1}] &= \int_0^1 [F_\varepsilon(b_\tau + \sigma_t^{-1}\nu(\boldsymbol{u})s) - F_\varepsilon(b_\tau + \sigma_t^{-1}q_{1t}(\boldsymbol{u})s)]ds \\ &= \sigma_t^{-1}q_{2t}(\boldsymbol{u})\int_0^1 f_\varepsilon(b^*)sds, \end{aligned}$$

where b^* is between $b_\tau + \sigma_t^{-1}q_{1t}(\boldsymbol{u})s$ and $b_\tau + \sigma_t^{-1}\nu(\boldsymbol{u})$. Then by the iterated-expectation and the Cauchy-Schiwarz inequality, together with Lemma 1 and $\sup_{x\in\mathbb{R}} f_\varepsilon(x) <$

∞ by Assumption 4, for any $\eta > 0$, it holds that

$$\begin{aligned} & E\left(\sup_{\|\boldsymbol{u}\|\leqslant\eta}\frac{|K_{6n}(\boldsymbol{u})|}{n\|\boldsymbol{u}\|^2}\right) \\ \leqslant\ & \frac{\eta}{n}\sum_{t=1}^{n}E\left\{\left\|\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial\boldsymbol{\theta}}\right\|\frac{1}{2}\sup_{x\in\mathbb{R}}f_{\varepsilon}(x)\sup_{\Theta}\left\|\frac{1}{\sigma_t}\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|\right\} \\ \leqslant\ & C\eta E\left\{\left\|\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial\boldsymbol{\theta}}\right\|\sup_{\Theta}\frac{\sigma_t(\boldsymbol{\lambda})}{\sigma_t}\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|\right\} \\ \leqslant\ & C\eta\left[E\left\|\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial\boldsymbol{\theta}}\right\|^4\right]^{1/4}\left[E\sup_{\Theta}\frac{\sigma_t^4(\boldsymbol{\lambda})}{\sigma_t^4}\right]^{1/4}\left[E\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|^2\right]^{1/2} \end{aligned} \quad \text{(S1.54)}$$

tends to 0 as $\eta \to 0$. Similar to (S1.50) and (S1.51), we can show that

$$K_{6n}(\boldsymbol{u}) = o_p(n\|\boldsymbol{u}\|^2). \quad \text{(S1.55)}$$

Finally, for $K_{7n}(\boldsymbol{u})$, it follows that

$$K_{7n}(\boldsymbol{u}) = \sqrt{n}\boldsymbol{u}'\Pi_{2n}(\boldsymbol{u})\sqrt{n}\boldsymbol{u} + \sqrt{n}\boldsymbol{u}'\Pi_{3n}(\boldsymbol{u})\sqrt{n}\boldsymbol{u}, \quad \text{(S1.56)}$$

where

$$\Pi_{2n}(\boldsymbol{u}) = \frac{1}{2n}\sum_{t=1}^{n}\frac{1}{\sigma_t}\frac{\partial^2 q_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\xi_{1t}(\boldsymbol{u}) \ \text{ and } \ \Pi_{3n}(\boldsymbol{u}) = \frac{1}{2n}\sum_{t=1}^{n}\frac{1}{\sigma_t}\frac{\partial^2 q_t(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\xi_{2t}(\boldsymbol{u}).$$

As for the proof in (S1.54), we can show that

$$\begin{aligned} & E\left(\sup_{\|\boldsymbol{u}\|\leqslant\eta}\|\Pi_{2n}(\boldsymbol{u})\|\right) \\ \leqslant\ & C\eta\left[E\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|^2\right]^{1/2}\left[E\sup_{\Theta}\frac{\sigma_t^4(\boldsymbol{\lambda})}{\sigma_t^4}\right]^{1/4}\left[E\left\|\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau0})}{\partial\boldsymbol{\theta}}\right\|^4\right]^{1/4} \end{aligned}$$

tends to 0 as $\eta\to 0$. Then similar to (S1.50) and (S1.51), we can prove that $\Pi_{2n}(\boldsymbol{u})=o_p(1)$. Similarly, for $\Pi_{3n}(\boldsymbol{u})$, by the Hölder inequality, for some $\delta>0$, we have

$$E\left(\sup_{\|\boldsymbol{u}\|\leqslant\eta}\|\Pi_{3n}(\boldsymbol{u})\|\right)\leqslant C\eta^2\left[E\sup_{\Theta}\left\|\frac{1}{\sigma_t(\boldsymbol{\lambda})}\frac{\partial^2 q_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right\|^{2+\delta}\right]^{\frac{2}{2+\delta}}\left\{E\sup_{\Theta}\left[\frac{\sigma_t(\boldsymbol{\lambda})}{\sigma_t}\right]^{\frac{2+\delta}{\delta}}\right\}^{\frac{\delta}{2+\delta}}$$

tends to 0 as $\eta\to 0$. This together with the proof similar to that of (S1.50) and (S1.51), implies that $\Pi_{3n}(\boldsymbol{u})=o_p(1)$. Therefore, by (S1.56), it follows that

$$K_{7n}(\boldsymbol{u})=o_p(n\|\boldsymbol{u}\|^2). \tag{S1.57}$$

Combing (S1.48), (S1.52), (S1.53), (S1.55) and (S1.57), we have

$$R_{2n}(\boldsymbol{u})=\sqrt{n}\boldsymbol{u}'J_n\sqrt{n}\boldsymbol{u}+o_p(\sqrt{n}\|\boldsymbol{u}\|+n\|\boldsymbol{u}\|^2). \tag{S1.58}$$

By (S1.45), (S1.47) and (S1.58), it follows that (S1.44) holds. In view of (S1.36) and (S1.44), the proof of this lemma is complete. □

Proof of Theorem 1. Denote $\ell_t(\boldsymbol{\theta}) = \rho_\tau[Y_t - q_t(\boldsymbol{\theta})]$ and $\widetilde{\ell}_t(\boldsymbol{\theta}) = \rho_\tau[Y_t - \widetilde{q}_t(\boldsymbol{\theta})]$, where $q_t(\boldsymbol{\theta}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b\sigma_t(\boldsymbol{\lambda})$ and $\widetilde{q}_t(\boldsymbol{\theta}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b\widetilde{\sigma}_t(\boldsymbol{\lambda})$. Define the following functions

$$\widehat{L}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\widehat{\sigma}_t^{-1}\widetilde{\ell}_t(\boldsymbol{\theta}), \quad \widetilde{L}_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\widetilde{\sigma}_t^{-1}\ell_t(\boldsymbol{\theta}) \text{ and } L_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n}\sigma_t^{-1}\ell_t(\boldsymbol{\theta}),$$

where $\widehat{\sigma}_t = \widetilde{\sigma}_t(\widehat{\boldsymbol{\lambda}}_n^{int})$, $\widetilde{\sigma}_t = \sigma_t(\widehat{\boldsymbol{\lambda}}_n^{int})$ and $\sigma_t = \sigma_t(\boldsymbol{\lambda}_0)$. To show the consistency of $\widehat{\boldsymbol{\theta}}_{\tau n}$, we first verify the following claims:

(i) $\sup_{\Theta} |\widehat{L}_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta})| = o_p(1)$;

(ii) $E[\sup_{\Theta} \sigma_t^{-1}\ell_t(\boldsymbol{\theta})] < \infty$;

(iii) $E[\sigma_t^{-1}\ell_t(\boldsymbol{\theta})]$ has a unique minimum at $\boldsymbol{\theta}_{\tau 0}$;

(iv) For any $\boldsymbol{\theta}^\dagger \in \Theta$, $E[\sup_{\boldsymbol{\theta}\in B_\eta(\boldsymbol{\theta}^\dagger)} \sigma_t^{-1}|\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}^\dagger)|] \to 0$ as $\eta \to 0$, where $B_\eta(\boldsymbol{\theta}^\dagger) = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta}^\dagger - \boldsymbol{\theta}\| < \eta\}$ is an open neighborhood of $\boldsymbol{\theta}^\dagger$ with radius $\eta > 0$.

