Submitted to the Annals of Statistics

SUPPLEMENT TO "INFERENCE FOR CONDITIONAL VALUE-AT-RISK OF A PREDICTIVE REGRESSION"

By Yi He[§], Yanxi Hou[‡,¶], Liang Peng[*,‖] and Haipeng Shen[†,**]

University of Amsterdam[§], *Fudan University*[¶], *Georgia State University*[‖] *and The University of Hong Kong*[**]

In this supplement, we provide additional simulation and empirical analysis results, and we prove all the theorems stated in the main document. Appendix A reports the empirical coverage probabilities for the 90% confidence intervals in our simulation study. Appendix B reports the time series and autocorrelations plots of the fitted residuals based on different models for our bank dataset. In Appendixes C and D, we provide the proofs from Section 2 and 3, respectively. In Appendix E, we develop the empirical likelihood method for the quantile regression model. Finally, in Appendix F, we prove the theorem for GARCH errors from Section 4.

APPENDIX A: MORE SIMULATION RESULTS

We repeat the simulation study in Section 5.2 for i.i.d. errors and Section 5.3 for GARCH errors, and report the empirical coverage probabilities for the 90% confidence intervals.

Tables 1 and 2 show the results for the empirical-likelihood confidence intervals based on the weighted least squares estimator (LSE), that based the weighted quantile regression estimator (QR) developed in Appendix E below, and the Gaussian approximations that based on the asymptotic variance in Chun et al. (2012) (CSU) which tends to over-cover in theory. Again we observe that the proposed empirical likelihood confidence interval based on least squares estimation has an accurate coverage probability in all cases, while that based on the quantile regression estimation tend to under-cover

[*]Research is partially supported by the Simons Foundation.

[†]Research is partially supported by the Ministry of Science and Technology Major Project of China 2017YFC1310903, University of Hong Kong Stanley Ho Alumni Challenge Fund, and HKU University Research Committee Seed Funding Award 104004215.

[‡]Corresponding author. Research is partially supported by National Natural Science Foundation of China 71803026 and 71991471.

MSC 2010 subject classifications: 62G20, 62M10, 62P20

Keywords and phrases: conditional risk measures, empirical likelihood, least squares estimation, interval estimation, quantile regression, risk management, value-at-risk

for the case $\alpha = 0.99$. Chun et al. (2012)'s confidence intervals show poor performance, especially in non-normal cases.

TABLE 1
Empirical coverage probabilities for the 90% confidence intervals of $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$ *and* $\mathrm{VaR}_{\boldsymbol{x}}(0.99)$ *for* N*(0,1) errors, with bandwidth* $h_i = 0.5i \times n^{-1/3}$.

Model	$t(1.5)$+AR(1)		$t(1.5)$+GARCH(1,1)		AR(1)+GARCH(1,1)	
90% CI of $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$, $\boldsymbol{x} = (0.1, 0.1)^T$						
CSU	0.9303	0.9291	0.9319	0.9314	0.9269	0.9283
LSE, h_1	0.9047	0.8983	0.9023	0.9012	0.8999	0.8992
LSE, h_2	0.9040	0.8978	0.9006	0.9006	0.8992	0.8990
LSE, h_3	0.9025	0.8992	0.9002	0.9007	0.8996	0.8975
QR, h_1	0.9035	0.9036	0.9058	0.9079	0.9102	0.9058
QR, h_2	0.9011	0.9006	0.9033	0.9035	0.9067	0.9017
QR, h_3	0.8980	0.8982	0.9011	0.9022	0.9020	0.8981
90% CI of $\mathrm{VaR}_{\boldsymbol{x}}(0.99)$, $\boldsymbol{x} = (0.1, 0.1)^T$						
CSU	0.8990	0.9031	0.9011	0.9052	0.8953	0.9101
LSE, h_1	0.9191	0.8960	0.9141	0.9024	0.9003	0.9030
LSE, h_2	0.9193	0.8954	0.9153	0.9041	0.9000	0.9013
LSE, h_3	0.9196	0.8945	0.9180	0.9038	0.9038	0.9015
QR, h_1	0.9021	0.8934	0.8848	0.8885	0.8779	0.8449
QR, h_2	0.8958	0.8889	0.8813	0.8824	0.8723	0.8428
QR, h_3	0.8918	0.8862	0.8798	0.8840	0.8671	0.8466

TABLE 2
Empirical coverage probabilities the 90% confidence intervals of $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$ *and* $\mathrm{VaR}_{\boldsymbol{x}}(0.99)$ *for centered-LN(0,1/16) errors, with bandwidth* $h_i = 0.5i \times n^{-1/3}$.

Model	$t(1.5)$+AR(1)		$t(1.5)$+GARCH(1,1)		AR(1)+GARCH(1,1)	
n	2000	5000	2000	5000	2000	5000
90% CI of $\mathrm{VaR}_{\boldsymbol{x}}(0.95)$, $\boldsymbol{x} = (0.1, 0.1)^T$						
CSU	0.9906	1.0000	0.9933	1.0000	0.9927	1.0000
LSE, h_1	0.9025	0.8985	0.8996	0.8999	0.8985	0.8968
LSE, h_2	0.8980	0.8944	0.8961	0.8982	0.8962	0.8925
LSE, h_3	0.8821	0.8878	0.8816	0.8818	0.8775	0.8747
QR, h_1	0.9013	0.8977	0.8813	0.8943	0.8862	0.8924
QR, h_2	0.8972	0.8944	0.8809	0.8912	0.8805	0.8857
QR, h_3	0.8799	0.8863	0.8680	0.8739	0.8655	0.8658
90% CI of $\mathrm{VaR}_{\boldsymbol{x}}(0.99)$, $\boldsymbol{x} = (0.1, 0.1)^T$						
CSU	0.9357	0.9396	0.9397	0.9397	0.9341	0.9458
LSE, h_1	0.8918	0.8931	0.9032	0.9038	0.8946	0.9001
LSE, h_2	0.8906	0.8935	0.9017	0.9020	0.8956	0.8979
LSE, h_3	0.8904	0.8904	0.8971	0.8956	0.8902	0.8945
QR, h_1	0.8686	0.8407	0.7902	0.6692	0.7456	0.6539
QR, h_2	0.8611	0.8391	0.8066	0.6811	0.7520	0.6573
QR, h_3	0.8528	0.8289	0.8029	0.6988	0.7640	0.6600

Table 3 shows the results for the regression models with GARCH regression errors. Our empirical likelihood confidence intervals show good coverage performance overall, while those with larger bandwidth seems (slightly) better.

TABLE 3
Empirical coverage probabilities for the 90% confidence intervals of $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.95)$ *and* $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.99)$ *for GARCH(1,1) errors* ($\epsilon_t = \sigma_t e_t$, $\sigma_t^2 = 0.1 + 0.3\sigma_{t-1}^2 + 0.3\epsilon_{t-1}^2$, $e_t \sim N(0,1)$)*, with bandwidth* $h_i = 0.5i \times n^{-1/3}$.

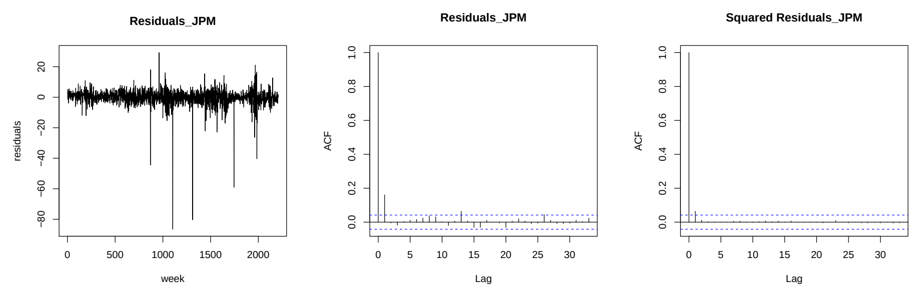
Model	$t(1.5)$+AR(1)		$t(1.5)$+GARCH(1,1)		AR(1)+GARCH(1,1)	
n	2000	5000	2000	5000	2000	5000
90% CI of $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.95)$, $\boldsymbol{x} = (0.1, 0.1)^T$ and $\sigma = \sqrt{0.25}$						
LSE, h_1	0.897	0.912	0.909	0.916	0.915	0.912
LSE, h_2	0.885	0.910	0.891	0.908	0.908	0.912
LSE, h_3	0.882	0.902	0.900	0.907	0.900	0.903
90% CI of $\mathrm{VaR}_{\boldsymbol{x},\sigma}(0.99)$, $\boldsymbol{x} = (0.1, 0.1)^T$ and $\sigma = \sqrt{0.25}$						
n	2000	5000	2000	5000	2000	5000
LSE, h_1	0.917	0.901	0.923	0.922	0.920	0.926
LSE, h_2	0.918	0.910	0.916	0.916	0.906	0.911
LSE, h_3	0.906	0.895	0.903	0.906	0.901	0.909

APPENDIX B: MORE EMPIRICAL ANALYSIS RESULTS

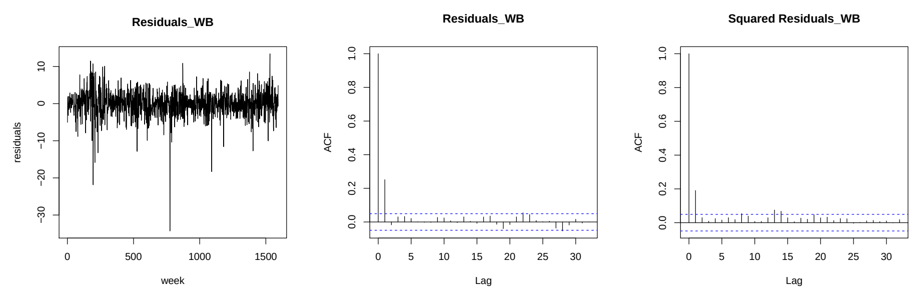
We repeat the model specification analysis for all four US banks: Citigroup, JPMorgan Chase, Wachovia, and Wells Fargo. Like in the main document, we compare the results based on the original regression model (6.1) proposed by Adrian and Brunnermeier (2016) and those based on the extended model (6.3), while the regression errors are always fitted by the GARCH model (6.2). In particular, we plot the time series of the least-squares and GARCH residuals and the autocorrelations of their nominal levels and squares. The plots for Citigroup are reported in Figures 2 and 3 in the main document, and we report the plots for other banks in this appendix.

Figure 1 shows the time series and autocorrelation plots of the least-squares residuals based on the original regression model (6.1). Like in the main document, we observe significant autocorrelations at lag one for all the banks. This issue cannot be removed by the GARCH model (6.2), as the corresponding GARCH residuals in Figure 2 still demonstrate non-trivial autocorrelations.

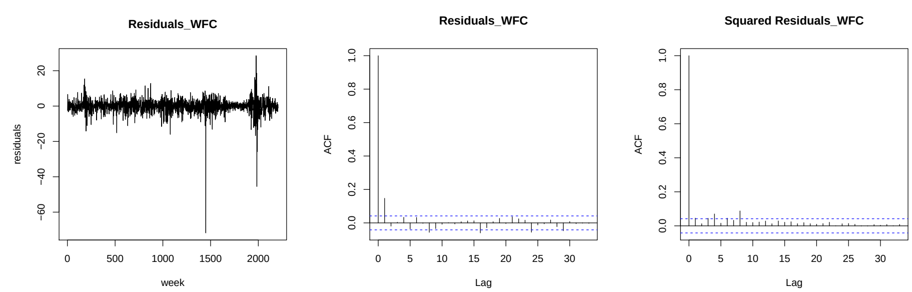
Figures 3 and 4 present similar plots for the least-squares residuals from the extended regression model (6.3) and the corresponding GARCH residuals, respectively. Now the GARCH residuals show insignificant autocorrelations, suggesting a good fitness of the extended model to our bank data.



(a) JPMorgan Chase

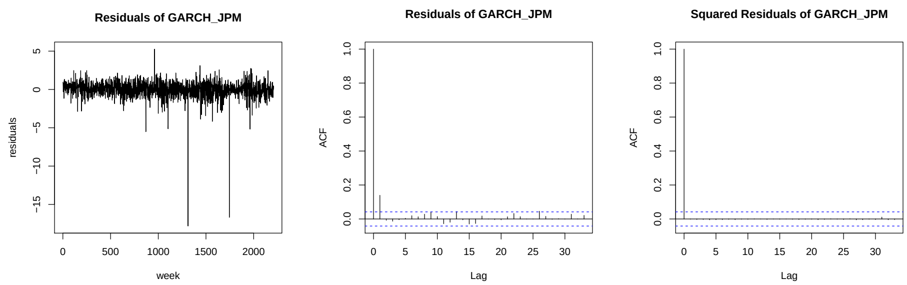


(b) Wachovia

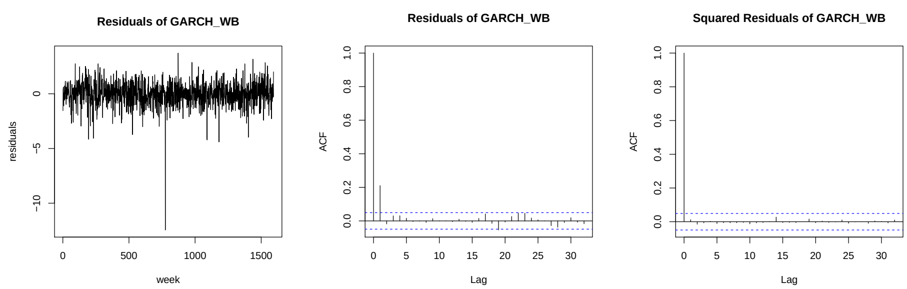


(c) Wells Fargo

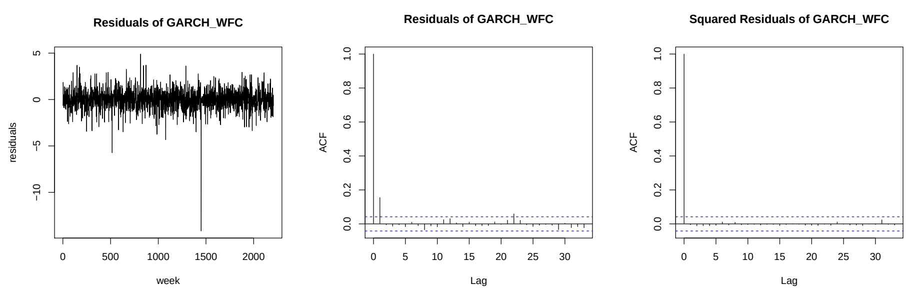
Fig 1: Time series and autocorrelation plots for least-squares residuals for JPMorgan Chase, Wachovia and Wells Fargo based on model (6.1). The middle plots are for their nominal levels, and right plots are for their squares.



(a) JPMorgan Chase

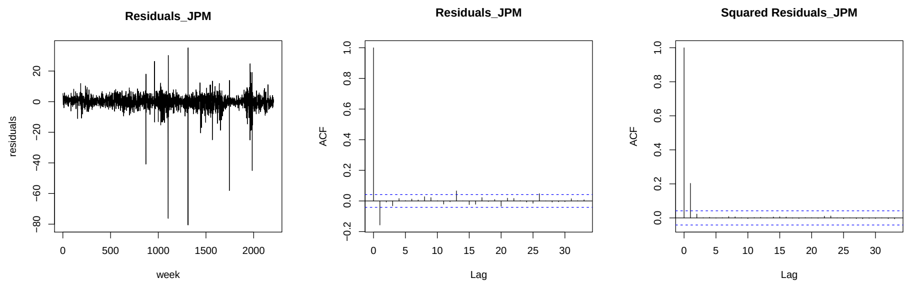


(b) Wachovia

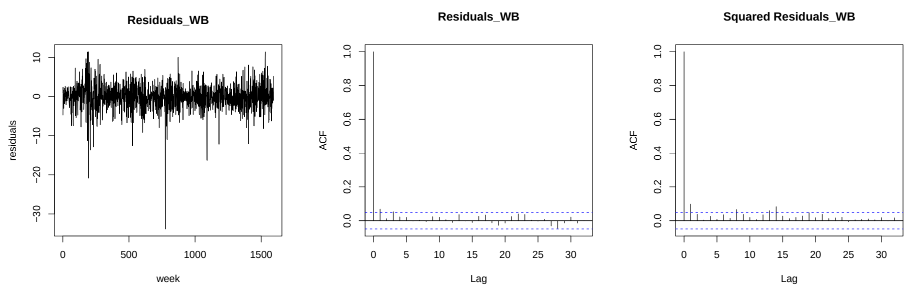


(c) Wells Fargo

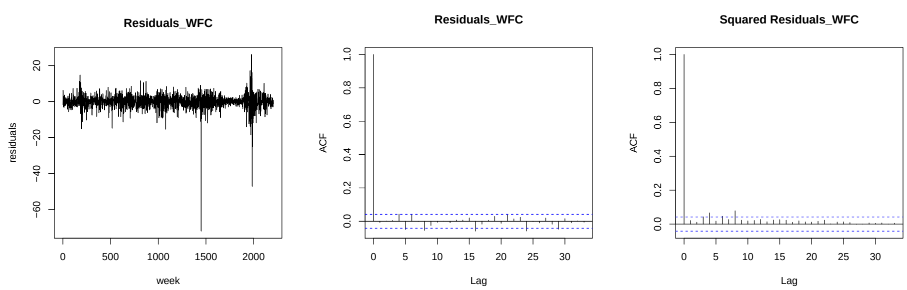
Fig 2: Time series and autocorrelation plots for GARCH residuals for JP-Morgan Chase, Wachovia and Wells Fargo based on model (6.1) and (6.2). The middle plots are for their nominal levels, and right plots are for their squares.