To prove Claim (i), we need to verify that

$$\sup_{\boldsymbol{\theta}\in\Theta} |\widehat{L}_n(\boldsymbol{\theta}) - \widetilde{L}_n(\boldsymbol{\theta})| = o_p(1) \tag{S1.59}$$

and

$$\sup_{\boldsymbol{\theta}\in\Theta} |L_n(\boldsymbol{\theta}) - \widetilde{L}_n(\boldsymbol{\theta})| = o_p(1). \tag{S1.60}$$

We first show that (S1.60) holds. By Taylor expansion and the fact that $\sigma_t^2(\boldsymbol{\lambda}) \geqslant 1$,

$$|\widetilde{\sigma}_t^{-1} - \sigma_t^{-1}| = |\sigma_t^{-1}(\widehat{\boldsymbol{\lambda}}_n^{int}) - \sigma_t^{-1}(\boldsymbol{\lambda}_0)| \leqslant \frac{1}{2}\sup_{\Theta}\left\|\frac{1}{\sigma_t^2(\boldsymbol{\lambda})}\frac{\partial \sigma_t^2(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}}\right\| \|\widehat{\boldsymbol{\lambda}}_n^{int} - \boldsymbol{\lambda}_0\|.$$

Moreover, by the fact that $|\rho_\tau(x)| \leqslant |x|$, together with $E(Y_t^2) < \infty$ and $E\sup_\Theta q_t^2(\boldsymbol{\theta}) < \infty$ implied by Assumption 1 and $E(u_t^2) < \infty$, we have

$$E\sup_{\Theta}\{\rho_\tau[Y_t - q_t(\boldsymbol{\theta})]\}^2 \leqslant 2E(Y_t^2) + 2E\sup_{\Theta} q_t^2(\boldsymbol{\theta}) < \infty. \tag{S1.61}$$

This, together with Lemma 1, the ergodic theorem and $\widehat{\boldsymbol{\lambda}}_n^{int} - \boldsymbol{\lambda}_0 = o_p(1)$, leads to

$$\sup_{\Theta}|L_n(\boldsymbol{\theta}) - \widetilde{L}_n(\boldsymbol{\theta})| \leqslant \frac{1}{n}\sum_{t=1}^{n}|\widetilde{\sigma}_t^{-1} - \sigma_t^{-1}|\sup_{\Theta}\rho_\tau[Y_t - q_t(\boldsymbol{\theta})] = o_p(1).$$

Hence (S1.60) holds. We next verify (S1.59). It can be verified that

$$\begin{aligned}\widehat{L}_n(\boldsymbol{\theta}) - \widetilde{L}_n(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{t=1}^{n}\widehat{\sigma}_t^{-1}\rho_\tau[Y_t - \widetilde{q}_t(\boldsymbol{\theta})] - \frac{1}{n}\sum_{t=1}^{n}\widetilde{\sigma}_t^{-1}\rho_\tau[Y_t - q_t(\boldsymbol{\theta})] \\ &= R_{1n}(\boldsymbol{\theta}) + R_{2n}(\boldsymbol{\theta}),\end{aligned} \tag{S1.62}$$

where

$$\begin{aligned}R_{1n}(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{t=1}^{n}(\widehat{\sigma}_t^{-1} - \widetilde{\sigma}_t^{-1})\rho_\tau[Y_t - q_t(\boldsymbol{\theta})] \quad \text{and} \\ R_{2n}(\boldsymbol{\theta}) &= \frac{1}{n}\sum_{t=1}^{n}\widehat{\sigma}_t^{-1}\{\rho_\tau[Y_t - \widetilde{q}_t(\boldsymbol{\theta})] - \rho_\tau[Y_t - q_t(\boldsymbol{\theta})]\}.\end{aligned}$$

Since $\sigma_t^2(\boldsymbol{\lambda}) \geqslant 1$ and $\widetilde{\sigma}_t^2(\boldsymbol{\lambda}) \geqslant 1$, by Lemma 2(i), it follows that

$$|\widehat{\sigma}_t^{-1} - \widetilde{\sigma}_t^{-1}| = |\widehat{\sigma}_t^{-1}\widetilde{\sigma}_t^{-1}[\widetilde{\sigma}_t(\widehat{\boldsymbol{\lambda}}_n^{int}) - \sigma_t(\widehat{\boldsymbol{\lambda}}_n^{int})]| \leqslant C\rho^t\xi_\rho. \tag{S1.63}$$

Recall that $\xi_{\rho,t} = \sum_{j=0}^{\infty}\rho^j(1 + \|\boldsymbol{X}_{t-j-1}\| + \|\boldsymbol{V}_{t-j-1}\|^{1/2} + |u_{t-j}|)$ and $\xi_\rho = \sum_{j=0}^{\infty}\rho^j(1 + \|\boldsymbol{X}_{-j-1}\| + |u_{-j}|)$, where $\rho \in (0,1)$ is a constant. By Assumption 1, it is clear that $E(\xi_{\rho,t}^2) < \infty$ and $E(\xi_\rho^2) < \infty$. Then similar to the proof in (S1.61), by Lemma 1(i), we can show that $E\{\sup_\Theta[Y_t - q_t(\boldsymbol{\theta})]^2\} < \infty$. This together with (S1.63), implies that

$$\begin{aligned}\sup_\Theta |R_{1n}(\boldsymbol{\theta})| &\leqslant \frac{1}{n}\sum_{t=1}^{n}\sup_\Theta|\widehat{\sigma}_t^{-1} - \widetilde{\sigma}_t^{-1}|\sup_\Theta \rho_\tau[Y_t - q_t(\boldsymbol{\theta})] \\ &\leqslant \frac{C\xi_\rho}{n}\sum_{t=1}^{n}\rho^t\sup_\Theta\rho_\tau[Y_t - q_t(\boldsymbol{\theta})] = o_p(1). \end{aligned}\tag{S1.64}$$

Note that $\widetilde{q}_t(\boldsymbol{\theta}) - q_t(\boldsymbol{\theta}) = b[\widetilde{\sigma}_t(\boldsymbol{\lambda}) - \sigma_t(\boldsymbol{\lambda})]$ and $b < \bar{b}$ for $\boldsymbol{\theta} \in \Theta$. By Lemma 2(i), the Lipschitz continuity of $\rho_\tau(x)$ and the fact that $\widehat{\sigma}_t^2 \geqslant 1$, it follows that

$$\begin{aligned}\sup_\Theta |R_{2n}(\boldsymbol{\theta})| &\leqslant \frac{C}{n}\sum_{t=1}^{n}\widehat{\sigma}_t^{-1}\sup_\Theta|\widetilde{q}_t(\boldsymbol{\theta}) - q_t(\boldsymbol{\theta})| \\ &\leqslant \frac{C}{n}\sum_{t=1}^{n}\sup_\Theta|\widetilde{\sigma}_t(\boldsymbol{\lambda}) - \sigma_t(\boldsymbol{\lambda})| \leqslant \frac{C\xi_\rho}{n}\sum_{t=1}^{n}\rho^t = o_p(1). \end{aligned}\tag{S1.65}$$

From (S1.62), (S1.64) and (S1.65), we show that (S1.59) holds. Combing (S1.59) and (S1.60), Claim (i) is verified. Moreover, Claim (ii) is implied by (S1.61) and the fact that $\sigma_t^2 \geqslant 1$.

We next prove Claim (iii). For $x \neq 0$, it holds that

$$\begin{aligned}\rho_\tau(x-y)-\rho_\tau(x) &= -y\psi_\tau(x)+y\int_0^1[I(x\leqslant ys)-I(x\leqslant 0)]ds\\ &= -y\psi_\tau(x)+(x-y)[I(0>x>y)-I(0<x<y)], \end{aligned} \quad (S1.66)$$

where $\psi_\tau(x)=\tau-I(x<0)$; see Knight (1998). Denote $\nu_t(\boldsymbol{\theta})=q_t(\boldsymbol{\theta})-q_t(\boldsymbol{\theta}_{\tau 0})$ and $\varepsilon_{t,\tau}=\varepsilon_t-b_\tau$. Note that $\psi_\tau(\varepsilon_{t,\tau}\sigma_t)=\psi_\tau(\varepsilon_{t,\tau})$ and $E[\psi_\tau(\varepsilon_{t,\tau})]=0$. Then by (S1.66), we have

$$\begin{aligned} & E[\sigma_t^{-1}\ell_t(\boldsymbol{\theta})]-E[\sigma_t^{-1}\ell_t(\boldsymbol{\theta}_{\tau 0})]\\ = & E\left\{\sigma_t^{-1}\left[\rho_\tau(\varepsilon_{t,\tau}-\sigma_t^{-1}\nu_t(\boldsymbol{\theta}))-\rho_\tau(\varepsilon_{t,\tau})\right]\right\}\\ = & E\left\{\sigma_t^{-1}[\varepsilon_{t,\tau}-\sigma_t^{-1}\nu_t(\boldsymbol{\theta})][I(0>\varepsilon_{t,\tau}>\sigma_t^{-1}\nu_t(\boldsymbol{\theta}))-I(0<\varepsilon_{t,\tau}<\sigma_t^{-1}\nu_t(\boldsymbol{\theta}))]\right\}\geqslant 0, \end{aligned}$$

and, by Assumption 2, the equality holds if and only if $\nu_t(\boldsymbol{\theta})=q_t(\boldsymbol{\theta})-q_t(\boldsymbol{\theta}_{\tau 0})=0$ with probability one. Note that

$$\begin{aligned}\nu_t(\boldsymbol{\theta}) &= b_\tau\sqrt{1+\sum_{i=1}^q\alpha_{i0}u_{t-i}^2+\sum_{j=1}^p\beta_{j0}\sigma_{t-j}^2+\sum_{k=1}^d\pi_{k0}\upsilon_{k,t-1}^2}\\ &\quad -b\sqrt{1+\sum_{i=1}^q\alpha_i u_{t-i}^2(\boldsymbol{\phi})+\sum_{j=1}^p\beta_j\sigma_{t-j}^2(\boldsymbol{\lambda})+\sum_{k=1}^d\pi_k\upsilon_{k,t-1}^2}+(\boldsymbol{\phi}-\boldsymbol{\phi}_0)'\boldsymbol{X}_{t-1},\end{aligned}$$

where $u_{t-1}=\sigma_{t-1}\varepsilon_{t-1}$, $u_{t-1}(\boldsymbol{\phi})=u_{t-1}-(\boldsymbol{\phi}-\boldsymbol{\phi}_0)'\boldsymbol{X}_{t-2}$, σ_t^2 and $\sigma_t^2(\boldsymbol{\lambda})$ are $\mathcal{F}_{t-1}$-

measurable, and $\{\boldsymbol{X}_{t-1}\}$ are independent of $\{\varepsilon_t\}$. As a result, the random variable ε_{t-1} in $\nu_t(\boldsymbol{\theta})$ is independent of all the others, and then it holds that $\boldsymbol{\phi} = \boldsymbol{\phi}_0$ and $b_\tau^2\alpha_{10} = b^2\alpha_1$. Sequentially we can show that $b_\tau^2\alpha_{i0} = b^2\alpha_i$ for $i \geqslant 2$, $b_\tau = b$ and $\pi_{k0} = \pi_k$ for $k = 1, \cdots, d$. Finally, we verify that $\beta_{j0} = \beta_j$ for $j = 1, \cdots, p$. Thus, the proof of Claim (iii) is accomplished.

Finally, we assert Claim (iv). By the Taylor expansion, it holds that

$$|q_t(\boldsymbol{\theta}) - q_t(\boldsymbol{\theta}^\dagger)| \leqslant \|\boldsymbol{\theta} - \boldsymbol{\theta}^\dagger\| \left\| \frac{\partial q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|,$$

where $\boldsymbol{\theta}$ is between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^\dagger$. This together with the Lipschitz continuity of $\rho_\tau(x)$, the Cauchy-Schwarz inequality and Lemma 1, implies that

$$E[\sup_{\boldsymbol{\theta}\in B_\eta(\boldsymbol{\theta}^\dagger)} \sigma_t^{-1}|\ell_t(\boldsymbol{\theta}) - \ell_t(\boldsymbol{\theta}_\tau^\dagger)|] \leqslant C\eta \left[E \sup_\Theta \frac{\sigma_t^2(\boldsymbol{\lambda})}{\sigma_t^2}\right]^{1/2} \left[E \sup_\Theta \left\| \frac{1}{\sigma_t(\boldsymbol{\lambda})} \frac{\partial q_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^2\right]^{1/2}$$

tends to 0 as $\eta \to 0$. Hence, Claim (iv) holds.

Based on Claims (i)-(iv), by a method similar to that in Huber (1973), we next verify the consistency. Let V be any open neighborhood of $\boldsymbol{\theta}_{\tau 0} \in \Theta$. By Claim (iv), for any $\boldsymbol{\theta}^\dagger \in V^c = \Theta/V$ and $\epsilon > 0$, there exists an $\eta_0 > 0$ such that

$$E[\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}^\dagger)} \sigma_t^{-1}\ell_t(\boldsymbol{\theta})] \geqslant E[\sigma_t^{-1}\ell_t(\boldsymbol{\theta}^\dagger)] - \epsilon. \tag{S1.67}$$

From Claim (ii), by the ergodic theorem, it follows that

$$\frac{1}{n}\sum_{t=1}^{n}\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}^{\dagger})}\sigma_t^{-1}\ell_t(\boldsymbol{\theta}) \geqslant E[\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}^{\dagger})}\sigma_t^{-1}\ell_t(\boldsymbol{\theta})] - \epsilon \tag{S1.68}$$

as n is large enough. Since V^c is compact, we can choose $\{B_{\eta_0}(\boldsymbol{\theta}_i) : \boldsymbol{\theta}_i \in V^c, i = 1, \ldots, k\}$ to be a finite covering of V^c. Then by (S1.67) and (S1.68), as n is large enough, we have

$$\begin{aligned}
\inf_{\boldsymbol{\theta}\in V^c} L_n(\boldsymbol{\theta}) &= \min_{1\leqslant i\leqslant k}\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}_i)} L_n(\boldsymbol{\theta}) \\
&\geqslant \min_{1\leqslant i\leqslant k}\frac{1}{n}\sum_{t=1}^{n}\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}_i)}\sigma_t^{-1}\ell_t(\boldsymbol{\theta}) \\
&\geqslant \min_{1\leqslant i\leqslant k} E[\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}_i)}\sigma_t^{-1}\ell_t(\boldsymbol{\theta})] - \epsilon.
\end{aligned} \tag{S1.69}$$

Moreover, for each $\boldsymbol{\theta}_i \in V^c$, by Claim (iii), there exists an $\epsilon_0 > 0$ such that

$$E[\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}_i)}\sigma_t^{-1}\ell_t(\boldsymbol{\theta})] \geqslant E[\sigma_t^{-1}\ell_t(\boldsymbol{\theta}_{\tau 0})] + 3\epsilon_0. \tag{S1.70}$$

Therefore, by (S1.69) and (S1.70), taking $\epsilon = \epsilon_0$, it holds that

$$\inf_{\boldsymbol{\theta}\in V^c} L_n(\boldsymbol{\theta}) \geqslant E[\sigma_t^{-1}\ell_t(\boldsymbol{\theta}_{\tau 0})] + 2\epsilon_0. \tag{S1.71}$$

Furthermore, by the ergodic theorem, it follows that

$$\inf_{\boldsymbol{\theta}\in V} L_n(\boldsymbol{\theta}) \leqslant L_n(\boldsymbol{\theta}_{\tau 0}) = \frac{1}{n}\sum_{t=1}^{n} \sigma_t^{-1}\ell_t(\boldsymbol{\theta}_{\tau 0}) \leqslant E[\sigma_t^{-1}\ell_t(\boldsymbol{\theta}_{\tau 0})] + \epsilon_0. \tag{S1.72}$$

Combing (S1.71) and (S1.72), we have

$$\inf_{\boldsymbol{\theta}\in V^c} L_n(\boldsymbol{\theta}) \geqslant E[\sigma_t^{-1}\ell_t(\boldsymbol{\theta}_{\tau 0})] + 2\epsilon_0 > E[\sigma_t^{-1}\ell_t(\boldsymbol{\theta}_{\tau 0})] + \epsilon_0 \geqslant \inf_{\boldsymbol{\theta}\in V} L_n(\boldsymbol{\theta}), \tag{S1.73}$$

which together with Claim (i), implies that

$$\widehat{\boldsymbol{\theta}}_{\tau n} \in V \quad \text{in probability for } \forall V, \text{ as } n \text{ is large enough.}$$

By the arbitrariness of V, it implies that $\widehat{\boldsymbol{\theta}}_{\tau n} \to \boldsymbol{\theta}_{\tau 0}$ in probability. The proof of this theorem is complete. □

Proof of Theorem 2. Denote $\widehat{\boldsymbol{u}}_n = \widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}$. From Theorem 1, we have $\widehat{\boldsymbol{u}}_n = o_p(1)$. Since $\widehat{\boldsymbol{\theta}}_{\tau n}$ minimizes $\widehat{L}_n(\boldsymbol{\theta})$, then $\widehat{\boldsymbol{u}}_n$ is the minimizer of $\widehat{H}_n(\boldsymbol{u}) = n[\widehat{L}_n(\boldsymbol{\theta}_{\tau 0} + \boldsymbol{u}) - \widehat{L}_n(\boldsymbol{\theta}_{\tau 0})]$. Define $J = f_\varepsilon(b_\tau)\Sigma(\tau)/2$. By Lemma 1 and the ergodic theorem, we have $J_n = J + o_p(1)$. Moreover, by Lemmas 4 and 5, it follows that

$$\begin{aligned} \widehat{H}_n(\widehat{\boldsymbol{u}}_n) &= -\sqrt{n}\widehat{\boldsymbol{u}}_n'\boldsymbol{T}_n + \sqrt{n}\widehat{\boldsymbol{u}}_n' J\sqrt{n}\widehat{\boldsymbol{u}}_n + o_p(\sqrt{n}\|\widehat{\boldsymbol{u}}_n\| + n\|\widehat{\boldsymbol{u}}_n\|^2) \\ &\geqslant -\sqrt{n}\|\widehat{\boldsymbol{u}}_n\|[\|\boldsymbol{T}_n\| + o_p(1)] + n\|\widehat{\boldsymbol{u}}_n\|^2[\lambda_{\min} + o_p(1)], \end{aligned} \tag{S1.74}$$

where $\lambda_{\min}$ is the smallest eigenvalue of J. Note that, as $n \to \infty$, $\boldsymbol{T}_n$ converges in distribution to a normal random variable with mean zero and variance matrix $\tau(1-\tau)\Sigma(\tau)$.