(a) JPMorgan Chase

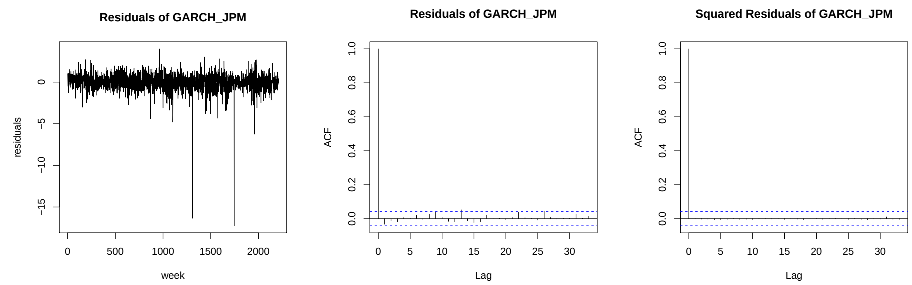


(b) Wachovia

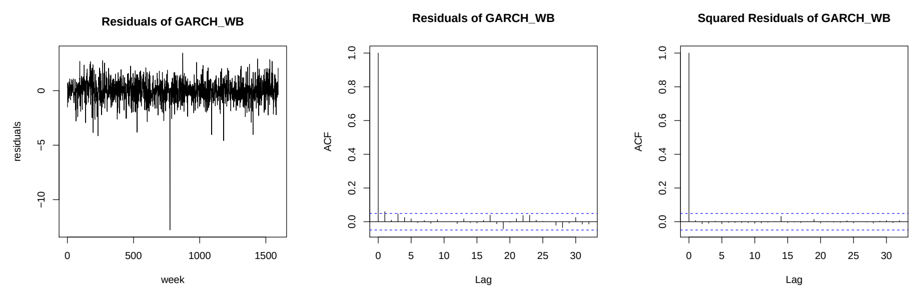


(c) Wells Fargo

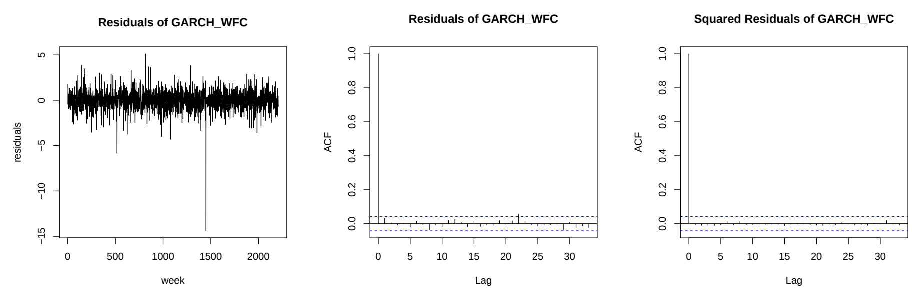
Fig 3: Time series and autocorrelation plots for least-squares residuals for JPMorgan Chase, Wachovia and Wells Fargo based on model (6.3). The middle plots are for their nominal levels, and right plots are for their squares.



(a) JPMorgan Chase



(b) Wachovia



(c) Wells Fargo

Fig 4: Time series and autocorrelation plots for GARCH residuals for JP-Morgan Chase, Wachovia and Wells Fargo based on model (6.3) and (6.2). The middle plots are for their nominal levels, and right plots are for their squares.

APPENDIX C: PROOFS FROM SECTION 2

PROOF OF THEOREM 1. Define $F_{n,\widehat{\epsilon}}(x) = \frac{1}{n}\sum_{t=1}^{n} I(\widehat{\epsilon}_t \le x)$, $F_{n,\epsilon}(x) = \frac{1}{n}\sum_{t=1}^{n} I(\epsilon_t \le x)$, and write $\theta = \mathrm{VaR}_{\boldsymbol{x}}(\alpha)$, $\widehat{\theta} = \widehat{\mathrm{VaR}}_{\boldsymbol{x}}(\alpha)$, $\gamma = F_{\epsilon}^{-1}(\alpha)$ and $\widehat{\gamma} = F_{n,\widehat{\epsilon}}^{-1}(\alpha)$. Then under Conditions A1–A4, we have

(C.1)

$$\widehat{\theta} = F_{n,\widehat{\epsilon}}^{-1}(\alpha) + \widehat{\boldsymbol{\beta}}^T \boldsymbol{z} \quad \text{and} \quad \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \Omega^{-1}\left\{\frac{1}{n}\sum_{t=1}^{n} \epsilon_t \boldsymbol{Z}_t\right\}\{1 + o_P(1)\} = O_P(n^{-1/2}).$$

By noting that $F_{n,\hat{\epsilon}}(x) = \frac{1}{n}\sum_{t=1}^{n} I(\epsilon_t \le x + (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{Z}_t)$, $\max_{1\le t\le n} |(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{Z}_t| = o_p(1)$ and $\{\epsilon_t\}$ is independent of $\{\boldsymbol{Z}_t\}$, it follows from approximation results of residual-based empirical process, see (2.3) of Mammen (1963), that

$$\begin{aligned}
&\sqrt{n}\{F_{n,\hat{\epsilon}}(x) - F_{\epsilon}(x)\} \\
&= \sqrt{n}\{F_{n,\epsilon}(x) - F_{\epsilon}(x)\} + F_{\epsilon}'(x)\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{Z}_t + o_p(1) \\
&=: \mathbb{F}_n(x) + F_{\epsilon}'(x)\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T E(\boldsymbol{Z}_1) + o_p(1),
\end{aligned}$$

where the $o_p(1)$ is uniform for x in a neighborhood around γ. Next, to obtain the approximation for quantiles, one may apply the Hadamard-differentiable result of the inverse function, see, e.g., Lemma 21.4 in van der Vaart (2000), after showing that $\mathbb{F}_n(x) + F_{\epsilon}'(x)\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T E(\boldsymbol{Z}_1)$ converges in a space. Because we are interested in the quantile at the particular level α, we employ the following delta method without deriving the above convergence in space and recalculating the limit for our conditional VaR.

Substituting x by the consistent estimator $\widehat{\gamma}$ and using the stochastic equicontinuity of the empirical process $\mathbb{F}_n$,

$$\begin{aligned}
\sqrt{n}(\alpha - F_{\epsilon}(\widehat{\gamma})) &= \mathbb{F}_n(\widehat{\gamma}) + F_{\epsilon}'(\widehat{\gamma})\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T E(\boldsymbol{Z}_1) + o_p(1) \\
&= \mathbb{F}_n(\gamma) + F_{\epsilon}'(\gamma)\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T E(\boldsymbol{Z}_1) + o_p(1).
\end{aligned}$$

Applying the delta method on the left-hand-side

$$\sqrt{n}(\alpha - F_{\epsilon}(\widehat{\gamma})) = \sqrt{n}(F_{\epsilon}(\gamma) - F_{\epsilon}(\widehat{\gamma})) = \sqrt{n}F_{\epsilon}'(\gamma)(\gamma - \widehat{\gamma}) + o_p(1),$$

and it follows that

$$\begin{aligned}
&\sqrt{n}\{\widehat{\gamma} - \gamma\} \\
&= -\frac{\sqrt{n}\{F_{n,\epsilon}(\gamma) - \alpha\}}{F_{\epsilon}'(\gamma)} - \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T E(\boldsymbol{Z}_{\boldsymbol{1}}) + o_P(1)
\end{aligned}$$

$$=-\frac{1}{F'_\epsilon(\gamma)}\cdot\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(I(\epsilon_t\le\gamma)-\alpha\right)-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\epsilon_t\boldsymbol{Z}_t^T\Omega^{-1}E(\boldsymbol{Z}_1)+o_P(1)$$
$$=-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{\frac{I(\epsilon_t\le\gamma)-\alpha}{F'_\epsilon(\gamma)}+\epsilon_t\boldsymbol{Z}_t^T\Omega^{-1}E(\boldsymbol{Z}_1)\right\}+o_P(1).$$

Hence,

$$\begin{aligned}
&\sqrt{n}\left(\widehat{\mathrm{VaR}}_{\boldsymbol{x}}(\alpha)-\mathrm{VaR}_{\boldsymbol{x}}(\alpha)\right)\\
=&\sqrt{n}\{\widehat{\gamma}-\gamma\}+\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^T\boldsymbol{z}\\
=&-\frac{\sqrt{n}\{F_{n,\epsilon}(\gamma)-\alpha\}}{F'_\epsilon(\gamma)}+\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)^T\boldsymbol{z}-\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})^TE(\boldsymbol{Z}_{\mathbf{1}})+o_P(1)\\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\{-\frac{I(\epsilon_t\le\gamma)-\alpha}{F'_\epsilon(\gamma)}+\epsilon_t\boldsymbol{Z}_t^T\Omega^{-1}\boldsymbol{z}-\epsilon_t\boldsymbol{Z}_t^T\Omega^{-1}E(\boldsymbol{Z}_1)\right\}+o_P(1)\\
=:&S_{nn}+o_P(1),
\end{aligned}$$

with, for $t=1,\dots,n$,

$$S_{nt}:=\frac{1}{\sqrt{n}}\sum_{m=1}^{t}\left\{-\frac{I(\epsilon_m\le\gamma)-\alpha}{F'_\epsilon(\gamma)}+\epsilon_m\boldsymbol{Z}_m^T\Omega^{-1}\boldsymbol{z}-\epsilon_m\boldsymbol{Z}_m^T\Omega^{-1}E(\boldsymbol{Z}_1)\right\}.$$

Define a filtration $\{\mathcal{F}_t\}_{t\ge0}=\{\sigma\left(\epsilon_1,\dots,\epsilon_t,\boldsymbol{Z}_{\mathbf{1}},\dots,\boldsymbol{Z}_t,\boldsymbol{Z}_{t+1}\right)\}_{t\ge0}$. Note that $(S_{nt},\mathcal{F}_t)$ is a zero-mean, squared integrable martingale with differences

$$\begin{aligned}
D_{nt}=&\frac{1}{\sqrt{n}}\left\{-\frac{I(\epsilon_t\le\gamma)-\alpha}{F'_\epsilon(\gamma)}+\epsilon_t\boldsymbol{Z}_t^T\Omega^{-1}\boldsymbol{z}-\epsilon_t\boldsymbol{Z}_t^T\Omega^{-1}E(\boldsymbol{Z}_1)\right\}\\
=:&D_{nt,1}+D_{nt,2}+D_{nt,3}.
\end{aligned}$$

Using Markov inequality, we have for all $\varepsilon>0$

$$\begin{aligned}
&P\left(\max_t|D_{nt}|>\varepsilon\right)\\
\le&\sum_{t=1}^{n}P\left(|D_{nt,1}|>\varepsilon/2\right)+\sum_{t=1}^{n}P\left(|D_{nt,2}+D_{nt,3}|>\varepsilon/2\right)\\
\le&n^{-\iota_0/2}\frac{1}{n}\sum_{t=1}^{n}\frac{(2/F'_\epsilon(\gamma))^{2+\iota_0}+E|\epsilon_t|^{2+\iota_0}E\left\|\boldsymbol{Z}_t\right\|^{2+\iota_0}\left\|\Omega^{-1}\right\|^{2+\iota_0}\left\|\boldsymbol{z}-E(\boldsymbol{Z}_1)\right\|^{2+\iota_0}}{(\varepsilon/2)^{2+\iota_0}}\\
\to&0,\quad\text{as } n\to\infty.
\end{aligned}$$

In other words, $\max_t |D_{nt}| \xrightarrow{P} 0$. Using the law of large numbers for martingale (e.g. Theorem 2.13 in Hall and Heyde (1980)) with the filtration $\{\mathcal{F}_t\}_{t\geq 0}$, under Conditions A1–A4 it is easy to verify that, as $n \to \infty$,

$$\begin{aligned}
\sum_{t=1}^n D_{nt}^2 &= \sum_{t=1}^n D_{nt,1}^2 + \sum_{t=1}^n D_{nt,2}^2 + \sum_{t=1}^n D_{nt,3}^2 \\
&\quad + 2\sum_{t=1}^n D_{nt,1}D_{nt,3} + 2\sum_{t=1}^n D_{nt,2}D_{nt,3} + 2\sum_{t=1}^n D_{nt,1}D_{nt,2} \\
&\xrightarrow{P} \omega^2 + \sigma^2 \boldsymbol{z}^T\Omega^{-1}\boldsymbol{z} + \sigma^2 E(\boldsymbol{Z}_1)^T\Omega^{-1}E(\boldsymbol{Z}_1) \\
&\quad + 2\frac{E(\epsilon_1 I(\epsilon_1 \leq F_\epsilon^{-1}(\alpha)))}{F_\epsilon'(F_\epsilon^{-1}(\alpha))} E(\boldsymbol{Z}_1^T)\Omega^{-1}E(\boldsymbol{Z}_1) - 2\sigma^2 E(\boldsymbol{Z}_1^T)\Omega^{-1}\boldsymbol{z} \\
&\quad - 2\frac{E(\epsilon_1 I(\epsilon_1 \leq F_\epsilon^{-1}(\alpha)))}{F_\epsilon'(F_\epsilon^{-1}(\alpha))} E(\boldsymbol{Z}_1^T)\Omega^{-1}\boldsymbol{z} \\
&= \omega^2 + \sigma^2 \boldsymbol{z}^T\Omega^{-1}\boldsymbol{z} + \Delta_1 + \Delta_2.
\end{aligned}$$

Furthermore, it is easy to verify that

$$E\left(\max_t \|D_{nt}\|^2\right) \leq E\left(\sum_{t=1}^n \|D_{nt}\|^2\right) \leq 3\sum_{i=1}^3 E\left(\sum_{t=1}^n \|D_{nt,i}\|^2\right) = O(1).$$

The theorem then follows from Theorem 3.2 in Hall and Heyde (1980). □

PROOF OF THEOREM 2. Define $F_{n,\widehat{\epsilon},k}(x) = \frac{1}{n}\sum_{t=1}^n I(\widehat{\epsilon}_t \leq x)I(|X_{t,k}| \leq n^{1/d^*})$ and $F_{n,\epsilon,k}(x) = \frac{1}{n}\sum_{t=1}^n I(\epsilon_t \leq x)I(|X_{t,k}| \leq n^{1/d^*})$ for some $d^* \in (d, 2)$ with d given in condition B6. Now, because $\widehat{\beta}_k - \beta_k$ has a faster rate of convergence than n^{-1/d^*} (see, e.g., Davis and Wu (1997), Section 4), we have $\max_{1\leq t\leq n} |(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \boldsymbol{Z}_i I(|X_{t,k}| \leq n^{1/d^*})| = o_p(1)$, which allows us to easily derive the following approximation results of residual-based empirical process by noting the independence of $\{\epsilon_t\}$ and $\{\boldsymbol{Z}_t\}$ like in Shorack and Wellner (1986), Chapter 4:

$$\begin{aligned}
&\sqrt{n}\left\{F_{n,\hat{\epsilon},k}(x) - F_\epsilon(x)\frac{1}{n}\sum_{t=1}^n I\big(|X_{t,k}| \leq n^{1/d^*}\big)\right\} \\
&= \sqrt{n}\left\{F_{n,\epsilon,k}(x) - F_\epsilon(x)\frac{1}{n}\sum_{t=1}^n I\big(|X_{t,k}| \leq n^{1/d^*}\big)\right\} \\
&\quad + F_\epsilon'(x)\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T\frac{1}{n}\sum_{t=1}^n \boldsymbol{Z}_t I(|X_{t,k}| \leq n^{1/d^*}) + o_p(1)
\end{aligned}$$

$$= \sqrt{n}\left\{F_{n,\epsilon,k}(x) - F_\epsilon(x)\frac{1}{n}\sum_{t=1}^{n} I\big(|X_{t,k}| \le n^{1/d^*}\big)\right\}$$
$$+ F_\epsilon'(x)\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T E(\tilde{\boldsymbol{Z}}_1) + o_p(1)$$

where the $o_p(1)$ is uniform for x in a neighborhood around γ. Because $d^* < 2 < 2d$, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} I(|X_{t,k}| > n^{1/d^*}) = o_p(1).$$

Therefore, the theorem can be shown in the same way as the proof of Theorem 1.

□

APPENDIX D: PROOFS FROM SECTION 3

Our proof of Theorem 3 exploits heavily the so-called local asymptotic normality (LAN) property (see, e.g., Chapter 7 in van der Vaart (2000)) of the empirical likelihood ratio function, for which we need some lemmas. For presentation convenience, we shall assume the conditions of Theorem 3 throughout this subsection. Denote $\|\cdot\|$ as the Frobenius norm, that is, for arbitrary matrix (or vector) A we define $\|A\|$ as the square root of the sum of the squares of its entries.

LEMMA D.1. *As $n \to \infty$,*

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\beta}^0, \theta^0) \xrightarrow{d} N(0, \Sigma_1),\ \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\beta}^0, \theta^0)\boldsymbol{W}_t^T(\boldsymbol{\beta}^0, \theta^0) \xrightarrow{P} \Sigma_1,\ \textit{and}$$
$$\frac{1}{n}\sum_{t=1}^{n} \frac{\partial}{\partial \boldsymbol{\beta}}\boldsymbol{W}_t(\boldsymbol{\beta}^0, \theta^0) \xrightarrow{P} \Sigma_2.$$

PROOF. It is straightforward to verify the above statements by using the weak law of large numbers and the central limit theorem for martingale differences array as in the proof of Theorem 1 with the same filtration $\{\mathcal{F}_t\}_{t\ge 0} = \{\sigma(\epsilon_1, \ldots, \epsilon_t, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_{t+1})\}_{t\ge 0}$; see, e.g., Theorem 2.13 and 3.2 in Hall and Heyde (1980). Note that for the first part we need to use the fact that

$$E\left(\boldsymbol{W}_{t,1}(\boldsymbol{\beta}^0, \theta^0)|\mathcal{F}_{t-1}\right) = O(h^2) = o(n^{-1/2}),$$
$$E\left(\boldsymbol{W}_{t,i}(\boldsymbol{\beta}^0, \theta^0)|\mathcal{F}_{t-1}\right) = 0, \quad i = 2, \ldots, k+2,$$

uniformly for $t \ge 1$. □

LEMMA D.2. *For all $a \in (0, r_0)$, as $n \to \infty$,*

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \le n^{-1/a}} \left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0)}{\partial \boldsymbol{\beta}} - \Sigma_2 \right\| \xrightarrow{P} 0.$$

PROOF. By Lemma D.1, it suffices to show that

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\partial \boldsymbol{W}_{t,1}(\boldsymbol{\beta}, \theta^0)}{\partial \boldsymbol{\beta}} - \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \boldsymbol{W}_{t,1}(\boldsymbol{\beta}^0, \theta^0)}{\partial \boldsymbol{\beta}} \xrightarrow{P} 0.$$

uniformly in $\{\boldsymbol{\beta}\colon \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \le n^{-1/a}\}$. All the statements below hold uniformly for $\boldsymbol{\beta}$ such that $\left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \le n^{-1/a}$; for presentation convenience, we do not repeat this argument.

Note that

$$\begin{aligned}
&\left\| \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \boldsymbol{W}_{t,1}(\boldsymbol{\beta}, \theta^0)}{\partial \boldsymbol{\beta}} - \frac{1}{n} \sum_{t=1}^{n} \frac{\partial \boldsymbol{W}_{t,1}(\boldsymbol{\beta}^0, \theta^0)}{\partial \boldsymbol{\beta}} \right\| \\
\le& \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{1}{h} g\left(\frac{F_\epsilon^{-1}(\alpha) - \epsilon_t + (\boldsymbol{\beta} - \boldsymbol{\beta}^0)^T (\boldsymbol{Z}_t - \boldsymbol{z})}{h} \right) - \frac{1}{h} g\left(\frac{F_\epsilon^{-1}(\alpha) - \epsilon_t}{h} \right) \right\| \frac{\|\boldsymbol{Z}_t - \boldsymbol{z}\|}{\|\boldsymbol{Z}_t\|^2} \\
\le& \frac{(1 + \|\boldsymbol{z}\|)}{nh} \sum_{t=1}^{n} \frac{1}{\|\boldsymbol{Z}_t\|} \left\| g\left(\frac{F_\epsilon^{-1}(\alpha) - \epsilon_t + (\boldsymbol{\beta} - \boldsymbol{\beta}^0)^T (\boldsymbol{Z}_t - \boldsymbol{z})}{h} \right) - g\left(\frac{F_\epsilon^{-1}(\alpha) - \epsilon_t}{h} \right) \right\| \\
&\cdot I\left[|\epsilon_t - F_\epsilon^{-1}(\alpha)| \le h \right] + \frac{(1 + \|\boldsymbol{z}\|)}{n \|\boldsymbol{Z}_t\|} \sum_{t=1}^{n} \left| \frac{1}{h} g\left(\frac{F_\epsilon^{-1}(\alpha) - \epsilon_t + (\boldsymbol{\beta} - \boldsymbol{\beta}^0)^T (\boldsymbol{Z}_t - \boldsymbol{z})}{h} \right) \right| \\
&\cdot I\left[|\epsilon_t - F_\epsilon^{-1}(\alpha)| > h \right] =: T_1 + T_2.
\end{aligned}$$

Applying Taylor expansion to get, for some large $C > 0$,

$$\begin{aligned}
T_1 \le& \sup_x |g'(x)| \cdot \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \frac{C}{nh^2} \sum_{t=1}^{n} \frac{\|\boldsymbol{Z}_t - \boldsymbol{z}\|}{\|\boldsymbol{Z}_t\|} I[|\epsilon_t - F_\epsilon^{-1}(\alpha)| \le h] \\
\le& \frac{C}{n^{1/a} h} \cdot \frac{1}{nh} \sum_{t=1}^{n} I[|\epsilon_t - F_\epsilon^{-1}(\alpha)| \le h] = o_P(1) \cdot (f(F_\epsilon^{-1}(\alpha)) + o_P(1)) = o_P(1),
\end{aligned}$$

noting that $n^{1/a} h \to \infty$, as $n \to \infty$.

Recall that $g(\cdot)$ is bounded and only supported on $[-1, 1]$. For some large $C > 0$, we have

$$T_2 \le \frac{C}{nh} \sum_{t=1}^{n} \frac{1}{\|\boldsymbol{Z}_t\|} I\left[0 \le \epsilon_t - F_\epsilon^{-1}(\alpha) - h \le \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \|\boldsymbol{Z}_t\| + \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \|\boldsymbol{z}\| \right]$$

$$
\begin{aligned}
&+\frac{C}{nh}\sum_{t=1}^{n}\frac{1}{\|\boldsymbol{Z}_t\|}I\left[0\geq \epsilon_t - F_\epsilon^{-1}(\alpha)+h\geq -\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\right\|\|\boldsymbol{Z}_t\|-\left\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\right\|\|\boldsymbol{z}\|\right]\\
\leq&\frac{C}{nh}\sum_{t=1}^{n}\frac{1}{\|\boldsymbol{Z}_t\|}I\left[0\leq \epsilon_t - F_\epsilon^{-1}(\alpha)-h\leq n^{-1/a}\|\boldsymbol{Z}_t\|+n^{-1/a}\|\boldsymbol{z}\|\right]\\
&+\frac{C}{nh}\sum_{t=1}^{n}\frac{1}{\|\boldsymbol{Z}_t\|}I\left[0\geq \epsilon_t - F_\epsilon^{-1}(\alpha)+h\geq -n^{-1/a}\|\boldsymbol{Z}_t\|-n^{-1/a}\|\boldsymbol{z}\|\right]\\
=:&T_{21}+T_{22}.
\end{aligned}
$$

In the sequel, we shall only prove that $T_{21}=o_P(1)$; the proof of $T_{22}=o_P(1)$ is completely analogous and therefore omitted. Observe that

$$
\max_{1\leq t\leq n} n^{-1/a}\|\boldsymbol{Z}_t\|\, I\left[\|\boldsymbol{Z}_t\|\leq n^{1/r_0}\right]\leq n^{-1/a}\cdot n^{1/r_0}=o(1).
$$

It follows that, for large n and large $C>0$,

$$
\begin{aligned}
T_{21}\leq&\frac{1}{nh}\sum_{t=1}^{n}\frac{1}{\|\boldsymbol{Z}_t\|}I\left[0\leq \epsilon_t - F_\epsilon^{-1}(\alpha)-h\leq n^{-1/a}\|\boldsymbol{Z}_t\|+n^{-1/a}\|\boldsymbol{z}\|\right]\\
&\cdot I\left[\|\boldsymbol{Z}_t\|\leq n^{1/r_0}\right]+\frac{1}{nh}\sum_{t=1}^{n}\frac{1}{\|\boldsymbol{Z}_t\|}I\left[\|\boldsymbol{Z}_t\|> n^{1/r_0}\right]=T_{21,1}+T_{21,2}
\end{aligned}
$$

where obviously $0\leq T_{21,2}\leq \frac{1}{n^{1/r_0}h}=o(1)$.

Using Taylor expansion and law of iterated expectations for each term in $T_{21,1}$, we obtain

$$
E(T_{21,1})\leq CE\left(\frac{1}{nh}\sum_{t=1}^{n}n^{-1/a}\cdot\frac{\|\boldsymbol{Z}_t\|+\|\boldsymbol{z}\|}{\|\boldsymbol{Z}_t\|}\right)\leq\frac{C(1+\|\boldsymbol{z}\|)}{n^{1/a}h}=o(1).
$$

It is then straightforward to show $T_{21}=o_P(1)$ using Markov inequality. □

LEMMA D.3. *Let $a\in(0,2+\iota_0)$. As $n\to\infty$,*

$$
\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|\leq n^{-1/a}}\max_{1\leq t\leq n}\left\|\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right\|=o_P\left(n^{\frac{1}{2+\iota_0}}\right).
$$

PROOF. Note that K is a distribution function and $\|\boldsymbol{Z}_t\|\geq 1$,

$$
\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|\leq n^{-1/a}}\max_{1\leq t\leq n}\left|W_{t,1}(\boldsymbol{\beta},\theta^0)\right|\leq 2=o_P\left(n^{\frac{1}{2+\iota_0}}\right).
$$

For $i \geq 2$, noting that $\|\boldsymbol{Z}_t\| \geq 1$ and $\|\boldsymbol{Z}_t\| \geq |X_{t,i}|$,

$$\begin{aligned}
&\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \leq n^{-1/a}} \max_{1\leq t\leq n} \left|W_{t,i}(\boldsymbol{\beta}, \theta^0)\right| \\
=& \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \leq n^{-1/a}} \max_{1\leq t\leq n} \left|\epsilon_t \frac{X_{t,i-2}}{\|\boldsymbol{Z}_t\|^2} - \left(\boldsymbol{\beta} - \boldsymbol{\beta}^0\right) \frac{\boldsymbol{Z}_t X_{t,i-2}}{\|\boldsymbol{Z}_t\|^2}\right| \\
\leq& \max_{1\leq t\leq n} |\epsilon_t| + n^{-\frac{1}{a}} = o_P\left(n^{\frac{1}{2+\iota_0}}\right),
\end{aligned}$$

where in the last step we use $\max_{1\leq t\leq n} |\epsilon_t| = o_P\left(n^{\frac{1}{2+\iota_0}}\right)$ by Lemma 3 in Owen (1990). Hence,

$$\begin{aligned}
\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \leq n^{-1/a}} \max_{1\leq t\leq n} \left\|\boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0)\right\| \leq& \sum_{i=1}^{k+2} \sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \leq n^{-1/a}} \max_{1\leq t\leq n} \left|W_{t,i}(\boldsymbol{\beta}, \theta^0)\right| \\
=& o_P\left(n^{\frac{1}{2+\iota_0}}\right).
\end{aligned}$$

□

LEMMA D.4. *For all $a > 0$, as $n \to \infty$,*

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \leq n^{-\frac{1}{a}}} \left\|\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0)\boldsymbol{W}_t^T(\boldsymbol{\beta}, \theta^0) - \Sigma_1\right\| \xrightarrow{P} 0.$$

PROOF. Recall from Lemma D.1, as $n \to \infty$,

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\beta}^0, \theta^0)\boldsymbol{W}_t^T(\boldsymbol{\beta}^0, \theta^0) - \Sigma_1 \xrightarrow{P} 0.$$

It suffices to verify that

$$\sup_{\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\| \leq n^{-\frac{1}{a}}} \frac{1}{n}\sum_{t=1}^{n} \left(W_{t,i}(\boldsymbol{\beta}, \theta^0) - W_{t,i}(\boldsymbol{\beta}^0, \theta^0)\right)^2 \xrightarrow{P} 0, \quad i = 1, \ldots, k+2.$$

This is trivial for $i \geq 2$ since uniformly for $\left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \leq n^{-\frac{1}{a}}$

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} \left(W_{t,i}(\boldsymbol{\beta}, \theta^0) - W_{t,i}(\boldsymbol{\beta}^0, \theta^0)\right)^2 \\
=& \frac{1}{n}\sum_{t=1}^{n} \left((\boldsymbol{\beta} - \boldsymbol{\beta}^0)^T \boldsymbol{Z}_t \frac{X_{t,i}}{\|\boldsymbol{Z}_t\|^2}\right)^2 \leq \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\|^2 \xrightarrow{P} 0.
\end{aligned}$$

On the other hand, taking some $r > a$ and noting that $\|\boldsymbol{Z}_t\| \geq 1$,

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left(W_{t,1}(\boldsymbol{\beta},\theta^0) - W_{t,1}(\boldsymbol{\beta}^0,\theta^0)\right)^2\\
\leq&\frac{1}{n}\sum_{t=1}^{n}\left\{K\left(\frac{F_\epsilon^{-1}(\alpha) - \epsilon_t + (\boldsymbol{\beta}-\boldsymbol{\beta}^0)^T(\boldsymbol{Z}_t - \boldsymbol{z})}{h}\right) - K\left(\frac{F_\epsilon^{-1}(\alpha) - \epsilon_t}{h}\right)\right\}^2\\
&\frac{1}{\|\boldsymbol{Z}_t\|^4}I(\|\boldsymbol{Z}_t\| \leq n^{\frac{1}{r}}) + \frac{4}{n}\sum_{t=1}^{n}\frac{1}{\|\boldsymbol{Z}_t\|^4}I(\|\boldsymbol{Z}_t\| > n^{\frac{1}{r}})\\
\leq&\frac{1}{n}\sum_{t=1}^{n}\left\{K\left(\frac{F_\epsilon^{-1}(\alpha) - \epsilon_t + n^{\frac{1}{r}-\frac{1}{a}} + n^{-\frac{1}{a}}\|\boldsymbol{z}\|}{h}\right)\right.\\
&\left.-K\left(\frac{F_\epsilon^{-1}(\alpha) - \epsilon_t - n^{\frac{1}{r}-\frac{1}{a}} - n^{-\frac{1}{a}}\|\boldsymbol{z}\|}{h}\right)\right\}^2 + 4n^{-\frac{4}{r}}\\
\leq&\frac{4}{n}\sum_{t=1}^{n}I\left[-h - n^{\frac{1}{r}-\frac{1}{a}} - n^{-\frac{1}{a}}\|\boldsymbol{z}\| < \epsilon_t - F_\epsilon^{-1}(\alpha) < h + n^{\frac{1}{r}-\frac{1}{a}} + n^{-\frac{1}{a}}\|\boldsymbol{z}\|\right] + 4n^{-\frac{4}{r}}\\
=&\left\{F_\epsilon\left(F_\epsilon^{-1}(\alpha) + h + n^{\frac{1}{r}-\frac{1}{a}} + n^{-\frac{1}{a}}\|\boldsymbol{z}\|\right)\right.\\
&\left.-F_\epsilon\left(F_\epsilon^{-1}(\alpha) - h - n^{\frac{1}{r}-\frac{1}{a}} - n^{-\frac{1}{a}}\|\boldsymbol{z}\|\right)\right\} + o_P(1) = o_P(1)
\end{aligned}
$$

uniformly in $\{\boldsymbol{\beta} : \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \leq n^{-\frac{1}{a}}\}$. □

The following lemma establishes the quadratic expansion of the empirical likelihood ratio function $-2\log L(\cdot,\theta^0)$ around the true parameters $\boldsymbol{\beta}^0$.

LEMMA D.5. *Let $a \in (2, \min\{2 + \iota_0/(3+\iota_0), r_0\})$. For all $\boldsymbol{\beta}$ such that $\left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \leq n^{-1/a}$,*

$$
\begin{aligned}
-2\log L(\boldsymbol{\beta},\theta^0) =&\boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\Sigma_2\boldsymbol{\nu} + 2\boldsymbol{\nu}^T\Sigma_2^T\Sigma_1^{-1}\mathbb{W}_n + \mathbb{W}_n^T\Sigma_1^{-1}\mathbb{W}_n\\
&+ o_P(1) + o_P(\|\boldsymbol{\nu}\|) + o_P(\|\boldsymbol{\nu}\|^2)
\end{aligned}
\tag{D.1}
$$

with $\mathbb{W}_n := \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{W}_t(\boldsymbol{\beta}^0,\theta^0)$ *and* $\boldsymbol{\nu} := \sqrt{n}(\boldsymbol{\beta} - \boldsymbol{\beta}^0)$.

PROOF. All the statements below hold uniformly in $\boldsymbol{\beta} \in \{\boldsymbol{\beta} : \left\|\boldsymbol{\beta} - \boldsymbol{\beta}^0\right\| \leq n^{-1/a}\}$; for presentation convenience, we do not repeat this argument.

Using the Lagrange multiplier method, it is straightforward to show that

$$
-2\log L(\boldsymbol{\beta},\theta^0) = 2\sum_{t=1}^{n}\log\left(1 + \boldsymbol{\lambda}(\boldsymbol{\beta})^T\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right), \tag{D.2}
$$

where the vector $\boldsymbol{\lambda}(\boldsymbol{\beta})$ solves the equations

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)}{1+\boldsymbol{\lambda}(\boldsymbol{\beta})^T\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)}=0;$$

for details, we refer to Owen (2001), Chapter 3. Moreover, with Lemmas D.2 and D.4, similar to the proof of (2.14) in Owen (1990) (see also (A.1) in Qin and Lawless, 1994), we can show that

$$\boldsymbol{\lambda}(\boldsymbol{\beta})=O_P(n^{-1/a}), \tag{D.3}$$

and, in conjunction with Lemma D.3, we may write

$$\begin{aligned}\boldsymbol{\lambda}(\boldsymbol{\beta}) =&\Sigma_1^{-1}\left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{W}_t(\boldsymbol{\beta}^0,\theta^0)+\Sigma_2(\boldsymbol{\beta}-\boldsymbol{\beta}^0)\right)+o_P(n^{-1/2})+o_P(\|\boldsymbol{\beta}-\boldsymbol{\beta}^0\|)\\ =&n^{-1/2}\Sigma_1^{-1}(\mathbb{W}_n+\Sigma_2\boldsymbol{\nu})+o_P(n^{-1/2})+o_P(n^{-1/2}\|\boldsymbol{\nu}\|)\end{aligned} \tag{D.4}$$

as in the proof of (2.17) in Owen (1990) by noting that

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^{n}\left\|\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right\|^3\|\boldsymbol{\lambda}(\boldsymbol{\beta})\|^2\frac{1}{|1-\boldsymbol{\lambda}(\boldsymbol{\beta})^T\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)|}\\ \le&\max_{1\le t\le n}\left\|\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right\|\cdot\frac{1}{n}\sum_{t=1}^{n}\left\|\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right\|^2\|\boldsymbol{\lambda}(\boldsymbol{\beta})\|^2\cdot\max_{1\le t\le n}\frac{1}{|1-\boldsymbol{\lambda}(\boldsymbol{\beta})^T\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)|}\\ =&o_P(n^{1/(2+\iota_0)})\cdot O_P(1)\cdot O_P(n^{-2/a})\cdot O_P(1)=o_P(n^{-1/2}).\end{aligned}$$

Recalling (D.2) we can expand

$$\begin{aligned}&-2\log L(\boldsymbol{\beta},\theta)\\ =&2n\boldsymbol{\lambda}(\boldsymbol{\beta})^T\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)-n\cdot\frac{1}{n}\sum_{t=1}^{n}\left(\boldsymbol{\lambda}(\boldsymbol{\beta})^T\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right)^2\\ &+\frac{2n}{3}\frac{1}{n}\sum_{t=1}^{n}\frac{(\boldsymbol{\lambda}(\boldsymbol{\beta})^T\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0))^3}{(1+\delta_t)}=:S_1-S_2+S_3\end{aligned}$$

where, recalling (D.3) and Lemma D.3,

$$\max_{1\le t\le n}|\delta_t|\le\|\boldsymbol{\lambda}(\boldsymbol{\beta})\|\max_{1\le t\le n}\left\|\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right\|=O_P(n^{-1/a})\cdot o_P(n^{1/(2+\iota_0)})=o_P(1)$$

and therefore, in conjunction with Lemma D.4, S_3 has a norm bounded by

$$n\|\boldsymbol{\lambda}(\boldsymbol{\beta})\|^3\cdot\max_{1\le t\le n}\left\|\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right\|\cdot\frac{1}{n}\sum_{t=1}^{n}\left\|\boldsymbol{W}_t(\boldsymbol{\beta},\theta^0)\right\|^2\max_{1\le t\le n}\frac{1}{|1+\delta_t|}$$

$$=n \cdot O_P(n^{-3/a}) \cdot o_P(n^{\frac{1}{2+\iota_0}}) \cdot O_P(1) \cdot O_P(1) = o_P(1).$$