Since $\widehat{H}_n(\widehat{\boldsymbol{u}}_n) \leqslant 0$, by (S1.74), it holds that

$$\sqrt{n}\|\widehat{\boldsymbol{u}}_n\| \leqslant [\lambda_{\min} + o_p(1)]^{-1}[\|\boldsymbol{T}_n\| + o_p(1)] = O_p(1). \tag{S1.75}$$

This together with Theorem 1, verifies the root-n consistency of $\widehat{\boldsymbol{\theta}}_{\tau n}$ in probability.

Let $\sqrt{n}\boldsymbol{u}_n = J^{-1}\boldsymbol{T}_n/2 = f_\varepsilon^{-1}(b_\tau)\Sigma^{-1}(\tau)\boldsymbol{T}_n$, then we have

$$\sqrt{n}\boldsymbol{u}_n \to N\left(\boldsymbol{0}, \frac{\tau(1-\tau)}{f_\varepsilon^2(b_\tau)}\Sigma^{-1}(\tau)\right)$$

in distribution as $n \to \infty$. Therefore, it suffices to show that $\sqrt{n}\boldsymbol{u}_n - \sqrt{n}\widehat{\boldsymbol{u}}_n = o_p(1)$. By (S1.74) and (S1.75), we have

$$\begin{aligned}\widehat{H}_n(\widehat{\boldsymbol{u}}_n) &= -\sqrt{n}\widehat{\boldsymbol{u}}_n'\boldsymbol{T}_n + \sqrt{n}\widehat{\boldsymbol{u}}_n'J\sqrt{n}\widehat{\boldsymbol{u}}_n + o_p(1) \\ &= -2\sqrt{n}\widehat{\boldsymbol{u}}_n'J\sqrt{n}\boldsymbol{u}_n + \sqrt{n}\widehat{\boldsymbol{u}}_n'J\sqrt{n}\widehat{\boldsymbol{u}}_n + o_p(1) \text{ and}\end{aligned} \tag{S1.76}$$

$$\widehat{H}_n(\boldsymbol{u}_n) = -\sqrt{n}\boldsymbol{u}_n'\boldsymbol{T}_n + \sqrt{n}\boldsymbol{u}_n'J\sqrt{n}\boldsymbol{u}_n + o_p(1) = -\sqrt{n}\boldsymbol{u}_n'J\sqrt{n}\boldsymbol{u}_n + o_p(1). \tag{S1.77}$$

From (S1.76) and (S1.77), it follows that

$$\begin{aligned}\widehat{H}_n(\widehat{\boldsymbol{u}}_n) - \widehat{H}_n(\boldsymbol{u}_n) &= (\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n)' J(\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n) + o_p(1) \\ &\geqslant \lambda_{\min}\|\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n\|^2 + o_p(1). \qquad \text{(S1.78)}\end{aligned}$$

Since $\widehat{H}_n(\widehat{\boldsymbol{u}}_n) - \widehat{H}_n(\boldsymbol{u}_n) = n[\widehat{L}_n(\boldsymbol{\theta}_{\tau 0}+\widehat{\boldsymbol{u}}_n) - \widehat{L}_n(\boldsymbol{\theta}_{\tau 0}+\boldsymbol{u}_n)] \leqslant 0$ a.s., then (S1.78) implies that $\|\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n\| = o_p(1)$. The proof of this theorem is hence accomplished. □

Proof of Corollary 1. First, we show the consistency of $\widetilde{\boldsymbol{\theta}}_{\tau n}$. The proof follows the same lines as that of Theorem 1, while functions $L(\boldsymbol{\theta})$, $\widetilde{L}_n(\boldsymbol{\theta})$ and $\widehat{L}_n(\boldsymbol{\theta})$ are defined as $L_n(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^n \rho_\tau[Y_t - q_t(\boldsymbol{\theta})]$, $\widetilde{L}_n(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^n \rho_\tau[Y_t - q_t(\boldsymbol{\theta})]$ and $\widehat{L}_n(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^n \rho_\tau[Y_t - \widetilde{q}_t(\boldsymbol{\theta})]$, respectively. Next, we show the root-n consistency and asymptotic normality of $\widetilde{\boldsymbol{\theta}}_{\tau n}$ as in Theorem 2, where functions $L(\boldsymbol{\theta})$, $\widetilde{L}_n(\boldsymbol{\theta})$ and $\widehat{L}_n(\boldsymbol{\theta})$ are defined as previous, the function $\zeta_n(\boldsymbol{u})$ in Lemma 3 is defined as $\zeta_n(\boldsymbol{u}) = \sum_{t=1}^n q_{1t}(\boldsymbol{u})\left\{\xi_{1t}(\boldsymbol{u}) - E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}]\right\}$, Lemma 4 remains unchanged while Lemma 5 is consequently revised as below.

Lemma 5'. Suppose $E(|u_t|^{2+\delta}) < \infty$ for some $\delta > 0$. By Assumptions 1, 3 and 4, we have

$$\begin{aligned}n[\widetilde{L}_n(\boldsymbol{\theta}) - \widetilde{L}_n(\boldsymbol{\theta}_{\tau 0})] &= -\sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0})'\boldsymbol{T}_n + \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0})' J_n \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}) \\ &\quad + o_p(\sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}\| + n\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}\|^2)\end{aligned}$$

for $\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0} = o_p(1)$, where $\widetilde{L}_n(\boldsymbol{\theta}) = n^{-1}\sum_{t=1}^n \rho_\tau[Y_t - \widetilde{q}_t(\boldsymbol{\theta})]$,

$$\boldsymbol{T}_n = \frac{1}{\sqrt{n}}\sum_{t=1}^n \frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}}\psi_\tau(\varepsilon_t - b_\tau) \text{ and } J_n = \frac{f_\varepsilon(b_\tau)}{2n}\sum_{t=1}^n \frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}'}.$$

Note that without weights σ_t^{-1}, additional moment condition on u_t will be needed in some intermediate steps of the proof. Therefore, instead of $E|u_t|^2 < \infty$, higher moment condition, $E|u_t|^{2+\delta} < \infty$ for some $\delta > 0$, is required for the proof of Corollary 1. □

Proof of Corollary 2. Denote $\underline{\boldsymbol{z}}_t(\boldsymbol{\gamma}) = (u_{t-1}^2, \ldots, u_{t-q}^2, \underline{\sigma}_{t-1}^2(\boldsymbol{\gamma}), \ldots, \underline{\sigma}_{t-p}^2(\boldsymbol{\gamma}), \upsilon_{1,t-1}^2, \ldots, \upsilon_{d,t-1}^2)'$ and $\underline{\check{\boldsymbol{z}}}_t(\boldsymbol{\gamma}) = (\check{u}_{t-1}^2, \ldots, \check{u}_{t-q}^2, \underline{\check{\sigma}}_{t-1}^2(\boldsymbol{\gamma}), \ldots, \underline{\check{\sigma}}_{t-p}^2(\boldsymbol{\gamma}), \upsilon_{1,t-1}^2, \ldots, \upsilon_{d,t-1}^2)'$, where $u_t = u_t(\boldsymbol{\phi}_0)$, $\check{u}_t = u_t(\check{\boldsymbol{\phi}}_n)$, $\underline{\sigma}_t^2(\boldsymbol{\gamma}) = 1 + \sum_{i=1}^q \alpha_i u_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2(\boldsymbol{\gamma}) + \sum_{k=1}^d \pi_k \upsilon_{k,t-1}^2$ and $\underline{\check{\sigma}}_t^2(\boldsymbol{\gamma}) = 1 + \sum_{i=1}^q \alpha_i \check{u}_{t-i}^2 + \sum_{j=1}^p \beta_j \check{\sigma}_{t-j}^2(\boldsymbol{\gamma}) + \sum_{k=1}^d \pi_k \upsilon_{k,t-1}^2$. Let $\underline{\sigma}_t^2(\boldsymbol{\gamma}) = 1 + \boldsymbol{\gamma}'\underline{\boldsymbol{z}}_t(\boldsymbol{\gamma})$ and $\underline{\check{\sigma}}_t^2(\boldsymbol{\gamma}) = 1 + \boldsymbol{\gamma}'\underline{\check{\boldsymbol{z}}}_t(\boldsymbol{\gamma})$. Note that $u_t = u_t(\boldsymbol{\phi}_0) = Y_t - \boldsymbol{\phi}_0'\boldsymbol{X}_{t-1}$ and $\check{u}_t = u_t(\check{\boldsymbol{\phi}}_n) = Y_t - \check{\boldsymbol{\phi}}_n'\boldsymbol{X}_{t-1}$. Let $\boldsymbol{\gamma}_\tau = (b, \boldsymbol{\gamma}')'$ and denote by $\boldsymbol{\gamma}_{\tau 0} = (b_\tau, \boldsymbol{\gamma}_0')'$ its true value. Denote by $\Theta' \subset \mathbb{R}^{p+q+d+1}$ the parameter space of $\boldsymbol{\gamma}_\tau$, which satisfies