Write

$$\begin{aligned} S_1 =& 2n^{1/2}\boldsymbol{\lambda}(\boldsymbol{\beta})^T \mathbb{W}_n + 2n\boldsymbol{\lambda}(\boldsymbol{\beta})^T \left\{ \frac{1}{n}\sum_{t=1}^n \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0) - \frac{1}{n}\sum_{t=1}^n \boldsymbol{W}_t(\boldsymbol{\beta}^0, \theta^0) \right\} \\ =&: S_{11} + S_{12}. \end{aligned}$$

Substituting (D.4) in above equation yields that

$$S_{11} = 2\mathbb{W}_n^T \Sigma_1^{-1} \mathbb{W}_n + 2\boldsymbol{\nu}^T \Sigma_2^T \Sigma_1^{-1} \mathbb{W}_n + o_P(1) + o_P(\|\boldsymbol{\nu}\|)$$

and, together with Lemma D.2 and Taylor expansion,

$$\begin{aligned} S_{12} =& 2n\boldsymbol{\lambda}(\boldsymbol{\beta})^T (\Sigma_2 + o_p(1))(\boldsymbol{\beta} - \boldsymbol{\beta}^0) \\ =& 2\boldsymbol{\nu}^T \Sigma_2^T \Sigma_1^{-1} \mathbb{W}_n + 2\boldsymbol{\nu}^T \Sigma_2^T \Sigma_1^{-1} \Sigma_2 \boldsymbol{\nu} + o_P(\|\boldsymbol{\nu}\|) + o_P(\|\boldsymbol{\nu}\|^2). \end{aligned}$$

Similarly, substituting (D.4) and using Lemma D.4 yields that

$$\begin{aligned} S_2 =& \boldsymbol{\lambda}(\boldsymbol{\beta})^T \left[\frac{1}{n}\sum_{t=1}^n \boldsymbol{W}_t(\boldsymbol{\beta}, \theta^0) \boldsymbol{W}_t^T(\boldsymbol{\beta}, \theta^0) \right] \boldsymbol{\lambda}(\boldsymbol{\beta}) \\ =& \boldsymbol{\lambda}(\boldsymbol{\beta})^T (\Sigma_1 + o_P(1)) \boldsymbol{\lambda}(\boldsymbol{\beta}) \\ =& \mathbb{W}_n^T \Sigma_1^{-1} \mathbb{W}_n + 2\boldsymbol{\nu}^T \Sigma_2^T \Sigma_1^{-1} \mathbb{W}_n + \boldsymbol{\nu}^T \Sigma_2^T \Sigma_1^{-1} \Sigma_2 \boldsymbol{\nu} \\ & + o_P(1) + o_P(\|\boldsymbol{\nu}\|) + o_P(\|\boldsymbol{\nu}\|^2). \end{aligned}$$

The lemma follows by collecting the substitutions above. □

LEMMA D.6. *Let $a \in (2, \min\{2 + \iota_0/(3 + \iota_0), r_0\})$. With probability tending to one, $L(\boldsymbol{\beta}, \theta^0)$ attains its maximum at some point $\widetilde{\boldsymbol{\beta}}$ in the interior of the ball $||\boldsymbol{\beta} - \boldsymbol{\beta}_0|| \le n^{-1/a}$ and, furthermore,*

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) = -(\Sigma_2^T \Sigma_1^{-1} \Sigma_2)^{-1} \Sigma_2^T \Sigma_1^{-1} \mathbb{W}_n + o_P(1), \tag{D.5}$$

where $\mathbb{W}_n$ is defined in Lemma D.5.

PROOF. The existence of $\widetilde{\boldsymbol{\beta}}$ can be proved similarly as Lemma 1 in Qin and Lawless (1994) using Lemmas D.1, D.2, and D.4 here, and therefore omitted. The proof of the second part is a (slight) modification of that of Theorem 2 in Sherman (1993).

Define $\boldsymbol{\nu}^* := -(\Sigma_2^T \Sigma_1^{-1} \Sigma_2)^{-1} \Sigma_2^T \Sigma_1^{-1} \mathbb{W}_n$, which is the minimum point of ν for the non-negligible (quadratic) part of the empirical likelihood ratio function in the expansion (D.1). By Lemma D.1, we have $\boldsymbol{\nu}^* = O_P(1)$, therefore $\left\|n^{-1/2}\boldsymbol{\nu}^*\right\| \leq n^{-1/a}$ with probability tending to 1. It follows that, with probability tending to 1, by definition we have

$$-2\log L(\widetilde{\boldsymbol{\beta}}, \theta^0) \leq -2\log L(\boldsymbol{\beta}^0 + n^{-1/2}\boldsymbol{\nu}^*, \theta^0).$$

Applying (D.1) twice in the last expression, consolidating terms, and using the facts that $\Sigma_2^T\Sigma_1^{-1}\Sigma_2$ is positive definite (by Condition C3) and $\boldsymbol{\nu}^* = O_P(1)$, to get

$$\begin{aligned} 0 \leq & \left(\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) - \boldsymbol{\nu}^*\right)^T \Sigma_2^T\Sigma_1^{-1}\Sigma_2 \left(\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) - \boldsymbol{\nu}^*\right) \\ \leq & o_P\left(\left\|\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) - \boldsymbol{\nu}^*\right\|^2\right). \end{aligned}$$

Hence, $\sqrt{n}(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) = \boldsymbol{\nu}^* + o_P(1)$ and the lemma follows. □

PROOF OF THEOREM 3. Recall the maximum empirical likelihood estimator $\widetilde{\boldsymbol{\beta}}$ from Lemma D.6. Substituting (D.5) into (D.1) in Lemma D.5 yields that

$$-2\log L^P(\theta^0) = -2\log L(\widetilde{\boldsymbol{\beta}}, \theta^0) = \mathbb{W}_n^T\Sigma_1^{-1/2}D\Sigma_1^{-1/2}\mathbb{W}_n + o_P(1)$$

where

$$D := I_{k+2} - \Sigma_1^{-1/2}\Sigma_2(\Sigma_2^T\Sigma_1^{-1}\Sigma_2)^{-1}\Sigma_2^T\Sigma_1^{-1/2}$$

and I_{k+2} denotes the identity $(k+2)\times(k+2)$ matrix. Hence the theorem follows from the facts that D is symmetric, idempotent,

$$\begin{aligned} & tr(I_{k+2} - \Sigma_1^{-1/2}\Sigma_2(\Sigma_2^T\Sigma_1^{-1}\Sigma_2)^{-1}\Sigma_2^T\Sigma_1^{-1/2}) \\ = & k+2 - tr((\Sigma_2^T\Sigma_1^{-1}\Sigma_2)^{-1}\Sigma_2^T\Sigma_1^{-1/2}\Sigma_1^{-1/2}\Sigma_2) \\ = & k+2 - tr((\Sigma_2^T\Sigma_1^{-1}\Sigma_2)^{-1}\Sigma_2^T\Sigma_1^{-1}\Sigma_2) \\ = & k+2-(k+1) = 1 \end{aligned}$$

and $\Sigma_1^{-1/2}\mathbb{W}_n \xrightarrow{d} N(0, I_{k+2})$ by Lemma D.1. □

APPENDIX E: EMPIRICAL LIKELIHOOD METHOD FOR QUANTILE REGRESSION MODEL

We can apply the same weighted idea to generalize the empirical likelihood method for quantile regression in Whang (2006) by allowing infinite variance predictors and constructing a confidence interval for the conditional VaR rather than a confidence region for the regression coefficient studied in Whang (2006). Define parameter $\boldsymbol{\varphi}$ for $(\beta_1, \ldots, \beta_k)^T$, the regression coefficients for predictors. We consider a smoothed but weighted estimating function for

$$\boldsymbol{G}_t(\boldsymbol{\varphi}, \theta) = \frac{\psi_t(\boldsymbol{\varphi}, \theta)}{\|\boldsymbol{Z}_t\|^2} \boldsymbol{Z}_t = \frac{\psi_t(\boldsymbol{\varphi}, \theta)}{\|\boldsymbol{Z}_t\|^2} (1, X_{t,1}, \ldots, X_{t,k})^T$$

with

$$\psi_t(\boldsymbol{\varphi}, \theta) = K\left(\frac{\theta - \boldsymbol{\varphi}^T \boldsymbol{x} - (Y_t - \boldsymbol{\varphi}^T \boldsymbol{X}_t)}{h}\right) - \alpha.$$

Again, the weight $\|\boldsymbol{Z}_t\|^{-2}$ is to remove the effect of infinite moments of $\boldsymbol{X}_t$. We now can define the empirical likelihood function for $(\boldsymbol{\varphi}^T, \theta)^T \in \mathbb{R}^{k+1}$ as

$$L_Q(\boldsymbol{\varphi}, \theta) = \sup\left\{\prod_{t=1}^n (np_t) : p_1 \geq 0, \ldots, p_n \geq 0, \sum_{t=1}^n p_t = 1, \sum_{t=1}^n p_t \boldsymbol{G}_t(\boldsymbol{\varphi}, \theta) = 0\right\}.$$

Profiling out the nuisance parameter $\boldsymbol{\varphi}$ yields the profile empirical likelihood function

$$L_Q^P(\theta) = \max_{\boldsymbol{\varphi}} L_Q(\boldsymbol{\varphi}, \theta).$$

We substitute the conditions C1-C3, with slight re-parametrization, in the setting of quantile regression as follows. Denote the conditional distribution function of ϵ_t given $\boldsymbol{X}_t = \boldsymbol{x}$ as $F_\epsilon(u|\boldsymbol{x}) = P(\epsilon_t \leq u | \boldsymbol{X}_t = \boldsymbol{x})$.

C1' $\{(\epsilon_t, \boldsymbol{X}_t)\}$ is a stationary sequence, satisfying $F_\epsilon(0|\boldsymbol{X}_t) = \alpha$; ϵ_t is conditionally independent of $\{\epsilon_1, \ldots, \epsilon_{t-1}, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_{t-1}\}$ given $\boldsymbol{X}_t$. There exists a neighborhood of zero and a constant $C_0 \in (0, \infty)$ such that $F'(u|\boldsymbol{X}_t) \leq C_0$ almost surely for all u in this neighborhood. For all u_1 and u_2 in this neighborhood, $|F'(u_1|\boldsymbol{x}) - F'(u_2|\boldsymbol{x})| \leq C(\boldsymbol{x})\,|u_1 - u_2|$, with some function $C(\boldsymbol{x})$ such that $E\left(C(\boldsymbol{X}_t)\right) < \infty$.

C2' The stationary sequence $\{\boldsymbol{X}_t\}$ is ergodic such that, as $n \to \infty$

$$\frac{1}{n}\sum_{t=1}^n \frac{\boldsymbol{Z}_t \boldsymbol{Z}_t^T}{\|\boldsymbol{Z}_t\|^4} \xrightarrow{P} E\left(\frac{\boldsymbol{Z}_1 \boldsymbol{Z}_1^T}{\|\boldsymbol{Z}_1\|^4}\right),$$

$$\frac{1}{n}\sum_{t=1}^n F_\varepsilon'(0|\boldsymbol{X}_t)\frac{\boldsymbol{Z}_t \boldsymbol{Z}_t^T}{\|\boldsymbol{Z}_t\|^2} \xrightarrow{P} E\left(F_\varepsilon'(0|\boldsymbol{X}_1)\frac{\boldsymbol{Z}_1 \boldsymbol{Z}_1^T}{\|\boldsymbol{Z}_1\|^2}\right).$$

C3' The matrix $\widetilde{\Sigma}_1 := \alpha(1-\alpha)E\left(\frac{\boldsymbol{Z}_1 \boldsymbol{Z}_1^T}{\|\boldsymbol{Z}_1\|^4}\right)$ is positive definite, and the matrix $\widetilde{\Sigma}_2 = E\left(F_\varepsilon'(0|\boldsymbol{X}_1)\frac{\boldsymbol{Z}_1(\boldsymbol{X}_1 - \boldsymbol{x})^T}{\|\boldsymbol{Z}_1\|^2}\right)$ has full rank.

THEOREM E.1. *Under conditions* C1'-C3' *and* C4 *in the main document,* $-2\log L_Q^P(\mathrm{VaR}_{\boldsymbol{x}}(\alpha))$ *converges in distribution to* $\chi^2(1)$*, a chi-squared limit with one degree of freedom as* $n\to\infty$ *if provided* $E\left(\|\boldsymbol{X}_1\|\right)<\infty$.

Our proof for Theorem E.1 is similar to that for Theorem 3 after substituting the moment conditions. However, establishing the uniform convergences of sample smoothed moments and gradients requires different techniques. In particular, the quantile regression method requires extra moment condition on $\|\boldsymbol{X}_t\|$ because of the nonlinearity of score functions for the nuisance parameter $\boldsymbol{\varphi}$.

Let $\boldsymbol{\varphi}^0$ denote the true values of regression coefficients. For simplicity, we assume the conditions of Theorem E.1 throughout this subsection. We need some lemmas first.

LEMMA E.1. *As* $n\to\infty$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{G}_t(\boldsymbol{\varphi}^0,\theta^0)\xrightarrow{d} N\left(0,\widetilde{\Sigma}_1\right),\ \frac{1}{n}\sum_{t=1}^n \boldsymbol{G}_t(\boldsymbol{\varphi}^0,\theta^0)\boldsymbol{G}_t^T(\boldsymbol{\varphi}^0,\theta^0)\xrightarrow{P}\widetilde{\Sigma}_1,\ \textit{and}$$

$$\frac{1}{n}\sum_{t=1}^n \frac{\partial \boldsymbol{G}_t(\boldsymbol{\varphi}^0,\theta^0)}{\partial\boldsymbol{\varphi}}\xrightarrow{P}\widetilde{\Sigma}_2.$$

PROOF. Again define the filtration $\{\mathcal{F}_t\}_{t\geq 0}=\{\sigma(\epsilon_1,\ldots,\epsilon_t,\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_{t+1})\}_{t\geq 0}$. By conditions C1' and C4, it is straightforward to verify that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^n\left|E\left(\boldsymbol{G}_t(\boldsymbol{\varphi}^0,\theta^0)|\mathcal{F}_{t-1}\right)\right| =&O_P(n^{\frac12}h^2)\cdot\frac{1}{n}\sum_{t=1}^n\frac{C(\boldsymbol{X}_t)\boldsymbol{Z}_t}{\|\boldsymbol{Z}_t\|^2}\\ =&o_P(1)\cdot\frac{1}{n}\sum_{t=1}^n C(\boldsymbol{X}_t)=o_P(1),\end{aligned}$$

$$\begin{aligned}&\frac{1}{n}\sum_{t=1}^n E\left(\boldsymbol{G}_t(\boldsymbol{\varphi}^0,\theta^0)\boldsymbol{G}_t^T(\boldsymbol{\varphi}^0,\theta^0)^T|\mathcal{F}_{t-1}\right)\\ =&\alpha(1-\alpha)\cdot\frac{1}{n}\sum_{t=1}^n\frac{\boldsymbol{Z}_t\boldsymbol{Z}_t^T}{\|\boldsymbol{Z}_t\|^4}+O_P(h)\cdot\frac{1}{n}\sum_{t=1}^n\frac{C(\boldsymbol{X}_t)\boldsymbol{Z}_t\boldsymbol{Z}_t^T}{\|\boldsymbol{Z}_t\|^4}\\ =&\widetilde{\Sigma}_1+o_P(1)+O_P(h)\cdot\frac{1}{n}\sum_{t=1}^n C(\boldsymbol{X}_t)=\widetilde{\Sigma}_1+o_P(1),\end{aligned}$$

and

$$\frac{1}{n}\sum_{t=1}^n E\left(\frac{\partial\boldsymbol{G}_t(\boldsymbol{\varphi}^0,\theta^0)}{\partial\boldsymbol{\varphi}}|\mathcal{F}_{t-1}\right)$$

$$
\begin{aligned}
=&\frac{1}{n}\sum_{t=1}^{n} F_{\epsilon}'(0|\boldsymbol{X}_t)\frac{\boldsymbol{Z}_t\left(\boldsymbol{X}_t-\boldsymbol{x}\right)^T}{\|\boldsymbol{Z}_t\|^2}+O_P(h)\cdot\frac{1}{n}\sum_{t=1}^{n} C(\boldsymbol{X}_t)\frac{\boldsymbol{Z}_t\left(\boldsymbol{X}_t-\boldsymbol{x}\right)^T}{\|\boldsymbol{Z}_t\|^2}\\
=&\widetilde{\Sigma}_2+o_P(1)+O_P(h)\cdot\frac{1}{n}\sum_{t=1}^{n} C(\boldsymbol{X}_t)=\widetilde{\Sigma}_2+o_P(1).
\end{aligned}
$$

The rest follows by the weak law of large numbers and the central limit theorem for martingale differences array. □

LEMMA E.2. *For all* $a\in(0,r_0)$, *as* $n\to\infty$,

$$
\sup_{\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\|\le n^{-\frac{1}{a}}}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial \boldsymbol{G}_t(\boldsymbol{\varphi},\theta^0)}{\partial\boldsymbol{\varphi}}-\widetilde{\Sigma}_2\right\|\xrightarrow{P}0.
$$

PROOF. By Lemma E.1 and a direct calculation, it suffices to show that

$$
\sup_{\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\|\le n^{-\frac{1}{a}}}\frac{1}{n}\sum_{t=1}^{n}\left|\frac{1}{h}g\left(\frac{-\epsilon_t-(\boldsymbol{\varphi}-\boldsymbol{\varphi}^0)^T(\boldsymbol{X}_t-\boldsymbol{x})}{h}\right)-\frac{1}{h}g\left(\frac{-\epsilon_t}{h}\right)\right|\xrightarrow{P}0.
$$

All the statements below hold uniformly for $\boldsymbol{\varphi}$ such that $\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\right\|\le n^{-1/a}$; for presentation convenience, we do not repeat this argument. Note that