$$\begin{aligned} \underline{b} \leqslant |b| \leqslant \overline{b}, \sum_{j=1}^p \beta_j \leqslant \rho_0,\ \underline{w} &\leqslant \min(\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p, \pi_1, \ldots, \pi_d) \\ &\leqslant \max(\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_p, \pi_1, \ldots, \pi_d) \leqslant \overline{w}, \end{aligned}$$

where $0 < \underline{b} < \overline{b}$, $0 < \underline{w} < \overline{w}$, $0 < \rho_0 < 1$ and $p\underline{w} < \rho_0$. Moreover, we assume Θ' is compact and $\boldsymbol{\gamma}_{\tau 0}$ is an interior of Θ'. Recall that $\check{\boldsymbol{\theta}}_{\tau n} = (\check{\boldsymbol{\gamma}}_{\tau n}', \check{\boldsymbol{\phi}}_n')'$, where

$\check{\boldsymbol{\gamma}}_{\tau n} = (\check{b}_{\tau n}, \check{\boldsymbol{\gamma}}_n')'$. For the least square estimator $\check{\boldsymbol{\phi}}_n$, by the model assumption in (2.1) and (2.2), we have

$$\sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{X}_{t-1}\sigma_t\varepsilon_t. \tag{S1.79}$$

Then $\sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = O_p(1)$ and $\sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) \to N(\mathbf{0}, \Sigma_{22})$ in distribution as $n \to \infty$, where $\Sigma_{22} = \omega^* D_0^{-1} D_2 D_0^{-1}$ with $\omega^* = \text{var}(\varepsilon_t)$ and $D_i = E(\sigma_t^i \boldsymbol{X}_{t-1}\boldsymbol{X}_{t-1}')$.

In the following proof, we focus on $\check{\boldsymbol{\gamma}}_{\tau n}$. First, we verify its consistency. Define

$$\check{L}_n(\boldsymbol{\gamma}_\tau) = \frac{1}{n}\sum_{t=1}^{n} \check{\ell}_t(\boldsymbol{\gamma}_\tau) \quad \text{and} \quad L_n(\boldsymbol{\gamma}_\tau) = \frac{1}{n}\sum_{t=1}^{n} \ell_t(\boldsymbol{\gamma}_\tau),$$

where $\check{\ell}_t(\boldsymbol{\gamma}_\tau) = \rho_\tau[Y_t - \check{\boldsymbol{\phi}}_n' \boldsymbol{X}_{t-1} - b\check{\underline{\sigma}}_t(\boldsymbol{\gamma})]$ and $\ell_t(\boldsymbol{\gamma}_\tau) = \rho_\tau[Y_t - \boldsymbol{\phi}_0' \boldsymbol{X}_{t-1} - b\underline{\sigma}_t(\boldsymbol{\gamma})]$. To show the consistency, we first verify the following claims:

(i) $\sup_{\Theta'} |\check{L}_n(\boldsymbol{\gamma}_\tau) - L_n(\boldsymbol{\gamma}_\tau)| = o_p(1)$;

(ii) $E[\sup_{\Theta'} \ell_t(\boldsymbol{\gamma}_\tau)] < \infty$;

(iii) $E[\ell_t(\boldsymbol{\gamma}_\tau)]$ has a unique minimum at $\boldsymbol{\gamma}_{\tau 0}$;

(iv) For any $\boldsymbol{\gamma}^\dagger \in \Theta'$, $E[\sup_{\boldsymbol{\gamma}_\tau \in B_\eta(\boldsymbol{\gamma}^\dagger)} |\ell_t(\boldsymbol{\gamma}_\tau) - \ell_t(\boldsymbol{\gamma}^\dagger)|] \to 0$ as $\eta \to 0$, where $B_\eta(\boldsymbol{\gamma}^\dagger) = \{\boldsymbol{\gamma}_\tau \in \Theta' : \|\boldsymbol{\gamma}^\dagger - \boldsymbol{\gamma}_\tau\| < \eta\}$ is an open neighborhood of $\boldsymbol{\gamma}^\dagger$ with radius $\eta > 0$.

We first prove Claim (i). Note that $\check{\underline{\sigma}}_t(\boldsymbol{\gamma}) - \underline{\sigma}_t(\boldsymbol{\gamma}) = [\check{\underline{\sigma}}_t^2(\boldsymbol{\gamma}) - \underline{\sigma}_t^2(\boldsymbol{\gamma})]/[\check{\underline{\sigma}}_t(\boldsymbol{\gamma}) + \underline{\sigma}_t(\boldsymbol{\gamma})]$.

Similar to the proof of Lemma 2(i), by the Taylor expansion, it can be verified that

$$\underline{\breve{\sigma}}_t^2(\boldsymbol{\gamma}) - \underline{\sigma}_t^2(\boldsymbol{\gamma}) = -2(\breve{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)' \sum_{i=1}^{\infty} a_\gamma(i) u_{t-i}(\boldsymbol{\phi}^*) \boldsymbol{X}_{t-i-1},$$

where $\boldsymbol{\phi}^*$ is between $\boldsymbol{\phi}_0$ and $\breve{\boldsymbol{\phi}}_n$. By Assumption 1, together with the facts that $\underline{\breve{\sigma}}_t^2(\boldsymbol{\gamma}) \geqslant a_\gamma(i)\breve{u}_{t-i}^2$ and $\underline{\sigma}_t^2(\boldsymbol{\gamma}) \geqslant a_\gamma(i)u_{t-i}^2$, we can show that

$$\sup_{\Theta'} \|\underline{\breve{\sigma}}_t(\boldsymbol{\gamma}) - \underline{\sigma}_t(\boldsymbol{\gamma})\| \leqslant 2\|\breve{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0\| \sum_{i=1}^{\infty} \frac{a_\gamma(i)|u_{t-i}(\boldsymbol{\phi}^*)|\|\boldsymbol{X}_{t-i-1}\|}{\underline{\breve{\sigma}}_t(\boldsymbol{\gamma}) + \underline{\sigma}_t(\boldsymbol{\gamma})} \leqslant 2\|\breve{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0\|\rho^t \xi_\rho.$$

Then by the Lipschitz continuity of $\rho_\tau(x)$ and Assumption 1, we have

$$\begin{aligned}\sup_{\Theta'} |\breve{L}_n(\boldsymbol{\gamma}_\tau) - L_n(\boldsymbol{\gamma}_\tau)| &\leqslant \frac{2}{n}\sum_{t=1}^{n} \sup_{\Theta'} |(\breve{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)'\boldsymbol{X}_{t-1} + b[\breve{\sigma}_t(\boldsymbol{\alpha}) - \sigma_t(\boldsymbol{\alpha})]| \\ &\leqslant 2\|\breve{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0\| \frac{1}{n}\sum_{t=1}^{n} [\|\boldsymbol{X}_{t-1}\| + C\bar{b}\rho^t\xi_\rho].\end{aligned}$$

This together with $\sqrt{n}(\breve{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = O_p(1)$, implies that Claim (i) holds. Similar to proofs of Claims (ii)-(iv) in the proof of Theorem 1, we can verify Claims (ii)-(iv). Finally, by Claims (i)-(iv) and the similar arguments as in the proof of Theorem 1, we can show that $\breve{\boldsymbol{\gamma}}_{\tau n} - \boldsymbol{\gamma}_{\tau 0} = o_p(1)$.