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left|\frac{1}{h}g\left(\frac{-\epsilon_t+(\boldsymbol{\varphi}-\boldsymbol{\varphi}^0)^T(\boldsymbol{X}_t-\boldsymbol{x})}{h}\right)-\frac{1}{h}g\left(\frac{-\epsilon_t}{h}\right)\right|\\
\le&\frac{1}{n}\sum_{t=1}^{n}\left|\frac{1}{h}g\left(\frac{-\epsilon_t+(\boldsymbol{\varphi}-\boldsymbol{\varphi}^0)^T(\boldsymbol{X}_t-\boldsymbol{x})}{h}\right)-\frac{1}{h}g\left(\frac{-\epsilon_t}{h}\right)\right|I\left(|\epsilon_t|\le h\right)\\
&+\frac{1}{n}\sum_{t=1}^{n}\left|\frac{1}{h}g\left(\frac{-\epsilon_t+(\boldsymbol{\varphi}-\boldsymbol{\varphi}^0)^T(\boldsymbol{X}_t-\boldsymbol{x})}{h}\right)\right|I\left(|\epsilon_t|>h\right)=:T_1+T_2.
\end{aligned}
$$

Recall that g' is bounded. We apply Taylor expansion to get, for some large $M>0$,

$$
\begin{aligned}
T_1\le&\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\right\|\frac{M}{nh^2}\sum_{t=1}^{n}\left(\|\boldsymbol{X}_t\|+\|\boldsymbol{x}\|\right)I\left(|\epsilon_t|\le h\right)\\
\le&\frac{M}{n^{1/a}h}\cdot\frac{1}{nh}\sum_{t=1}^{n}\|\boldsymbol{X}_t\|\,I\left(|\epsilon_t|\le h\right)+\frac{M\,\|\boldsymbol{x}\|}{n^{1/a}h}\cdot\frac{1}{nh}\sum_{t=1}^{n}I\left(|\epsilon_t|\le h\right)\xrightarrow{P}0,
\end{aligned}
$$

where the convergence can be shown by Markov inequality using $n^{1/a}h\to\infty$ and for large n

$$
E\left(\frac{1}{nh}\sum_{t=1}^{n}\|\boldsymbol{X}_t\|\,I\left(|\epsilon_t|\le h\right)\right)=E\left(\frac{P(|\epsilon_1|\le h)}{h}\,\|\boldsymbol{X}_1\|\right)\le C_0E\left(\|\boldsymbol{X}_1\|\right)<\infty,
$$

$$E\left(\frac{1}{nh}\sum_{t=1}^{n} I\left(|\epsilon_t| \le h\right)\right) = E\left(\frac{P(|\epsilon_1| \le h)}{h}\right) \le C_0.$$

By definition, g is bounded and has only support on $[-1,1]$. For some large $M>0$, we have

$$\begin{aligned} T_2 \le & \frac{M}{nh}\sum_{t=1}^{n} I\left[0 \le \epsilon_t - h \le \left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\right\| \|\boldsymbol{X}_t\| + \left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\right\| \|\boldsymbol{x}\|\right] \\ & + \frac{M}{nh}\sum_{t=1}^{n} I\left[0 \ge \epsilon_t + h \ge -\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\right\| \|\boldsymbol{X}_t\| - \left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\right\| \|\boldsymbol{x}\|\right] \\ =: & T_{21} + T_{22}. \end{aligned}$$

To complete the proof, we only prove $T_{21} = o_P(1)$; the proof of $T_{22} = o_P(1)$ is completely analogous and omitted. It is easy to verify that

$$\begin{aligned} T_{21} \le & \frac{1}{nh}\sum_{t=1}^{n} I\left(0 \le \epsilon_t - h \le n^{-1/a}\|\boldsymbol{X}_t\| + n^{-1/a}\|\boldsymbol{x}\|\right) I\left[\|\boldsymbol{X}_t\| \le n^{1/r_0}\right] \\ & + \frac{1}{nh}\sum_{t=1}^{n} I\left[\|\boldsymbol{X}_t\| > n^{1/r_0}\right] = T_{21,1} + T_{22,2}. \end{aligned}$$

Note that $\max_{1\le t\le n} n^{-1/a}\|\boldsymbol{X}_t\| I\left[\|\boldsymbol{X}_t\| \le n^{1/r_0}\right] \le n^{-1/a}\cdot n^{1/r_0} \to 0$. For large n, using Taylor expansion, we can show that

$$E\left(T_{21,1}\right) \le \frac{C_0}{n^{1/a}h}\left(E\left(\|\boldsymbol{X}_t\|\right) + \|\boldsymbol{x}\|\right) \to 0,$$

and so it follows that $T_{21,1} \xrightarrow{P} 0$. Similarly, it remains to verify that

$$E\left(T_{21,2}\right) \le \frac{1}{n^{1/r_0}h}\cdot n^{1/r_0}P(\|\boldsymbol{X}_t\| > n^{1/r_0}) \to 0\cdot 0 = 0,$$

where the convergence of the second term holds because $E(\|\boldsymbol{X}_t\|) < \infty$. □

LEMMA E.3. *For all $a>0$, as $n\to\infty$,*

$$\sup_{\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\|\le n^{-1/a}}\left\|\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{G}_t(\boldsymbol{\varphi},\theta^0)\boldsymbol{G}_t^T(\boldsymbol{\varphi},\theta^0) - \widetilde{\Sigma}_1\right\| = o_P(1).$$

PROOF. By Lemma E.1 and a direct calculation, it suffices to show that

$$\sup_{\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\|\le n^{-1/a}}\frac{1}{n}\sum_{t=1}^{n}\left\{K\left(\frac{-\epsilon_t + (\boldsymbol{\varphi}-\boldsymbol{\varphi}^0)^T(\boldsymbol{X}_t - \boldsymbol{x})}{h}\right) - K\left(\frac{-\epsilon_t}{h}\right)\right\}^2\frac{1}{\|\boldsymbol{Z}_t\|^2} = o_P(1).$$

All the statements below hold uniformly for $\boldsymbol{\varphi}$ such that $\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\right\| \leq n^{-1/a}$; for presentation convenience we do not repeat this argument. Take some $r > a$, and we have

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left\{K\left(\frac{-\epsilon_t+(\boldsymbol{\varphi}-\boldsymbol{\varphi}^0)^T(\boldsymbol{X}_t-\boldsymbol{x})}{h}\right)-K\left(\frac{-\epsilon_t}{h}\right)\right\}^2\frac{1}{\|\boldsymbol{Z}_t\|^2}\\
\leq&\frac{1}{n}\sum_{t=1}^{n}\left\{K\left(\frac{-\epsilon_t+n^{-1/a+1/r}+n^{-1/a}\|\boldsymbol{z}\|}{h}\right)\right.\\
&\left.-K\left(\frac{-\epsilon_t-n^{-1/a+1/r}-n^{-1/a}\|\boldsymbol{z}\|}{h}\right)\right\}^2 I\left(\|\boldsymbol{Z}_t\|\leq n^{1/r}\right)\\
&+\frac{4}{n}\sum_{t=1}^{n}\frac{1}{\|\boldsymbol{Z}_t\|^2}I\left(\|\boldsymbol{Z}_t\|>n^{1/r}\right)\\
\leq&\frac{4}{n}\sum_{t=1}^{n}I\left[|\epsilon_t|<n^{-1/a+1/r}+n^{-1/a}\|\boldsymbol{z}\|\right]+4n^{-\frac{2}{r}}\xrightarrow{P}0,
\end{aligned}$$

where the convergence is straightforward by Markov inequality. □

Using Lemmas E.2, E.3 and noting $\boldsymbol{G}_t$'s are bounded, we give the quadratic expression of the empirical likelihood ratio function $-2\log L_Q(\cdot,\theta^0)$ around the true parameters $\boldsymbol{\varphi}^0$ in the following lemma. The proof is analogous to that of Lemma D.5 and that of Lemma D.6, and therefore we omit the details.

Lemma E.4. *Let $a \in (2, r_0)$. For all $\boldsymbol{\varphi}$ such that $\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\right\| \leq n^{-1/a}$,*

$$\begin{aligned}
-2\log L_Q(\boldsymbol{\varphi},\theta^0)=&\boldsymbol{\nu}^T\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1}\widetilde{\Sigma}_2\boldsymbol{\nu}+2\boldsymbol{\nu}^T\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1}\mathbb{G}_n+\mathbb{G}_n^T\widetilde{\Sigma}_1^{-1}\mathbb{G}_n\\
&+o_P(1)+o_P(\|\boldsymbol{\nu}\|)+o_P(\|\boldsymbol{\nu}\|^2)
\end{aligned} \tag{E.1}$$

with $\mathbb{G}_n := \frac{1}{\sqrt{n}}\sum_{t=1}^n \boldsymbol{G}_t(\boldsymbol{\varphi}^0,\theta^0)$ and $\boldsymbol{\nu} := \sqrt{n}(\boldsymbol{\varphi}-\boldsymbol{\varphi}^0)$. With probability tending to one, $L_Q(\boldsymbol{\varphi},\theta^0)$ attains its maximum at some point $\widetilde{\boldsymbol{\varphi}}$ in the interior of the ball $\left\|\boldsymbol{\varphi}-\boldsymbol{\varphi}^0\right\| \leq n^{-1/a}$ and

$$\sqrt{n}(\widetilde{\boldsymbol{\varphi}}-\boldsymbol{\varphi}^0) = -(\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1}\widetilde{\Sigma}_2)^{-1}\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1}\mathbb{G}_n+o_P(1). \tag{E.2}$$

Proof of Theorem E.1. Substituting (E.2) into (E.1) in Lemma E.4 yields that

$$-2\log L_Q^P(\theta^0) = -2\log L_Q(\widetilde{\boldsymbol{\varphi}},\theta^0) = \mathbb{G}_n^T\widetilde{\Sigma}_1^{-1/2}\widetilde{D}\widetilde{\Sigma}_1^{-1/2}\mathbb{G}_n + o_P(1)$$

where

$$\widetilde{D} := I_{k+1} - \widetilde{\Sigma}_1^{-1/2}\widetilde{\Sigma}_2(\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1}\widetilde{\Sigma}_2)^{-1}\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1/2}.$$

Hence the theorem follows from the facts that $\widetilde{D}$ is symmetric, idempotent,

$$\begin{aligned} & tr(I_{k+1} - \widetilde{\Sigma}_1^{-1/2}\widetilde{\Sigma}_2(\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1}\widetilde{\Sigma}_2)^{-1}\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1/2}) \\ = \; & k+1 - tr((\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1}\widetilde{\Sigma}_2)^{-1}\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1/2}\widetilde{\Sigma}_1^{-1/2}\widetilde{\Sigma}_2) \\ = \; & k+1 - tr((\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1}\widetilde{\Sigma}_2)^{-1}\widetilde{\Sigma}_2^T\widetilde{\Sigma}_1^{-1}\widetilde{\Sigma}_2) = k+1-k = 1 \end{aligned}$$

and $\widetilde{\Sigma}_1^{-1/2}\mathbb{G}_n \xrightarrow{d} N(0, I_{k+1})$ by Lemma E.1. □

APPENDIX F: PROOFS FROM SECTION 4

We denote the nuisance parameters by $\boldsymbol{\upsilon} = (\boldsymbol{\zeta}^T, \boldsymbol{\beta}^T)^T$ with true values $\boldsymbol{\upsilon}^0 = (\boldsymbol{\zeta}^{0T}, \boldsymbol{\beta}^{0T})^T$. Define a local parameter space $\Theta_{\boldsymbol{\upsilon}}^a = \left\{\boldsymbol{\upsilon} : \left\|\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^0\right\| \le n^{-1/a}\right\}$ for all $a > 0$. Recall that $\xi_{\rho,t} = 1 + \sum_{j=0}^{\infty} \rho^j \left\|\boldsymbol{X}_{t-j}\right\|$ and $\widetilde{\xi}_{\rho,t} = 1 + \sum_{j=0}^{\infty} \rho^j |\epsilon_{t-j}|$ for $\rho \in (0,1)$. Unless specified otherwise, we allow the initial values $\{(Y_s, \boldsymbol{X}_s, h_s)\}$, $s \le 0$ to be substituted by a known constant independent of the parameters. We denote the *unobserved* parametric models with *true* initial values $\{(Y_s, \boldsymbol{X}_s)\}_{s\le 0}$ by $\widetilde{h}_t(\boldsymbol{\upsilon})$ and $\tilde{\eta}_t(\boldsymbol{\upsilon}) = \epsilon_t(\boldsymbol{\beta})/\sqrt{\tilde{h}_t(\boldsymbol{\upsilon})}$ with log-likelihood function $\widetilde{l}_t(\boldsymbol{\zeta}, \boldsymbol{\beta}) = -\frac{1}{2}\log \tilde{h}_t(\boldsymbol{\zeta}, \boldsymbol{\beta}) - \frac{\epsilon_t^2(\boldsymbol{\beta})}{2\tilde{h}_t(\boldsymbol{\zeta}, \boldsymbol{\beta})}$ for all $t \in \mathbb{Z}$. The constant $C > 0$ and $\rho \in (0,1)$ may be different in different inequalities below, but they are independent of t and n, and the inequalities hold uniformly for $t = 1, \ldots, n$. For simplicity, we assume all the conditions of Theorem 4 throughout this section.

LEMMA F.1. *There exists a constant $\rho \in (0,1)$ such that*

(i) $\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a} \left\| \frac{1}{h_t(\boldsymbol{\upsilon})} \frac{\partial h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\zeta}} \right\| \le C\left(\xi_{\rho,t-1}^{\iota} + \widetilde{\xi}_{\rho,t-1}^{\iota}\right)$,

(ii) $\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a} \left\| \frac{1}{h_t(\boldsymbol{\upsilon})} \frac{\partial^2 h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\zeta}^T} \right\| \le C\left(\xi_{\rho,t-1}^{\iota} + \widetilde{\xi}_{\rho,t-1}^{\iota}\right)$,

(iii) $\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a} \left\| \frac{1}{\sqrt{h_t(\boldsymbol{\upsilon})}} \frac{\partial h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\beta}} \right\| \le C\xi_{\rho,t-1}$,

(iv) $\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a} \left\| \frac{1}{\sqrt{h_t(\boldsymbol{\upsilon})}} \frac{\partial^2 h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\zeta}\partial \boldsymbol{\beta}^T} \right\| \le C\xi_{\rho,t-1}$,

(v) $\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a} \left\| \frac{1}{\sqrt{h_t(\boldsymbol{\upsilon})}} \frac{\partial^2 h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}^T} \right\| \le C\xi_{\rho,t-1}^2$

for any $\iota \in (0,1)$, where C is a constant independent of ι.

Proof. Note that $|\epsilon_t(\gamma)| \leq C(|\epsilon_t| + \|X_t\|)$ for large constant $C > 0$. Following the proof of Lemma A.2 in Ling (2007) we have

$$\left\| \frac{1}{h_t(\boldsymbol{\upsilon})} \frac{\partial h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\zeta}} \right\| \leq C \left(\xi_{\rho,t-1} + \widetilde{\xi}_{\rho,t-1} \right)^{\iota}$$

and

$$\left\| \frac{1}{h_t(\boldsymbol{\upsilon})} \frac{\partial^2 h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\zeta}^T} \right\| \leq C \left(\xi_{\rho,t-1} + \widetilde{\xi}_{\rho,t-1} \right)^{\iota}.$$

Statements (i) and (ii) follow because $\left(\xi_{\rho,t-1} + \widetilde{\xi}_{\rho,t-1} \right)^{\iota} \leq \xi^{\iota}_{\rho,t-1} + \widetilde{\xi}^{\iota}_{\rho,t-1}$. Moreover, using the fact that $\frac{\partial \epsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\boldsymbol{Z}_t$ and following the proof of Lemma A.3 in Ling (2007),

$$\sup_{\boldsymbol{\upsilon} \in \Theta^a_{\boldsymbol{\upsilon}}} \left\| \frac{1}{\sqrt{h_t(\boldsymbol{\upsilon})}} \frac{\partial h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\beta}} \right\| \leq C \left(1 + \sum_{j=0}^{\infty} \rho^j \|\boldsymbol{Z}_t\| \right) \leq C \xi_{\rho,t-1},$$

i.e., statement (iii) holds. Similarly, we can prove statements (iv) and (v). □

From now on we take ρ as that in Lemma F.1.

Lemma F.2. *For all small $\iota \in (0, \iota_0/a)$,*

$$\sup_{\boldsymbol{\upsilon} \in \Theta^a_{\boldsymbol{\upsilon}}} \left| \sqrt{\frac{h_t}{h_t(\boldsymbol{\upsilon})}} - 1 \right| \leq C n^{-1/a} (\xi_{\rho,t-1} + \widetilde{\xi}^{\iota}_{\rho,t-1})$$

where the constant C is independent of t and n.

Proof. By Taylor expansion and Lemma F.1,

$$\begin{aligned} \max_{1 \leq t \leq n} \sup_{\boldsymbol{\upsilon} \in \Theta^a_{\boldsymbol{\upsilon}}} \left| \log \frac{h_t(\boldsymbol{\zeta}, \boldsymbol{\beta})}{h_t(\boldsymbol{\zeta}^0, \boldsymbol{\beta})} \right| &\leq n^{-1/a} \max_{1 \leq t \leq n} \sup_{\boldsymbol{\upsilon} \in \Theta^a_{\boldsymbol{\upsilon}}} \left\| \frac{1}{h_t(\boldsymbol{\upsilon})} \frac{\partial h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\zeta}} \right\| \\ &\leq C n^{-1/a} \left(\max_{1 \leq t \leq n} \xi^{\iota}_{\rho,t-1} + \max_{1 \leq t \leq n} \widetilde{\xi}^{\iota}_{\rho,t-1} \right) = o(1), \end{aligned}$$

where the last equality holds for sufficiently small $\iota \in (0, \iota_0/a)$ using the facts that

$$\max_{1 \leq t \leq n} \widetilde{\xi}_{\rho,t-1} = o(n^{1/(2+\iota_0)}) = o(n^{1/\iota_0}), \quad \max_{1 \leq t \leq n} \xi_{\rho,t-1} = o(n^{1/\iota_0})$$

by a standard Borel-Cantelli argument; see, e.g., Lemma 11.2 in Owen (2001). Applying the Delta method yields that

$$\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\left|\sqrt{\frac{h_t(\boldsymbol{\zeta}^0,\boldsymbol{\beta})}{h_t(\boldsymbol{\zeta},\boldsymbol{\beta})}}-1\right|\le Cn^{-1/a}\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)=o(1)$$

uniformly for $t=1,\ldots,n$. On the other hand, by Taylor expansion and Lemma F.1 we have

$$\begin{aligned}\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\left|\sqrt{\frac{h_t}{h_t(\boldsymbol{\zeta}_0,\boldsymbol{\beta})}}-1\right| \le& n^{-1/a}\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\frac{1}{h_t^{1/2}(\boldsymbol{\upsilon})}\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\left\|\frac{1}{\sqrt{h_t(\boldsymbol{\upsilon})}}\frac{\partial h_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\beta}}\right\|\\ \le& Cn^{-1/a}\xi_{\rho,t-1}.\end{aligned}$$

Hence,

$$\begin{aligned}&\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\left|\sqrt{\frac{h_t}{h_t(\boldsymbol{\upsilon})}}-1\right|\\ \le& \sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\left|\sqrt{\frac{h_t}{h_t(\boldsymbol{\zeta}^0,\boldsymbol{\beta})}}-1\right|\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\left|\sqrt{\frac{h_t(\boldsymbol{\zeta}^0,\boldsymbol{\beta})}{h_t(\boldsymbol{\upsilon})}}\right|+\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\left|\sqrt{\frac{h_t(\boldsymbol{\zeta}^0,\boldsymbol{\beta})}{h_t(\boldsymbol{\upsilon})}}-1\right|\\ \le& Cn^{-1/a}\xi_{\rho,t-1}+Cn^{-1/a}\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)\\ \le& Cn^{-1/a}\left(\xi_{\rho,t-1}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)\end{aligned}$$

uniformly for $t=1,\ldots,n$. □

LEMMA F.3. *For all $\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a$ and small $\iota\in(0,\iota_0/a)$*

(i) $|\eta_t(\boldsymbol{\upsilon})-\eta_t|\le Cn^{-1/a}\left\{\|\boldsymbol{Z}_t\|+|\eta_t|\xi_{\rho,t-1}+|\eta_t|\widetilde{\xi}_{\rho,t-1}^{\iota}\right\}$,

(ii) $|\eta_t(\boldsymbol{\upsilon})|\le C\left\{n^{-1/a}\|\boldsymbol{Z}_t\|+|\eta_t|(1+n^{-1/a}\xi_{\rho,t-1})\right\}$,

(iii)

$$\begin{aligned}\left\|\frac{\partial\eta_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|\le& Cn^{-1/a}\|\boldsymbol{Z}_t\|\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)\\ &+C|\eta_t|\left(1+n^{-\frac{1}{a}}\xi_{\rho,t-1}\right)\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right),\end{aligned}$$

(iv) $\left\|\frac{\partial\eta_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\beta}}\right\|\le C(1+n^{-1/a}\xi_{\rho,t-1})\|\boldsymbol{Z}_t\|+C(1+n^{-1/a}\xi_{\rho,t-1})\xi_{\rho,t-1}|\eta_t|$,

(v) $\left\|\frac{\partial\eta_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\beta}}-\frac{\partial\eta_t}{\partial\boldsymbol{\beta}}\right\|\le Cn^{-1/a}(1+|\eta_t|)\left(\xi_{\rho,t}^2+\xi_{\rho,t}\widetilde{\xi}_{\rho,t-1}^{\iota}\right)$,

(vi) $\left\|\frac{\partial\eta_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}-\frac{\partial\eta_t}{\partial\boldsymbol{\zeta}}\right\|\le Cn^{-1/a}(1+|\eta_t|)\left(\xi_{\rho,t}^{1+\iota}+\xi_{\rho,t}\widetilde{\xi}_{\rho,t-1}^{2\iota}\right)$,

where the constant C is independent of t and n.

PROOF. We have the statement (i) by Lemma F.2 since

$$\begin{aligned}|\eta_t(\boldsymbol{\upsilon}) - \eta_t| \leq & \frac{|\epsilon_t(\boldsymbol{\beta}) - \epsilon_t|}{h_t(\boldsymbol{\upsilon})} + |\eta_t| \left| \sqrt{\frac{h_t}{h_t(\boldsymbol{\upsilon})}} - 1 \right| \\ \leq & C n^{-1/a} \|\boldsymbol{Z}_t\| + C|\eta_t| n^{-1/a} (\xi_{\rho,t-1} + \widetilde{\xi}^{\iota}_{\rho,t-1}).\end{aligned}$$

Statement (ii) immediately follows because

$$\begin{aligned}|\eta_t(\boldsymbol{\upsilon})| \leq & |\eta_t(\boldsymbol{\upsilon}) - \eta_t| + |\eta_t| \\ \leq & C n^{-1/a} \|\boldsymbol{Z}_t\| + C|\eta_t| \left(1 + n^{-1/a}\xi_{\rho,t-1} + n^{-1/a}\widetilde{\xi}^{\iota}_{\rho,t-1}\right)\end{aligned}$$

and $\max_{1\leq t\leq n} \widetilde{\xi}^{\iota}_{\rho,t-1} = o\left(n^{\iota/(2+\iota_0)}\right) = o(n^{1/a})$ by a standard Borel-Cantelli argument; see, e.g., Lemma 11.2 in Owen (2001). For statement (iii), apply statement (ii) and Lemma F.1 to get

$$\begin{aligned}\left\| \frac{\partial \eta_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\zeta}} \right\| = & \left\| -\frac{1}{2}\eta_t(\boldsymbol{\upsilon}) \frac{1}{h_t(\boldsymbol{\upsilon})} \frac{\partial h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\zeta}} \right\| \\ \leq & C\,|\eta_t(\boldsymbol{\upsilon})| \left\| \frac{1}{h_t(\boldsymbol{\upsilon})} \frac{\partial h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\zeta}} \right\| \\ \leq & C \left\{ n^{-1/a} \|\boldsymbol{Z}_t\| + |\eta_t|(1 + n^{-1/a}\xi_{\rho,t-1}) \right\} \left(\xi^{\iota}_{\rho,t-1} + \widetilde{\xi}^{\iota}_{\rho,t-1} \right) \\ \leq & C n^{-1/a} \left(\xi^{\iota}_{\rho,t-1} + \widetilde{\xi}^{\iota}_{\rho,t-1} \right) \|\boldsymbol{Z}_t\| \\ & + C \left(1 + n^{-1/a}\xi_{\rho,t-1}\right) \left(\xi^{\iota}_{\rho,t-1} + \widetilde{\xi}^{\iota}_{\rho,t-1} \right) |\eta_t|.\end{aligned}$$

Similarly, we have

$$\begin{aligned}\left\| \frac{\partial \eta_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\beta}} \right\| \leq & \frac{1}{\sqrt{h_t(\boldsymbol{\upsilon})}} \left\| -\boldsymbol{Z}_t - \frac{1}{2}\eta_t(\boldsymbol{\upsilon}) \frac{1}{\sqrt{h_t(\boldsymbol{\upsilon})}} \frac{\partial h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\beta}} \right\| \\ \leq & C \|\boldsymbol{Z}_t\| + C \left\{ n^{-1/a} \|\boldsymbol{Z}_t\| + |\eta_t|(1 + n^{-1/a}\xi_{\rho,t-1}) \right\} \xi_{\rho,t-1} \\ \leq & C(1 + n^{-1/a}\xi_{\rho,t-1}) \|\boldsymbol{Z}_t\| + C(1 + n^{-1/a}\xi_{\rho,t-1})\xi_{\rho,t-1}|\eta_t|.\end{aligned}$$

Furthermore, invoking Lemmas F.1 and F.2 and applying Taylor expansion,

$$\begin{aligned}& \left\| \frac{\partial \eta_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\beta}} - \frac{\partial \eta_t}{\partial \boldsymbol{\beta}} \right\| \\ \leq & \frac{1}{\sqrt{h_t}} \|\boldsymbol{Z}_t\| \left| \sqrt{\frac{h_t}{h_t(\boldsymbol{\upsilon})}} - 1 \right| + |\eta_t(\boldsymbol{\upsilon}) - \eta_t| \frac{1}{h_t(\boldsymbol{\upsilon})} \left\| \frac{\partial h_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\beta}} \right\|\end{aligned}$$

$$
\begin{aligned}
&+ n^{-1/a}|\eta_t| \left\{ \frac{1}{h_t^2(\boldsymbol{v})} \left\| \frac{\partial h_t(\boldsymbol{v})}{\partial \boldsymbol{\beta}} \frac{\partial h_t(\boldsymbol{v})}{\partial \boldsymbol{v}}^T \right\| + \frac{1}{h_t(\boldsymbol{v})} \left\| \frac{\partial^2 h_t(\boldsymbol{v})}{\partial \boldsymbol{\beta} \partial \boldsymbol{v}^T} \right\| \right\} \\
\leq & Cn^{-1/a} \left\{ \|\boldsymbol{Z}_t\| \left(\xi_{\rho,t-1} + \widetilde{\xi}^{\iota}_{\rho,t-1} \right) + \left(\|\boldsymbol{Z}_t\| + |\eta_t|\xi_{\rho,t-1} + |\eta_t|\widetilde{\xi}^{\iota}_{\rho,t-1} \right) \xi_{\rho,t-1} \right. \\
&\left. + |\eta_t|\xi^2_{\rho,t-1} \right\} \leq Cn^{-1/a}(1+|\eta_t|) \left(\xi^2_{\rho,t} + \xi_{\rho,t}\widetilde{\xi}^{\iota}_{\rho,t-1} \right).
\end{aligned}
$$

Similarly, with small ι,

$$
\begin{aligned}
&\left\| \frac{\partial \eta_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta}} - \frac{\partial \eta_t}{\partial \boldsymbol{\zeta}} \right\| \\
\leq & |\eta_t(\boldsymbol{v}) - \eta_t| \frac{1}{h_t(\boldsymbol{v})} \left\| \frac{\partial h_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta}} \right\| \\
&+ n^{-1/a}|\eta_t| \left\{ \frac{1}{h_t^2(\boldsymbol{v})} \left\| \frac{\partial h_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta}} \frac{\partial h_t(\boldsymbol{v})}{\partial \boldsymbol{v}}^T \right\| + \frac{1}{h_t(\boldsymbol{v})} \left\| \frac{\partial^2 h_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta} \partial \boldsymbol{v}^T} \right\| \right\} \\
\leq & Cn^{-1/a} \left\{ \left(\|\boldsymbol{Z}_t\| + |\eta_t|\xi_{\rho,t-1} + |\eta_t|\widetilde{\xi}^{\iota}_{\rho,t-1} \right) \left(\xi^{\iota}_{\rho,t-1} + \widetilde{\xi}^{\iota}_{\rho,t-1} \right) \right. \\
&\left. + |\eta_t| \left(\xi_{\rho,t-1} + \widetilde{\xi}^{\iota}_{\rho,t-1} \right) \right\} \\
\leq & Cn^{-1/a}(1+|\eta_t|) \left(\xi^{1+\iota}_{\rho,t} + \xi_{\rho,t}\widetilde{\xi}^{2\iota}_{\rho,t-1} \right),
\end{aligned}
$$

where in the last line we use the fact that $\widetilde{\xi}^{\iota}_{\rho,t-1} \leq \widetilde{\xi}^{2\iota}_{\rho,t-1}$. □

LEMMA F.4. *For all small positive $\delta \leq \iota_0$, uniformly for $t = 1, \ldots, n$,*

(i) $E\left(w_t \sup_{\Theta^a_{\boldsymbol{v}}} \left\| \frac{\partial l_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta} \partial \boldsymbol{v}^T} \right\| \right)^{2(1+\delta)} < \infty,$

(ii) $E\left(w_t \sup_{\Theta^a_{\boldsymbol{v}}} \left\| \frac{\partial l_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta}} \right\| \right)^{2(1+\delta)} < \infty,$

when n is large enough.

PROOF. All the inequalities below hold uniformly for $t = 1, \ldots, n$ and sufficiently small ι; we do not repeat this statement for presentation convenience. Write

$$
\begin{aligned}
\left\| \frac{\partial l_t(\boldsymbol{v}, \theta^0)}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\beta}^T} \right\| \leq & |\eta_t(\boldsymbol{v})| \left\| \frac{\partial \eta_t(\boldsymbol{v})}{\partial \boldsymbol{\beta}} \right\| \left\| \frac{1}{h_t(\boldsymbol{v})} \frac{\partial h_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta}} \right\| \\
&+ \left| \eta_t^2(\boldsymbol{v}) - 1 \right| \left\{ \frac{1}{h_t^2(\boldsymbol{v})} \left\| \frac{\partial h_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta}} \right\| \left\| \frac{\partial h_t(\boldsymbol{v})}{\partial \boldsymbol{\beta}} \right\| + \frac{1}{h_t(\boldsymbol{v})} \left\| \frac{\partial^2 h_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta} \partial \boldsymbol{\beta}^T} \right\| \right\} \\
=: & \Delta_{t,1} + \Delta_{t,2}.
\end{aligned}
$$

Applying Lemmas F.1 and F.3 we obtain, uniformly for $\boldsymbol{\upsilon} \in \Theta_{\boldsymbol{\upsilon}}^a$,

$$\begin{aligned}
\Delta_{t,1} \leq & C\left\{n^{-1/a}\left\|\boldsymbol{Z}_t\right\|+|\eta_t|(1+n^{-1/a}\xi_{\rho,t-1})\right\}(1+n^{-1/a}\xi_{\rho,t-1}) \\
& (\|\boldsymbol{Z}_t\|+\xi_{\rho,t-1}|\eta_t|)\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right) \\
\leq & Cn^{-1/a}\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)(1+n^{-1/a}\xi_{\rho,t-1})\left\|\boldsymbol{Z}_t\right\|^2 \\
& + Cn^{-2/a}\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)\xi_{\rho,t-1}^2\left\|\boldsymbol{Z}_t\right\||\eta_t| \\
& + Cn^{-2/a}\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)\xi_{\rho,t-1}^3\eta_t^2 \\
& + C\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)(\left\|\boldsymbol{Z}_t\right\||\eta_t|+\xi_{\rho,t-1}\eta_t^2) \\
\leq & C\xi_{\rho,t}^3+C\xi_{\rho,t}^3|\eta_t|+C\xi_{\rho,t}^3+C\xi_{\rho,t}^3\eta_t^2+C\xi_{\rho,t}^{1+\iota}(|\eta_t|+\eta_t^2) \\
& + C\widetilde{\xi}_{\rho,t-1}^{\iota}\xi_{\rho,t}(|\eta_t|+\eta_t^2) \\
\leq & C\xi_{\rho,t}^3(1+\eta_t^2)+C\xi_{\rho,t}\widetilde{\xi}_{\rho,t-1}^{\iota}(|\eta_t|+\eta_t^2),
\end{aligned}$$

where in the penultimate line we use $\|\boldsymbol{Z}_t\| \leq \xi_{\rho,t}$ and $\xi_{\rho,t-1} \leq C\xi_{\rho,t}$.

Since $w_t\xi_{\rho,t}^3 \leq C$ by construction,

$$\begin{aligned}
& E\left(w_t\sup_{\Theta_{\boldsymbol{\upsilon}}^a}\Delta_{t,1}\right)^{2(1+\delta)} \\
\leq & C(1+E\eta_t^{4(1+\delta)})+CE\left(\widetilde{\xi}_{\rho,t-1}^{2(1+\delta)\iota}\right)(E|\eta_t|^{2(1+\delta)}+E\eta_t^{4(1+\delta)})<\infty.
\end{aligned}$$

Similarly,

$$\begin{aligned}
& \sup_{\Theta_{\boldsymbol{\upsilon}}^a}\Delta_{t,2} \\
\leq & C\left\{n^{-2/a}\left\|\boldsymbol{Z}_t\right\|^2+\eta_t^2(1+n^{-2/a}\xi_{\rho,t-1}^2)+1\right\}\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)\xi_{\rho,t-1} \\
\leq & Cn^{-2/a}\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)\xi_{\rho,t-1}\left(\left\|\boldsymbol{Z}_t\right\|^2+\xi_{\rho,t-1}^2\eta_t^2\right) \\
& + C\left(\xi_{\rho,t-1}^{\iota}+\widetilde{\xi}_{\rho,t-1}^{\iota}\right)\xi_{\rho,t-1}(1+\eta_t^2) \\
\leq & C\xi_{\rho,t}^3(1+\eta_t^2)+C\xi_{\rho,t-1}^{1+\iota}(1+\eta_t^2)+C\widetilde{\xi}_{\rho,t-1}^{\iota}\xi_{\rho,t-1}(1+\eta_t^2) \\
\leq & C\xi_{\rho,t}^3(1+\eta_t^2)+C\widetilde{\xi}_{\rho,t-1}^{\iota}\xi_{\rho,t}(1+\eta_t^2).
\end{aligned}$$

Therefore,

$$E\left(w_t\sup_{\Theta_{\boldsymbol{\upsilon}}^a}\Delta_{t,2}\right)^{2(1+\delta)}$$

$$\leq C(1+E\eta_t^{4(1+\delta)}) + CE\left(\widetilde{\xi}_{\rho,t-1}^{2(1+\delta)\iota}\right)(E|\eta_t|^{2(1+\delta)} + E\eta^{4(1+\delta)}) < \infty,$$

and it follows that

$$E\left(w_t \sup_{\Theta_{\boldsymbol{v}}^a}\left\|\frac{\partial l_t(\boldsymbol{v},\theta^0)}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\beta}^T}\right\|\right)^{2(1+\delta)}$$
$$\leq CE\left(w_t \sup_{\Theta_{\boldsymbol{v}}^a}\Delta_{t,1}\right)^{2(1+\delta)} + CE\left(w_t \sup_{\Theta_{\boldsymbol{v}}^a}\Delta_{t,2}\right)^{2(1+\delta)} < \infty.$$

Even easier, we can show that $E\left(w_t \sup_{\Theta_{\boldsymbol{v}}^a}\left\|\frac{\partial l_t(\boldsymbol{v},\theta^0)}{\partial\boldsymbol{\zeta}\partial\boldsymbol{\zeta}^T}\right\|^{2(1+\delta)}\right) < \infty$ and omit the details. This completes the proof of statement (i).