We next prove the asymptotic normality of $\breve{\boldsymbol{\gamma}}_{\tau n}$. We use the same technical tools as in the proof of Theorem 2. We only sketch the key steps as below. Denote $\boldsymbol{u} = \boldsymbol{\gamma}_\tau - \boldsymbol{\gamma}_{\tau 0}$. Define $q_t(\boldsymbol{\gamma}_{\tau 0}) = \boldsymbol{\phi}_0'\boldsymbol{X}_{t-1} + b_\tau\underline{\sigma}_t(\boldsymbol{\gamma}_0)$ and $\breve{q}_t(\boldsymbol{\gamma}_\tau) = \breve{\boldsymbol{\phi}}_n'\boldsymbol{X}_{t-1} + b\underline{\breve{\sigma}}_t(\boldsymbol{\gamma})$. Denote

$\nu_t(\boldsymbol{u}) = q_t(\boldsymbol{\gamma}_\tau) - \check{q}_t(\boldsymbol{\gamma}_{\tau 0})$. This together with $\check{u}_t = Y_t - \check{\boldsymbol{\phi}}_n' \boldsymbol{X}_{t-1}$ and $u_t = Y_t - \boldsymbol{\phi}_0' \boldsymbol{X}_{t-1}$, implies that $\nu_t(\boldsymbol{u}) = -(\check{u}_t - u_t) + [b \check{\underline{\sigma}}_t(\boldsymbol{\gamma}) - b_\tau \underline{\sigma}_t(\boldsymbol{\gamma}_0)]$. Note that $\sigma_t = \underline{\sigma}_t(\boldsymbol{\gamma}_0)$ and $u_t = \sigma_t \varepsilon_t$, then $u_t - b_\tau \underline{\sigma}_t(\boldsymbol{\gamma}_0) = \sigma_t(\varepsilon_t - b_\tau)$ and hence $\psi_\tau(u_t - b_\tau \underline{\sigma}_t(\boldsymbol{\gamma}_0)) = \psi_\tau(\varepsilon_t - b_\tau)$. This together with the Knight equation (S1.66), we can show that

$$n[\check{L}_n(\boldsymbol{\gamma}_\tau) - L_n(\boldsymbol{\gamma}_{\tau 0})] = R_{1n}(\boldsymbol{u}) + R_{2n}(\boldsymbol{u}), \tag{S1.80}$$

where $R_{1n}(\boldsymbol{u}) = -\sum_{t=1}^n \nu_t(\boldsymbol{u}) \psi_\tau(\varepsilon_t - b_\tau)$ and

$$R_{2n}(\boldsymbol{u}) = \sum_{t=1}^n \nu_t(\boldsymbol{u}) \int_0^1 [I(\varepsilon_t \leqslant b_\tau + \sigma_t^{-1} \nu_t(\boldsymbol{u}) s) - I(\varepsilon_t \leqslant b_\tau)] ds.$$

Since $\check{u}_t - u_t = -(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)' \boldsymbol{X}_{t-1}$, it can be verified that $\nu_t(\boldsymbol{u}) = q_{1t}(\boldsymbol{u}) + q_{2t}(\boldsymbol{u})$, where

$$\begin{aligned}
q_{1t}(\boldsymbol{u}) &= (\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)' \boldsymbol{X}_{t-1} + b_\tau [\check{\underline{\sigma}}_t(\boldsymbol{\alpha}_0) - \underline{\sigma}_t(\boldsymbol{\gamma}_0)] + (b - b_\tau)\sigma_t + b_\tau [\underline{\sigma}_t(\boldsymbol{\gamma}) - \underline{\sigma}_t(\boldsymbol{\gamma}_0)], \\
q_{2t}(\boldsymbol{u}) &= b_\tau \{\check{\underline{\sigma}}_t(\boldsymbol{\gamma}) - \underline{\sigma}_t(\boldsymbol{\gamma}) - [\check{\underline{\sigma}}_t(\boldsymbol{\gamma}_0) - \underline{\sigma}_t(\boldsymbol{\gamma}_0)]\} + (b - b_\tau)[\underline{\sigma}_t(\boldsymbol{\gamma}) - \underline{\sigma}_t(\boldsymbol{\gamma}_0)] \\
&\quad + (b - b_\tau)\{\check{\underline{\sigma}}_t(\boldsymbol{\gamma}) - \underline{\sigma}_t(\boldsymbol{\gamma}) - [\check{\underline{\sigma}}_t(\boldsymbol{\gamma}_0) - \underline{\sigma}_t(\boldsymbol{\gamma}_0)]\} + (b - b_\tau)[\check{\underline{\sigma}}_t(\boldsymbol{\gamma}_0) - \underline{\sigma}_t(\boldsymbol{\gamma}_0)].
\end{aligned}$$

Recall that $\boldsymbol{M}_t = \boldsymbol{X}_{t-1} + 0.5 \sigma_t^{-1} b_\tau \partial \sigma_t^2(\boldsymbol{\lambda}_0)/\partial \boldsymbol{\phi}'$ and $\boldsymbol{W}_t = (\sigma_t, 0.5 \sigma_t^{-1} b_\tau \partial \sigma_t^2(\boldsymbol{\lambda}_0)/\partial \boldsymbol{\gamma}')'$. Following the lines in the proof of Theorem 2, if Assumptions 1, 3 and 4 hold and

$E|u_t|^{2+\delta} < \infty$ for some $\delta > 0$, we can show that

$$R_{1n}(\boldsymbol{u}) = -\sqrt{n}\boldsymbol{u}'\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{W}_t\psi_\tau(\varepsilon_t - b_\tau) - T_{2n} + o_p(n\|\boldsymbol{u}\|^2), \quad \text{with} \qquad \text{(S1.81)}$$

$$T_{2n} = \sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)'\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{M}_t\psi_\tau(\varepsilon_t - b_\tau).$$

For $R_{2n}(\boldsymbol{u})$, we have

$$\begin{aligned} R_{2n}(\boldsymbol{u}) &= \sqrt{n}\boldsymbol{u}'\frac{f_\varepsilon(b_\tau)}{2}\frac{1}{n}\sum_{t=1}^{n}\sigma_t^{-1}\boldsymbol{W}_t\boldsymbol{W}_t'\sqrt{n}\boldsymbol{u} \\ &+\sqrt{n}\boldsymbol{u}'f_\varepsilon(b_\tau)\frac{1}{n}\sum_{t=1}^{n}\sigma_t^{-1}\boldsymbol{W}_t\boldsymbol{M}_t'\sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) \\ &+T_{3n} + o_p(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2), \quad \text{with} \end{aligned} \qquad \text{(S1.82)}$$

$$T_{3n} = \sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)'\frac{f_\varepsilon(b_\tau)}{2}\frac{1}{n}\sum_{t=1}^{n}\sigma_t^{-1}\boldsymbol{M}_t\boldsymbol{M}_t'\sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0).$$

Hence, by (S1.80)-(S1.82), we have

$$\begin{aligned} & n[\check{L}_n(\boldsymbol{\gamma}_\tau) - L_n(\boldsymbol{\gamma}_{\tau 0})] \\ = \quad & -\sqrt{n}\boldsymbol{u}'\left[\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{W}_t\psi_\tau(\varepsilon_{t,\tau}) - f_\varepsilon(b_\tau)\frac{1}{n}\sum_{t=1}^{n}\sigma_t^{-1}\boldsymbol{W}_t\boldsymbol{M}_t'\sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)\right] \\ & +\sqrt{n}\boldsymbol{u}'\frac{f_\varepsilon(b_\tau)}{2}\frac{1}{n}\sum_{t=1}^{n}\sigma_t^{-1}\boldsymbol{W}_t\boldsymbol{W}_t'\sqrt{n}\boldsymbol{u} - T_{2n} + T_{3n} + o_p(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2). \end{aligned}$$

By the consistency, we have $\check{\boldsymbol{\gamma}}_{\tau n} - \boldsymbol{\gamma}_{\tau 0} = o_p(1)$. Moreover, $\check{\boldsymbol{\gamma}}_{\tau n}$ is the minimizer of

$\check{L}_n(\boldsymbol{\gamma}_\tau)$. Then it follows that

$$\sqrt{n}(\check{\boldsymbol{\gamma}}_{\tau n}-\boldsymbol{\gamma}_{\tau 0})=\frac{\Omega_1^{-1}}{f_\varepsilon(b_\tau)}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{W}_t\psi_\tau(\varepsilon_t-b_\tau)-\Omega_1^{-1}\Gamma_1\sqrt{n}(\check{\boldsymbol{\phi}}_n-\boldsymbol{\phi}_0)+o_p(1),$$

where $\Omega_1=E(\sigma_t^{-1}\boldsymbol{W}_t\boldsymbol{W}_t')$ and $\Gamma_1=E(\sigma_t^{-1}\boldsymbol{W}_t\boldsymbol{M}_t')$. Similar to the proof of Theorem 2, we can verify that $\sqrt{n}(\check{\boldsymbol{\gamma}}_{\tau n}-\boldsymbol{\gamma}_{\tau 0})=O_p(1)$ and $\sqrt{n}(\check{\boldsymbol{\gamma}}_{\tau n}-\boldsymbol{\gamma}_{\tau 0})\to N(\mathbf{0},\Sigma_{11}(\tau))$ in distribution as $n\to\infty$, where $\Sigma_{11}(\tau)$ is defined in Section 2.2. This together with (S1.79), we complete the proof by the central limit theorem and the Cramér-Wold device. $\square$