Applying Lemmas F.1 and F.3 again yields

$$\begin{aligned}
\sup_{\Theta_{\boldsymbol{v}}^a}\left\|\frac{\partial l_t(\boldsymbol{v})}{\partial\boldsymbol{\zeta}}\right\|^2 &\leq C\left(\eta_t^2(\boldsymbol{v})-1\right)^2\left\|\frac{1}{h_t(\boldsymbol{v})}\frac{\partial h_t(\boldsymbol{v})}{\partial\boldsymbol{\zeta}}\right\|^2 \\
&\leq C\left(\eta_t^4(\boldsymbol{v})+1\right)\left(\xi_{\rho,t-1}^{2\iota}+\widetilde{\xi}_{\rho,t-1}^{2\iota}\right) \\
&\leq C\left(\eta_t^4(1+n^{-4/a}\xi_{\rho,t-1}^4)+n^{-4/a}\|\boldsymbol{Z}_t\|^4+1\right)\left(\xi_{\rho,t-1}^{2\iota}+\widetilde{\xi}_{\rho,t-1}^{2\iota}\right) \\
&\leq C\left(\eta_t^4+1\right)\left(\xi_{\rho,t-1}^{2\iota}+\widetilde{\xi}_{\rho,t-1}^{2\iota}\right) \\
&\quad + Cn^{-4/a}\left(\xi_{\rho,t-1}^{2\iota}+\widetilde{\xi}_{\rho,t-1}^{2\iota}\right)\left(\xi_{\rho,t-1}^4+\|\boldsymbol{Z}_t\|^4\right) \\
&\leq C\left(\eta_t^4+1\right)\left(\xi_{\rho,t-1}^{2\iota}+\widetilde{\xi}_{\rho,t-1}^{2\iota}\right)+C\xi_{\rho,t}^4.
\end{aligned}$$

Statement (ii) then follows since

$$E\left(w_t^2 \sup_{\Theta_{\boldsymbol{v}}^a}\left\|\frac{\partial l_t(\boldsymbol{v})}{\partial\boldsymbol{\zeta}}\right\|^2\right)^{1+\delta}$$
$$\leq CE\left(\eta_t^{4(1+\delta)}+1\right)\left(E\xi_{\rho,t-1}^{2\iota(1+\delta)}+E\widetilde{\xi}_{\rho,t-1}^{2\iota(1+\delta)}\right)+C<\infty.$$

The proof is now complete. □

LEMMA F.5. *For all* $\delta \in (0,1)$,

(i) $\max_{t\geq n^\delta}\frac{|w_t-\widetilde{w}_t|}{w_t} = o_P(1)$,
(ii) $\max_{1\leq t\leq n^\delta}\frac{|w_t-\widetilde{w}_t|}{w_t} = O_P(1)$.

PROOF. By definition, it is easy to show that

$$\frac{|w_t - \widetilde{w}_t|}{w_t} \le C \sum_{i=t}^{\infty} e^{-\log^2(i+1)} \|\boldsymbol{X}_{t-i}\| = C \sum_{i=0}^{\infty} e^{-\log^2(t+1+i)} \|\boldsymbol{X}_{-i}\|, \tag{F.1}$$

where C is independent of t. Therefore, for large n

$$\begin{aligned} E\left(\max_{t\ge n^\delta}\left(\frac{|w_t - \widetilde{w}_t|}{w_t}\right)^{\iota_0'}\right) \le& C\sum_{i=0}^{\infty} e^{-\iota_0' \log^2(n^\delta+1+i)} E\|\boldsymbol{X}_{-i}\|^{\iota_0'} \\ \le& C \sum_{k=n^\delta+1}^{\infty} k^{-\iota_0'\log(n^\delta+1+i)} \le C \sum_{k=n^\delta+1}^{\infty} k^{-2} \to 0, \end{aligned}$$

where $\iota_0' = \min\{\iota_0, 1\}$. Statement (i) then follows by Markov inequality.

Similarly, from (F.1),

$$\begin{aligned} & E\left(\max_{1\le t\le n^\delta}\left(\frac{|w_t - \widetilde{w}_t|}{w_t}\right)^{\iota_0'}\right) \\ \le& C\sum_{i=0}^{\infty} e^{-\iota_0'\log^2(1+i)} E\|\boldsymbol{X}_{-i}\|^{\iota_0'} \\ \le& C\sum_{i=0}^{\infty} e^{-\iota_0'\log^2(1+i)} \\ \le& C\left\{\sum_{i=0}^{\exp(2/\iota_0')} e^{-\iota_0'\log^2(1+i)} + \sum_{i=\exp(2/\iota_0')}^{\infty} (i+1)^{-2}\right\} < \infty, \end{aligned}$$

which implies statement (ii). □

LEMMA F.6. *There exists $\rho \in (0,1)$ such that for all small $\iota > 0$*

(i) $\left|\widetilde{l}_t(\boldsymbol{v}) - l_t(\boldsymbol{v})\right| \le O(\rho^t)\left(\xi_{\rho,t}^2 + \widetilde{\xi}_{\rho,t-1}^2\right)(1+\eta_t^2)$,

(ii) $\left\|\frac{\partial \widetilde{l}_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta}} - \frac{\partial l_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta}}\right\| \le O(\rho^t)\left(\xi_{\rho,t}^{2+\iota} + \widetilde{\xi}_{\rho,t-1}^{2+\iota}\right)(1+\eta_t^2)$,

(iii) $\left\|\frac{\partial^2 \widetilde{l}_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta}\partial \boldsymbol{v}^T} - \frac{\partial^2 l_t(\boldsymbol{v})}{\partial \boldsymbol{\zeta}\partial \boldsymbol{v}^T}\right\| \le O(\rho^t)\left(\xi_{\rho,t}^{3+\iota} + \widetilde{\xi}_{\rho,t-1}^{2+\iota}\xi_{\rho,t}\right)(1+\eta_t^2)$,

(iv) $\left|K\left(\frac{(\theta-\boldsymbol{\beta}^T\boldsymbol{z})-\widetilde{\eta}_t(\boldsymbol{v})}{h}\right) - K\left(\frac{(\theta-\boldsymbol{\beta}^T\boldsymbol{z})-\eta_t(\boldsymbol{v})}{h}\right)\right| \le O\left(\frac{\rho^t}{h}\right)\xi_{\rho,t}(1+|\eta_t|)$,

(v) $\left\|\frac{\partial}{\partial \boldsymbol{v}} K\left(\frac{(\theta-\boldsymbol{\beta}^T\boldsymbol{z})-\widetilde{\eta}_t(\boldsymbol{v})}{h}\right) - \frac{\partial}{\partial \boldsymbol{v}} K\left(\frac{(\theta-\boldsymbol{\beta}^T\boldsymbol{z})-\eta_t(\boldsymbol{v})}{h}\right)\right\|$
$\le O\left(\frac{\rho^t}{h^2}\right)\left(\xi_{\rho,t}^2 + \xi_{\rho,t}\widetilde{\xi}_{\rho,t-1}^{\iota}\right)(1+|\eta_t|)$

uniformly in $\Theta_{\boldsymbol{v}}^a$.

PROOF. Note that $\epsilon_t(\boldsymbol{\beta})$ does not depend on the initial values, and recall that $|\epsilon_t(\gamma)| \le C(|\epsilon_t| + \|X_t\|)$ for large constant $C > 0$. Using the GARCH equations, it is straightforward to show that there exists $\rho \in (0,1)$ such that

$$\sup_{\Theta_{\boldsymbol{v}}^a} \left|\widetilde{h}_t(\boldsymbol{v}) - h_t(\boldsymbol{v})\right| \le O(\rho^t), \quad \sup_{\Theta_{\boldsymbol{v}}^a} \left\|\frac{\partial \widetilde{h}_t(\boldsymbol{v})}{\partial \boldsymbol{v}} - \frac{\partial h_t(\boldsymbol{v})}{\partial \boldsymbol{v}}\right\| \le O(\rho^t)$$

and

$$\sup_{\Theta_{\boldsymbol{v}}^a} \left\|\frac{\partial^2 \widetilde{h}_t(\boldsymbol{v})}{\partial \boldsymbol{v} \partial \boldsymbol{v}^T} - \frac{\partial^2 h_t(\boldsymbol{v})}{\partial \boldsymbol{v} \partial \boldsymbol{v}^T}\right\| \le O(\rho^t).$$

The proofs of the statements (i)-(iii) are then completely analogous to the proof of Lemma A.4 in Ling (2007), and we omit the details. For the rest, we need to note that the preceding inequalities together with Lemma F.3 imply that

$$|\widetilde{\eta}_t(\boldsymbol{v}) - \eta_t(\boldsymbol{v})| \le C\left|\widetilde{h}_t(\boldsymbol{v}) - h_t(\boldsymbol{v})\right| \eta_t(\boldsymbol{v}, \boldsymbol{\beta}) \le O(\rho^t)\xi_{\rho,t}(1 + |\eta_t|),$$

and then statements (iv) and (v) follow easily by Taylor expansion. □

Henceforth we take $\rho \in (0,1)$ as the maximum of that in Lemma F.1 and that in Lemma F.6.

LEMMA F.7. *As $n \to \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \bar{\boldsymbol{W}}_t \xrightarrow{d} N(0, \Omega_1), \quad \frac{1}{n}\sum_{t=1}^n \bar{\boldsymbol{W}}_t \bar{\boldsymbol{W}}_t^T \xrightarrow{P} \Omega_1, \quad \frac{1}{n}\sum_{t=1}^n \frac{\partial \bar{\boldsymbol{W}}_t}{\partial \boldsymbol{v}^T} \xrightarrow{P} \Omega_2$$

where $\bar{\boldsymbol{W}}_t := \bar{\boldsymbol{W}}_t(\boldsymbol{v}^0, \theta^0)$.

PROOF. If the initial value $\left\{(Y_s, \boldsymbol{X}_s^T)^T\right\}_{s \le 0}$ are observed and $\widetilde{w}_t$ were used instead of w_t, we can prove the lemma by martingale limit theorems as in Lemma D.1 using conditions G5 and G6. We only need to show that substituting $\widetilde{w}_t$ by w_t and using constant initial values do not change the limits. Let $\delta \in (0,1)$, and we shall split the summations into two parts for $t \le n^\delta$ and $t \ge n^\delta + 1$ respectively. Note that, for any $\rho \in (0,1)$, $\rho^t/h = o(\rho^{t/2})$ uniformly for $t \ge n^\delta$ as $n \to \infty$. By Lemmas F.5–F.6, uniformly for $\boldsymbol{v} \in \Theta_{\boldsymbol{v}}^a$,

$$\frac{1}{n}\sum_{t=n^\delta+1}^n \left| w_t \bar{W}_{t,1}(\boldsymbol{v}, \theta^0) - \widetilde{w}_t \left(K\left(\frac{(\theta^0 - \boldsymbol{\beta}^T \boldsymbol{z}) - \widetilde{\eta}_t(\boldsymbol{v})}{h}\right) - \alpha\right)\right|^2$$

$$
\begin{aligned}
\leq & \frac{8}{n}\sum_{t=n^\delta}^{n}|w_t-\widetilde{w}_t|^2 \\
& + \frac{2}{n}\sum_{t=n^\delta+1}^{n}\widetilde{w}_t^2\left|K\left(\frac{(\theta^0-\boldsymbol{\beta}^T\boldsymbol{z})-\widetilde{\eta}_t(\boldsymbol{\upsilon})}{h}\right)-K\left(\frac{(\theta^0-\boldsymbol{\beta}^T\boldsymbol{z})-\eta_t(\boldsymbol{\upsilon})}{h}\right)\right|^2 \\
= & o_p(1)+O_p(1)\cdot\frac{1}{n}\sum_{t=n^\delta}^{n}\rho^{t/2}(1+\eta_t^2)=o_p(1).
\end{aligned}
$$

On the other hand, uniformly for $\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a$,

$$
\begin{aligned}
& \frac{1}{n}\sum_{t=1}^{n^\delta}\left|w_t\bar{W}_{t,1}(\boldsymbol{\upsilon},\theta^0)-\widetilde{w}_t\left(K\left(\frac{(\theta^0-\boldsymbol{\beta}^T\boldsymbol{z})-\widetilde{\eta}_t(\boldsymbol{\upsilon})}{h}\right)-\alpha\right)\right|^2 \\
\leq & \frac{8}{n}\sum_{t=1}^{n^\delta}|w_t-\widetilde{w}_t|^2+\frac{8}{n}\sum_{t=1}^{n^\delta}\widetilde{w}_t^2=o(1).
\end{aligned}
$$

Summing up the above inequalities yields

$$
\frac{1}{n}\sum_{t=1}^{n}\left|w_t\bar{W}_{t,1}(\boldsymbol{\upsilon},\theta^0)-\widetilde{w}_t\left(K\left(\frac{(\theta-\boldsymbol{\beta}^T\boldsymbol{z})-\widetilde{\eta}_t(\boldsymbol{\upsilon})}{h}\right)-\alpha\right)\right|^2\xrightarrow{P}0.
$$

uniformly for $\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a$.

Similarly, it is easy to verify that

$$
\begin{aligned}
& \frac{1}{n}\sum_{t=1}^{n}\left\|w_t\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}-\widetilde{w}_t\frac{\partial\widetilde{l}_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|^2 \\
\leq & \frac{2}{n}\sum_{t=1}^{n}\left\{\left(\frac{w_t-\widetilde{w}_t}{w_t}\right)^2\left(w_t^2\left\|\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|^2\right)+\widetilde{w}_t^2\left\|\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}-\frac{\partial\widetilde{l}_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|^2\right\}.
\end{aligned}
$$

Again splitting the sum for $t\leq n^\delta$ and $t\geq n^\delta+1$ and applying Lemmas F.4–F.6 yield that

$$
\begin{aligned}
& \frac{2}{n}\sum_{t=1}^{n}\left(\frac{w_t-\widetilde{w}_t}{w_t}\right)^2\left(w_t^2\left\|\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|^2\right) \\
= & O_p(1)\cdot\frac{1}{n}\sum_{t=1}^{n^\delta}w_t^2\left\|\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|^2+o_p(1)\cdot\frac{1}{n}\sum_{t=n^\delta+1}^{n}w_t^2\left\|\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|^2 \\
= & O_p(1)\cdot O_p(n^{\delta-1})+o_p(1)\cdot O_p(1)=o_p(1)
\end{aligned}
$$

and, with sufficiently small $\iota > 0$,

$$
\begin{aligned}
&\frac{2}{n}\sum_{t=1}^{n}\widetilde{w}_t^2\left\|\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}-\frac{\partial\widetilde{l}_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|^2\\
&\le\frac{2}{n}\sum_{t=1}^{n^\delta}\left(w_t^2\left\|\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|^2+\widetilde{w}_t^2\left\|\frac{\partial\widetilde{l}_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|^2\right)\\
&+\frac{1}{n}\sum_{t=n^\delta+1}^{n}O(\rho^t)(1+\widetilde{\xi}_{\rho,t-1}^{2+\iota})(1+\eta_t^2)\\
&=O_p(n^{\delta-1})+o_p(1)=o_p(1)
\end{aligned}
$$

uniformly for $\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a$. Hence, uniformly for $\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a$,

$$
\frac{1}{n}\sum_{t=1}^{n}\left\|w_t\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}-\widetilde{w}_t\frac{\partial\widetilde{l}_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|^2\xrightarrow{P}0. \tag{F.2}
$$

Next, taking $\boldsymbol{\upsilon}=\boldsymbol{\upsilon}_0$ and applying Cauchy-Schwartz inequality, we can show that constant initial values do not change the limit in the second statement. The proof for the third statement is similar, and we omit the details. Like the proof of Lemma D.1, the first statement follows by the second one using martingale central limit theorem and our proof is complete now. □

LEMMA F.8. *For all $a\in(0,\min\{r_0,2(1+\iota_0)\})$, as $n\to\infty$,*

$$
\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\bar{\boldsymbol{W}}_t(\boldsymbol{\upsilon},\theta^0)}{\partial\boldsymbol{\upsilon}^T}-\Omega_2\right\|\xrightarrow{P}0.
$$

PROOF. It is equivalent to show that

$$
\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\bar{\boldsymbol{W}}_t(\boldsymbol{\upsilon},\theta^0)}{\partial\boldsymbol{\upsilon}^T}-\frac{1}{n}\sum_{t=1}^{n}\frac{\partial\bar{\boldsymbol{W}}_t}{\partial\boldsymbol{\upsilon}^T}\right\|\xrightarrow{P}0.
$$

We prove the convergence coordinate-wisely. For the least-squares estimating functions, the convergence is trivial since its derivative does not depend on $\boldsymbol{\upsilon}$. Like in Ling (2007), invoking (F.2) and ergodic theorem it is easy to verify the convergence for GARCH estimating functions; for details, we refer to the last paragraph in Proof of Theorem 3.1(ii) therein. It remains to show that

$$
\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^a}\left\|\frac{1}{n}\sum_{t=1}^{n}w_t\left(\frac{\partial\bar{W}_{t,1}(\boldsymbol{\zeta},\boldsymbol{\beta},\theta^0)}{\partial\boldsymbol{\upsilon}^T}-\frac{\partial\bar{W}_{t,1}}{\partial\boldsymbol{\upsilon}^T}\right)\right\|\xrightarrow{P}0.
$$

We only prove the hardest part, that is, the convergence in the coordinates for $\boldsymbol{\beta}$; the proofs for other coordinates are similar. All the inequalities below hold uniformly for $t = 1, \dots, n$ and sufficiently small ι; we do not repeat this statement for presentation convenience.