Proof of Corollary 3. First, consider the conditional quantile $q_{n+1}(\widehat{\boldsymbol{\theta}}_{\tau n})$ for the jointly weighted estimator $\widehat{\boldsymbol{\theta}}_{\tau n}=(\widehat{\boldsymbol{\gamma}}_{\tau n}',\widehat{\boldsymbol{\phi}}_n')'$, where $\widehat{\boldsymbol{\gamma}}_{\tau n}=(\widehat{b}_{\tau n},\widehat{\boldsymbol{\gamma}}_n')'$. By Theorem 2, we have $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n}-\boldsymbol{\theta}_{\tau 0})=O_p(1)$. Then by the Taylor expansion and Lemma 1, given $\mathcal{F}_n$, it follows that

$$\begin{aligned}q_{n+1}(\widehat{\boldsymbol{\theta}}_{\tau n})-q_{n+1}(\boldsymbol{\theta}_{\tau 0}) &= \frac{\partial q_{n+1}(\boldsymbol{\theta}_{\tau 0})}{\partial\boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}_{\tau n}-\boldsymbol{\theta}_{\tau 0})+\frac{1}{2}(\widehat{\boldsymbol{\theta}}_{\tau n}-\boldsymbol{\theta}_{\tau 0})'\frac{\partial^2 q_{n+1}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}(\widehat{\boldsymbol{\theta}}_{\tau n}-\boldsymbol{\theta}_{\tau 0})\\ &= \boldsymbol{W}_{n+1}'(\widehat{\boldsymbol{\gamma}}_{\tau n}-\boldsymbol{\gamma}_{\tau 0})+\boldsymbol{M}_{n+1}'(\widehat{\boldsymbol{\phi}}_n-\boldsymbol{\phi}_0)+o_p(n^{-1/2}).\end{aligned}$$

Similarly, by Corollary 1 and the Taylor expansion, we can verify the representation of $q_{n+1}(\widetilde{\boldsymbol{\theta}}_{\tau n})$ for the jointly weighted estimator $\widetilde{\boldsymbol{\theta}}_{\tau n}=(\widetilde{\boldsymbol{\gamma}}_{\tau n}',\widetilde{\boldsymbol{\phi}}_n')'$, where $\widetilde{\boldsymbol{\gamma}}_{\tau n}=(\widetilde{b}_{\tau n},\widetilde{\boldsymbol{\gamma}}_n')'$.

Finally, consider $q_{n+1}(\check{\boldsymbol{\theta}}_{\tau n})$ for the two-step estimator $\check{\boldsymbol{\theta}}_{\tau n}=(\check{\boldsymbol{\gamma}}_{\tau n}',\check{\boldsymbol{\phi}}_n')'$, where $\check{\boldsymbol{\gamma}}_{\tau n}=(\check{b}_{\tau n},\check{\boldsymbol{\gamma}}_n')'$. Recall that $\underline{\sigma}_t^2(\boldsymbol{\gamma})=1+\boldsymbol{\gamma}'\underline{\boldsymbol{z}}_t(\boldsymbol{\gamma})$ and $\underline{\check{\sigma}}_t^2(\boldsymbol{\gamma})=1+\boldsymbol{\gamma}'\underline{\check{\boldsymbol{z}}}_t(\boldsymbol{\gamma})$ in the proof of Corollary 2. Note that $\sigma_t = \underline{\sigma}_t(\boldsymbol{\gamma}_0) = \sigma_t(\boldsymbol{\lambda}_0)$. By Corollary 2, we have $\sqrt{n}(\check{\boldsymbol{\gamma}}_{\tau n} - \boldsymbol{\gamma}_{\tau 0}) = O_p(1)$ and $\sqrt{n}(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) = O_p(1)$. Then by the Taylor expansion, Lemma 1 and the proof of Corollary 2, given $\mathcal{F}_n$, it holds that

$$\begin{aligned} q_{n+1}(\check{\boldsymbol{\theta}}_{\tau n}) - q_{n+1}(\boldsymbol{\theta}_{\tau 0}) &= (\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)'\boldsymbol{X}_n + [\check{b}_{\tau n}\check{\underline{\sigma}}_{n+1}(\check{\boldsymbol{\gamma}}_n) - b_\tau \underline{\sigma}_{n+1}(\boldsymbol{\gamma}_0)] \\ &= (\check{b}_{\tau n} - b_\tau)\underline{\sigma}_{n+1}(\boldsymbol{\gamma}_0) + b_\tau[\underline{\sigma}_{n+1}(\check{\boldsymbol{\gamma}}_n) - \underline{\sigma}_{n+1}(\boldsymbol{\gamma}_0)] \\ &\quad + b_\tau[\check{\underline{\sigma}}_{n+1}(\boldsymbol{\gamma}_0) - \underline{\sigma}_{n+1}(\boldsymbol{\gamma}_0)] + (\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)'\boldsymbol{X}_n + o_p(n^{-1/2}) \\ &= (\check{b}_{\tau n} - b_\tau)\sigma_{n+1} + (\check{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0)'\frac{b_\tau}{2\sigma_{n+1}}\frac{\partial \sigma_t^2(\boldsymbol{\lambda}_0)}{\partial \boldsymbol{\gamma}'} \\ &\quad + (\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0)'\left(\boldsymbol{X}_n - \frac{b_\tau}{2\sigma_{n+1}}\frac{\partial \sigma_t^2(\boldsymbol{\lambda}_0)}{\partial \boldsymbol{\phi}'}\right) + o_p(n^{-1/2}) \\ &= \boldsymbol{W}_{n+1}'(\check{\boldsymbol{\gamma}}_{\tau n} - \boldsymbol{\gamma}_{\tau 0}) + \boldsymbol{M}_{n+1}'(\check{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_0) + o_p(n^{-1/2}), \end{aligned}$$

where $\boldsymbol{M}_t = \boldsymbol{X}_{t-1} + 0.5\sigma_t^{-1}b_\tau \partial\sigma_t^2(\boldsymbol{\lambda}_0)/\partial\boldsymbol{\phi}'$ and $\boldsymbol{W}_t = (\sigma_t, 0.5\sigma_t^{-1}b_\tau\partial\sigma_t^2(\boldsymbol{\lambda}_0)/\partial\boldsymbol{\gamma}')'$. Hence, the proof of this corollary is accomplished. □

Proof of Theorem 3. Since the proofs for (ii) and (iii) are similar to the proof of (i), in below we only provide detailed proof for establishing the bootstrap consistency for the joint weighed estimator $\widehat{\boldsymbol{\theta}}_{\tau n}$ in (i). Recall that $\ell_t(\boldsymbol{\theta}) = \rho_\tau[Y_t - q_t(\boldsymbol{\theta})]$ and $\widetilde{\ell}_t(\boldsymbol{\theta}) = \rho_\tau[Y_t - \widetilde{q}_t(\boldsymbol{\theta})]$, where $q_t(\boldsymbol{\theta}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b\sigma_t(\boldsymbol{\lambda})$ and $\widetilde{q}_t(\boldsymbol{\theta}) = \boldsymbol{\phi}'\boldsymbol{X}_{t-1} + b\widetilde{\sigma}_t(\boldsymbol{\lambda})$. Define the following functions

$$\widehat{L}_n^*(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \omega_t\widehat{\sigma}_t^{-1}\widetilde{\ell}_t(\boldsymbol{\theta}), \quad \widetilde{L}_n^*(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \omega_t\widetilde{\sigma}_t^{-1}\ell_t(\boldsymbol{\theta}) \text{ and } L_n^*(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^n \omega_t\sigma_t^{-1}\ell_t(\boldsymbol{\theta}),$$

where $\widehat{\sigma}_t = \widetilde{\sigma}_t(\widehat{\boldsymbol{\lambda}}_n^{int})$, $\widetilde{\sigma}_t = \sigma_t(\widehat{\boldsymbol{\lambda}}_n^{int})$ and $\sigma_t = \sigma_t(\boldsymbol{\lambda}_0)$.

Similar to the consistency proof of Theorem 1 and Lemma A.3 of Zhu, Zeng, and Li (2020), by Assumptions 1, 3-5, we can show that

$$\widehat{\boldsymbol{\theta}}_{\tau n}^* - \boldsymbol{\theta}_{\tau 0} = o_p^*(1). \tag{S1.83}$$

Let $\zeta_n^*(\boldsymbol{u}) = \sum_{t=1}^n \omega_t \sigma_t^{-1} q_{1t}(\boldsymbol{u}) \{\xi_{1t}(\boldsymbol{u}) - E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}]\}$, where $q_{1t}(\boldsymbol{u})$ and $\xi_{1t}(\boldsymbol{u})$ are defined as in Lemma 3. In line with the proof of Lemma 3, by Assumptions 1, 3-5, for $\boldsymbol{u} = o_p(1)$, we can show that

$$\zeta_n^*(\boldsymbol{u}) - \zeta_n(\boldsymbol{u}) = \sum_{t=1}^n (\omega_t - 1)\sigma_t^{-1} q_{1t}(\boldsymbol{u}) \{\xi_{1t}(\boldsymbol{u}) - E[\xi_{1t}(\boldsymbol{u})|\mathcal{F}_{t-1}]\} = o_p^*(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2),$$

which implies that

$$\zeta_n^*(\boldsymbol{u}) = o_p^*(\sqrt{n}\|\boldsymbol{u}\| + n\|\boldsymbol{u}\|^2). \tag{S1.84}$$

In line with the proof of Lemma 4, by Assumption 5, for $\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0} = o_p(1)$ we have

$$n[\widehat{L}_n^*(\boldsymbol{\theta}) - \widehat{L}_n^*(\boldsymbol{\theta}_{\tau 0})] - n[\widetilde{L}_n^*(\boldsymbol{\theta}) - \widetilde{L}_n^*(\boldsymbol{\theta}_{\tau 0})] = o_p^*(\sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}\| + n\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}\|^2). \tag{S1.85}$$

Similar to the proof of Lemma 5, together with (S1.84), we can verify that

$$n[\widetilde{L}_n^*(\boldsymbol{\theta}) - \widetilde{L}_n^*(\boldsymbol{\theta}_{\tau 0})] = -\sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0})'\boldsymbol{T}_n^* + \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0})' J_n \sqrt{n}(\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}) + o_p^*(\sqrt{n}\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}\| + n\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0}\|^2), \tag{S1.86}$$

where $\boldsymbol{\theta} - \boldsymbol{\theta}_{\tau 0} = o_p(1)$ and $\boldsymbol{T}_n^* = n^{-1/2}\sum_{t=1}^n \omega_t \sigma_t^{-1} \partial q_t(\boldsymbol{\theta}_{\tau 0})/\partial \boldsymbol{\theta} \psi_\tau(\varepsilon_{t,\tau})$. By methods similar to (S1.74)-(S1.77), together with (S1.83), (S1.85) and (S1.86), it can be shown that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n}^* - \boldsymbol{\theta}_{\tau 0}) = f_\varepsilon^{-1}(b_\tau)\Sigma^{-1}(\tau)\boldsymbol{T}_n^* + o_p^*(1).$$

This together with $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n} - \boldsymbol{\theta}_{\tau 0}) = f_\varepsilon^{-1}(b_\tau)\Sigma^{-1}(\tau)\boldsymbol{T}_n + o_p(1)$, implies that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\tau n}^* - \widehat{\boldsymbol{\theta}}_{\tau n}) = f_\varepsilon^{-1}(b_\tau)\Sigma^{-1}(\tau)\frac{1}{\sqrt{n}}\sum_{t=1}^n(\omega_t - 1)\frac{1}{\sigma_t}\frac{\partial q_t(\boldsymbol{\theta}_{\tau 0})}{\partial \boldsymbol{\theta}}\psi_\tau(\varepsilon_{t,\tau}) + o_p^*(1).$$

This together with Assumption 5, we complete the proof by applying Lindeberg's central limit theorem and the Cramér-Wold device.

□

S2 Additional simulation results

Tables 1-3 report additional simulation results for the first experiment in Section 3.1 with sample size $n = 1000$, which aims to evaluate the finite-sample performance of

Table 1: Biases, ESDs, ASDs, and ECRs of the 95% confidence intervals for $\widehat{\boldsymbol{\theta}}_{\tau n}$ at $\tau = 0.05$ and 0.10, for normally distributed X_{t-1} and υ_{t-1}. The innovations follow a normal or a Student's t_5 distribution.

			Normal				t_5			
τ	n		Bias	ESD	ASD	ECR	Bias	ESD	ASD	ECR
0.05	1000	b_τ	-0.227	0.418	0.376	0.942	-0.192	0.361	0.344	0.947
		α	-0.012	0.053	0.052	0.909	-0.010	0.062	0.059	0.892
		β	-0.041	0.074	0.079	0.952	-0.041	0.080	0.089	0.963
		π	0.086	0.221	0.201	0.931	0.109	0.270	0.234	0.953
		ϕ	0.007	0.251	0.250	0.950	-0.004	0.277	0.287	0.959
0.10	1000	b_τ	-0.185	0.334	0.347	0.954	-0.135	0.294	0.290	0.965
		α	-0.014	0.056	0.053	0.898	-0.010	0.058	0.055	0.884
		β	-0.040	0.074	0.089	0.958	-0.038	0.084	0.093	0.968
		π	0.081	0.234	0.207	0.930	0.089	0.248	0.221	0.933
		ϕ	0.001	0.212	0.215	0.951	0.005	0.205	0.212	0.950

the three proposed estimators $\widehat{\boldsymbol{\theta}}_{\tau n}$, $\widetilde{\boldsymbol{\theta}}_{\tau n}$, and $\check{\boldsymbol{\theta}}_{\tau n}$, and their bootstrap approximations. Simulation findings are provided in Section 3.1 of the main manuscript.

Table 2: Biases, ESDs, ASDs, and ECRs of the 95% confidence intervals for $\widetilde{\boldsymbol{\theta}}_{\tau n}$ at $\tau = 0.05$ and 0.10, for normally distributed X_{t-1} and υ_{t-1}. The innovations follow a normal or a Student's t_5 distribution.

			Normal				t_5			
τ	n		Bias	ESD	ASD	ECR	Bias	ESD	ASD	ECR
0.05	1000	b_τ	-0.134	0.336	0.307	0.952	-0.144	0.341	0.314	0.949
		α	-0.009	0.053	0.052	0.927	-0.010	0.064	0.060	0.886
		β	-0.024	0.064	0.069	0.953	-0.029	0.077	0.085	0.952
		π	0.089	0.249	0.234	0.940	0.110	0.296	0.264	0.964
		ϕ	0.007	0.259	0.260	0.947	-0.003	0.284	0.301	0.958
0.10	1000	b_τ	-0.132	0.314	0.294	0.971	-0.102	0.284	0.267	0.961
		α	-0.011	0.055	0.054	0.908	-0.008	0.061	0.058	0.878
		β	-0.028	0.076	0.079	0.956	-0.028	0.084	0.090	0.963
		π	0.082	0.256	0.245	0.949	0.089	0.266	0.251	0.946
		ϕ	0.004	0.219	0.222	0.947	0.005	0.211	0.219	0.961

Table 3: Biases, ESDs, ASDs, and ECRs of the 95% confidence intervals for $\check{\boldsymbol{\theta}}_{\tau n}$ at $\tau = 0.05$ and 0.10, for normally distributed X_{t-1} and υ_{t-1}. The innovations follow a normal or a Student's t_5 distribution.

			Normal				t_5			
τ	n		Bias	ESD	ASD	ECR	Bias	ESD	ASD	ECR
0.05	1000	b_τ	-0.131	0.332	0.297	0.964	-0.129	0.324	0.304	0.955
		α	-0.010	0.052	0.052	0.932	-0.009	0.064	0.060	0.887
		β	-0.023	0.066	0.068	0.952	-0.027	0.076	0.083	0.958
		π	0.086	0.233	0.238	0.956	0.107	0.292	0.264	0.957
		ϕ	0.002	0.151	0.147	0.953	-0.004	0.149	0.145	0.944
0.10	1000	b_τ	-0.129	0.318	0.288	0.957	-0.099	0.275	0.261	0.974
		α	-0.011	0.055	0.053	0.911	-0.008	0.062	0.057	0.879
		β	-0.027	0.076	0.078	0.955	-0.026	0.080	0.088	0.959
		π	0.087	0.256	0.248	0.951	0.090	0.272	0.254	0.947
		ϕ	0.002	0.151	0.147	0.953	-0.004	0.149	0.145	0.944

Bibliography

Berkes, I., L. Horváth, and P. Kokoszka (2003). GARCH processes: structure and estimation. *Bernoulli 9*, 201–227.

Huber, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability 1*, 221–233.

Huber, P. (1973). Robust regression: Asymptotics, conjectures and monte carlo. *Annals of Statistics 1*, 799–821.

Knight, K. (1998). Limiting distributions for L1 regression estimators under general conditions. *The Annals of Statistics 26*, 755–770.

Lee, S. and J. Noh (2013). Quantile regression estimator for GARCH models. *Scandinavian Journal of Statistics 40*, 2–20.

Ling, S. and M. McAleer (2003). Asymptotic theory for a vector ARMA–GARCH model. *Econometric Theory 19*, 280–310.

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory 1*, 295–314.

Zheng, Y., Q. Zhu, G. Li, and Z. Xiao (2018). Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *Journal of the Royal Statistical Society, Series B 80*, 975–993.

Zhu, K. and S. Ling (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA–GARCH/IGARCH models. *Annal of Statistics 39*, 2131–2163.