We first bound the difference as follows:

$$\left\| \frac{1}{n} \sum_{t=1}^{n} w_t \left(\frac{\partial \bar{W}_{t,1}(\boldsymbol{\zeta}, \boldsymbol{\beta}, \theta^0)}{\partial \boldsymbol{\beta}} - \frac{\partial \bar{W}_{t,1}}{\partial \boldsymbol{\beta}} \right) \right\| \leq \frac{1}{n} \sum_{t=1}^{n} w_t \left\| \frac{\partial \bar{W}_{t,1}(\boldsymbol{\zeta}, \boldsymbol{\beta}, \theta^0)}{\partial \boldsymbol{\beta}} - \frac{\partial \bar{W}_{t,1}}{\partial \boldsymbol{\beta}} \right\|.$$

By a direct calculation,

$$\begin{aligned} &\frac{\partial \bar{W}_{t,1}(\boldsymbol{\zeta}, \boldsymbol{\beta}, \theta^0)}{\partial \boldsymbol{\beta}} \\ &= -\frac{1}{h} g \left(\frac{F_\eta^{-1}(\alpha) - \eta_t - (\boldsymbol{\beta} - \boldsymbol{\beta}^0)^T \boldsymbol{z}/\sigma + (\eta_t - \eta_t(\boldsymbol{\upsilon}))}{h} \right) \left(\frac{\partial \eta_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\beta}} + \frac{\boldsymbol{z}}{\sigma} \right). \end{aligned}$$

Hence,

$$\begin{aligned} &\left\| \frac{\partial \bar{W}_{t,1}(\varsigma, \boldsymbol{\zeta}, \boldsymbol{\beta}, \theta)}{\partial \boldsymbol{\beta}} - \frac{\partial \bar{W}_{t,1}}{\partial \boldsymbol{\beta}} \right\| \\ \leq &\left\| \frac{1}{h} g \left(\frac{F_\eta^{-1}(\alpha) - \eta_t - (\boldsymbol{\beta} - \boldsymbol{\beta}^0)^T \boldsymbol{z}/\sigma + (\eta_t - \eta_t(\boldsymbol{\upsilon}))}{h} \right) - \frac{1}{h} g \left(\frac{F_\eta^{-1}(\alpha) - \eta_t}{h} \right) \right\| \\ &\cdot \left\| \frac{\partial \eta_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\beta}} + \frac{\boldsymbol{z}}{\sigma} \right\| + \frac{1}{h} g \left(\frac{F_\eta^{-1}(\alpha) - \eta_t}{h} \right) \left\| \frac{\partial \eta_t(\boldsymbol{\upsilon})}{\partial \boldsymbol{\beta}} - \frac{\partial \eta_t}{\partial \boldsymbol{\beta}} \right\| =: \Gamma_t + \Delta_t. \end{aligned}$$

Using Taylor expansion with bounded g' and applying (i) and (iv) in Lemma F.3, we have

$$\begin{aligned} \Gamma_t =& \Gamma_t I(|F_\eta^{-1}(\alpha) - \eta_t| \leq h) + \Gamma_t I(|F_\eta^{-1}(\alpha) - \eta_t| > h) \\ \leq& \frac{C}{n^{1/a} h^2} \left(\|\boldsymbol{Z}_t\|^2 + \xi_{\rho,t-1}^2 \right) \left(1 + n^{-1/a} \xi_{\rho,t-1} \right) I(|F_\eta^{-1}(\alpha) - \eta_t| \leq h) \\ &+ \frac{C \widetilde{\xi}_{\rho,t-1}^{\iota}}{n^{1/a} h^2} (\|\boldsymbol{Z}_t\| + \xi_{\rho,t-1}) \left(1 + n^{-1/a} \xi_{\rho,t-1} \right) I(|F_\eta^{-1}(\alpha) - \eta_t| \leq h) \\ &+ \frac{C}{h} (\|\boldsymbol{Z}_t\| + \xi_{\rho,t-1}) \left(1 + n^{-1/a} \xi_{\rho,t-1} \right) \\ &\times I \left(h < \left| F_\eta^{-1}(\alpha) - \eta_t \right| < h + C n^{-1/a} (\|\boldsymbol{Z}_t\| + |\eta_t| \xi_{\rho,t-1} + |\eta_t| \widetilde{\xi}_{\rho,t-1}^{\iota}) \right) \\ =:& \Gamma_{t,1} + \Gamma_{t,2} + \Gamma_{t,3} \end{aligned}$$

uniformly for $\boldsymbol{\upsilon} \in \Theta_{\boldsymbol{\upsilon}}^a$. It is easy to verify that

$$w_t \Gamma_{t,1} \leq w_t \xi_{\rho,t}^3 \cdot \frac{C}{n^{1/a} h} \cdot \frac{1}{h} I(|F_\eta^{-1}(\alpha) - \eta_t| \leq h),$$

and therefore

$$E\left(w_t\Gamma_{t,1}\right) = o(1)\cdot\frac{1}{h}P(|F_\eta^{-1}(\alpha)-\eta_t|\le h) = o(1).$$

Similarly, with sufficiently small ι, we can show that

$$E\left(w_t\Gamma_{t,2}\right) = o(1)\cdot E(\widetilde{\xi}^\iota_{\rho,t-1})\cdot\frac{1}{h}P(|F_\eta^{-1}(\alpha)-\eta_t|\le h) = o(1).$$

Furthermore, for large n,

$$\begin{aligned}
&I(|F_\eta^{-1}(\alpha)-\eta_t|>h)I\left(\left|F_\eta^{-1}(\alpha)-\eta_t\right|<h+Cn^{-1/a}(\|\boldsymbol{Z}_t\|+|\eta_t|\xi_{\rho,t-1}+|\eta_t|\widetilde{\xi}^\iota_{\rho,t-1})\right)\\
\le& I(|F_\eta^{-1}(\alpha)-\eta_t|>h)I\left(\left|F_\eta^{-1}(\alpha)-\eta_t\right|<h+Cn^{-1/a}((1+|\eta_t|)\xi_{\rho,t}+|\eta_t|\widetilde{\xi}^\iota_{\rho,t-1})\right)\\
\le& I\left(0<F_\eta^{-1}(\alpha)-\eta_t-h<Cn^{-1/a}((1+|\eta_t|)\xi_{\rho,t}+|\eta_t|\widetilde{\xi}^\iota_{\rho,t-1})\right)\\
&+I\left(0>F_\eta^{-1}(\alpha)-\eta_t+h>-Cn^{-1/a}((1+|\eta_t|)\xi_{\rho,t}+|\eta_t|\widetilde{\xi}^\iota_{\rho,t-1})\right)\\
\le& I\left(0<F_\eta^{-1}(\alpha)-\eta_t-h<Cn^{-1/a}(\xi_{\rho,t}+\widetilde{\xi}^\iota_{\rho,t-1})\right)I\left(\xi_{\rho,t}+\widetilde{\xi}^\iota_{\rho,t-1}\le 2n^{1/r_0}\right)\\
&+I\left(0>F_\eta^{-1}(\alpha)-\eta_t+h>-Cn^{-1/a}(\xi_{\rho,t}+\widetilde{\xi}^\iota_{\rho,t-1})\right)I\left(\xi_{\rho,t}+\widetilde{\xi}^\iota_{\rho,t-1}\le 2n^{1/r_0}\right)\\
&+I\left(\xi_{\rho,t}>n^{1/r_0}\right)=:I_{t,1}+I_{t,2}+I_{t,3},
\end{aligned}$$

where in the last inequality we use the fact $\max_{1\le t\le n}\widetilde{\xi}^\iota_{\rho,t-1}=o(n^{1/r_0})$ for small ι; see Lemma 11.2 in Owen (2001).

Note that by assumptions, η_t is independent of $\|\boldsymbol{Z}_t\|+\xi_{\rho,t-1}+\widetilde{\xi}^\iota_{\rho,t-1}$, and F'_η is bounded in a neighborhood of $F_\eta^{-1}(\alpha)$. Using the law of iterated expectations and Taylor expansion, we can show that

$$\begin{aligned}
\frac{1}{h}E\left(w_t\xi^2_{\rho,t}I_{t,i}\right)\le&\frac{1}{n^{1/a}h}E\left\{w_t\xi^2_{\rho,t}\left(\xi_{\rho,t}+\widetilde{\xi}^\iota_{\rho,t-1}\right)\right\}\\
\le&\frac{C}{n^{1/a}h}+\frac{C}{n^{1/a}h}E\widetilde{\xi}^\iota_{\rho,t-1}=o(1),
\end{aligned}$$

for $i=1,2$. Again using the boundedness of $w_t\xi^3_{\rho,t}$,

$$\frac{1}{h}E\left(w_t\xi^2_{\rho,t}I_{t,3}\right)=\frac{1}{h}E\left(w_t\xi^3_{\rho,t}\frac{1}{\xi_{\rho,t}}I\left(\xi_{\rho,t}>n^{1/r_0}\right)\right)\le\frac{C}{n^{1/r_0}h}=o(1).$$

Hence,

$$E\left(w_t\Gamma_{t,3}\right)\le\sum_{i=1}^3E\left(w_t\xi^2_{\rho,t}I_{t,i}\right)=o(1).$$

Now we can conclude that

$$E\left(w_t\Gamma_t\right) \leq \sum_{i=1}^{3} E\left(w_t\Gamma_{t,i}\right) = o(1),$$

and therefore by Markov inequality

$$\frac{1}{n}\sum_{t=1}^{n} w_t\Gamma_t \xrightarrow{P} 0.$$

Invoking statement (v) in Lemma F.3 and using the law of iterated expectations and Taylor expansion, it is easy to show that

$$\begin{aligned} E\left(w_t\Delta_t\right) \leq & Cn^{-1/a}E\left\{\frac{1}{h}g\left(\frac{F_{\eta_t}^{-1}(\alpha)-\eta_t}{h}\right)(1+|\eta_t|)\right\}E\left(\widetilde{\xi}_{\rho,t-1}^{\iota}\right) \\ = & O(n^{-1/a}) = o(1). \end{aligned}$$

Applying Markov inequality yields that

$$\frac{1}{n}\sum_{i=1}^{n} w_t\Delta_t \xrightarrow{P} 0.$$

This completes the proof. □

Lemma F.9. *For all $a \in (0, 2(1+\iota_0))$, as $n \to \infty$,*

$$\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^{a}} \left\| \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{W}_t(\boldsymbol{\upsilon},\theta^0)\boldsymbol{W}_t^T(\boldsymbol{\upsilon},\theta^0) - \Omega_1 \right\| \xrightarrow{P} 0.$$

Proof. From Lemma F.7 and Cauchy-Swarchz inequality, it suffices to show that

$$\sup_{\boldsymbol{\upsilon}\in\Theta_{\boldsymbol{\upsilon}}^{a}} \frac{1}{n}\sum_{t=1}^{n} \left\|\boldsymbol{W}_t(\boldsymbol{\upsilon},\theta^0) - \boldsymbol{W}_t\right\|^2 \xrightarrow{P} 0.$$

Clearly, we only need to verify the convergence coordinate-wisely. The convergence in the second to $(k+2)$th coordinates are known in Lemma D.4, and that in the coordinates for GARCH parameters follows by a similar argument as in the proof of Lemma F.8. It remains to prove the convergence in the first coordinate, and our proof is very similar to that in Lemma D.4.

Take some $r > a$ and apply (i) in Lemma F.3 to get

$$\frac{1}{n}\sum_{t=1}^{n}\left(\bar{W}_{t,1}(\boldsymbol{\upsilon},\theta^0) - \bar{W}_{t,1}\right)^2$$

$$
\begin{aligned}
=&\frac{1}{n}\sum_{t=1}^{n}\left\{K\left(\frac{F_\eta^{-1}(\alpha)-\eta_t-(\boldsymbol{\beta}-\boldsymbol{\beta}^0)^T\boldsymbol{z}/\sigma+(\eta_t-\eta_t(\boldsymbol{\upsilon}))}{h}\right)-K\left(\frac{F_\eta^{-1}(\alpha)-\eta_t}{h}\right)\right\}^2\\
\leq&\frac{1}{n}\sum_{t=1}^{n}\left(K\left(\frac{F_\eta^{-1}(\alpha)-\eta_t+n^{-1/a}\left\|\boldsymbol{z}\right\|/\sigma+Cn^{-1/a}\left\{\|\boldsymbol{Z}_t\|+|\eta_t|\xi_{\rho,t-1}+|\eta_t|\widetilde{\xi}^{\iota}_{\rho,t-1}\right\}}{h}\right)\right.\\
&\left.-K\left(\frac{F_\eta^{-1}(\alpha)-\eta_t-n^{-1/a}\left\|\boldsymbol{z}\right\|/\sigma-Cn^{-1/a}\left\{\|\boldsymbol{Z}_t\|+|\eta_t|\xi_{\rho,t-1}+|\eta_t|\widetilde{\xi}^{\iota}_{\rho,t-1}\right\}}{h}\right)\right)^2\\
\leq&\frac{1}{n}\sum_{t=1}^{n}\left\{K\left(\frac{F_\eta^{-1}(\alpha)-\eta_t+Cn^{-1/a+1/r}}{h}\right)-K\left(\frac{F_\eta^{-1}(\alpha)-\eta_t-Cn^{-1/a+1/r}}{h}\right)\right\}^2\\
&+\frac{4}{n}\sum_{t=1}^{n}I(\|\boldsymbol{Z}_t\|+|\eta_t|\xi_{\rho,t-1}+|\eta_t|\widetilde{\xi}^{\iota}_{\rho,t-1}>n^{1/r})\\
\leq&\frac{4}{n}\sum_{t=1}^{n}I\left(|F_\eta^{-1}(\alpha)-\eta_t|\leq Cn^{-1/a+1/r}\right)\\
&+\frac{4}{n}\sum_{t=1}^{n}I(\|\boldsymbol{Z}_t\|+|\eta_t|\xi_{\rho,t-1}+|\eta_t|\widetilde{\xi}^{\iota}_{\rho,t-1}>n^{1/r})=o_P(1).
\end{aligned}
$$

Hence the proof is complete by noting that w_t is bounded. □

LEMMA F.10. *For all $a\in(0,2+\iota_0)$,*

$$\sup_{\Theta_{\boldsymbol{\upsilon}}^a}\max_{1\leq t\leq n}\left\|\bar{\boldsymbol{W}}_t(\boldsymbol{\upsilon},\theta^0)\right\|=o(n^{1/(2+\iota_0)}).$$

PROOF. Invoking Lemma D.3, we only need to verify that

$$\max_{1\leq t\leq n}w_t\sup_{\Theta_{\boldsymbol{\upsilon}}^a}\left\|\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|=o(n^{1/(2+\iota_0)}).\tag{F.3}$$

From Lemma F.3, uniformly for small ι and $t=1,\ldots,n$,

$$
\begin{aligned}
&w_t\sup_{\Theta_{\boldsymbol{\upsilon}}^a}\left\|\frac{\partial l_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|\\
\leq&w_t\sup_{\Theta_{\boldsymbol{\upsilon}}^a}\left\|\frac{1}{h_t(\boldsymbol{\upsilon})}\frac{\partial h_t(\boldsymbol{\upsilon})}{\partial\boldsymbol{\zeta}}\right\|\left(\sup_{\Theta_{\boldsymbol{\upsilon}}^a}\eta_t^2(\boldsymbol{\upsilon})+1\right)\\
\leq&Cw_t\left(\xi^{\iota}_{\rho,t-1}+\widetilde{\xi}^{\iota}_{\rho,t-1}\right)\left(\left\{n^{-1/a}\left\|\boldsymbol{Z}_t\right\|+|\eta_t|(1+n^{-1/a}\xi_{\rho,t-1})\right\}^2+1\right)
\end{aligned}
$$

$$\leq C w_t \left(\xi_{\rho,t}^{2+\iota} + \xi_{\rho,t}^2 \widetilde{\xi}_{\rho,t-1}^{\iota} \right) (1+\eta_t^2) \leq C \widetilde{\xi}_{\rho,t-1}^{\iota} (1+\eta_t^2)$$

where the last inequality is due to the boundedness of $w_t \xi_{\rho,t}^3$. By a standard Borel-Cantelli argument, see Lemma 11.2 in Owen (2001), we can show that

$$\max_{1\leq t\leq n} \widetilde{\xi}_{\rho,t-1}^{\iota} = o(n^{\iota/(2+2\iota_0)}), \text{ and } \max_{1\leq t\leq n} \eta_t^2 = o(n^{1/(2+2\iota_0)}).$$

Statement (F.3) then follows by taking ι sufficiently small. □

Using Lemmas F.8–F.10, we give the quadratic expression of the empirical likelihood ratio function $-2\log \bar{L}(\cdot, \theta^0)$ around the true parameters $\boldsymbol{\upsilon}^0$ in the following lemma. The proof is completely analogous to that of Lemma D.5 and that of Lemma D.6, and therefore we omit the details.

LEMMA F.11. *Let $a \in (2, \min\{r_0, 2+\iota_0/(2+\iota_0), 15/7\})$. For all $\boldsymbol{\upsilon}$ such that $\left\|\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^0\right\| \leq n^{-1/a}$,*

$$\begin{aligned} -2\log \bar{L}(\boldsymbol{\upsilon}, \theta^0) =& \bar{\boldsymbol{\nu}}^T \Omega_2^T \Omega_1^{-1} \Omega_2 \bar{\boldsymbol{\nu}} + 2\bar{\boldsymbol{\nu}}^T \Omega_2^T \Omega_1^{-1} \bar{\mathbb{W}}_n + \bar{\mathbb{W}}_n^T \Omega_1^{-1} \bar{\mathbb{W}}_n \\ &+ o_P(1) + o_P(\|\bar{\boldsymbol{\nu}}\|) + o_P(\|\bar{\boldsymbol{\nu}}\|^2) \end{aligned} \tag{F.4}$$

with $\bar{\mathbb{W}}_n := \frac{1}{\sqrt{n}} \sum_{t=1}^n \bar{\boldsymbol{W}}_t(\boldsymbol{\upsilon}^0, \theta^0)$ and $\bar{\boldsymbol{\nu}} := \sqrt{n}(\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^0)$. With probability tending to one, $\bar{L}(\boldsymbol{\upsilon}, \theta^0)$ attains its maximum at some point $\widetilde{\boldsymbol{\upsilon}}$ in the interior of the ball $\left\|\boldsymbol{\upsilon} - \boldsymbol{\upsilon}^0\right\| \leq n^{-1/a}$ and

$$\sqrt{n}(\widetilde{\boldsymbol{\upsilon}} - \boldsymbol{\upsilon}^0) = -(\Omega_2^T \Omega_1^{-1} \Omega_2)^{-1} \Omega_2^T \Omega_1^{-1} \bar{\mathbb{W}}_n + o_P(1). \tag{F.5}$$

PROOF OF THEOREM 4. Substituting (F.5) into (F.4) in Lemma F.11 yields that

$$-2\log \bar{L}^P(\theta^0) = -2\log \bar{L}(\widetilde{\boldsymbol{\upsilon}}, \theta^0) = \bar{\mathbb{W}}_n^T \Omega_1^{-1/2} D \Omega_1^{-1/2} \bar{\mathbb{W}}_n + o_P(1)$$

where

$$D := I_{k+r+s+3} - \Omega_1^{-1/2} \Omega_2 (\Omega_2^T \Omega_1^{-1} \Omega_2)^{-1} \Omega_2^T \Omega_1^{-1/2}.$$

Hence the theorem follows from the facts that $\widetilde{D}$ is symmetric, idempotent,

$$\begin{aligned} & tr(I_{k+r+s+3} - \Omega_1^{-1/2} \Omega_2 (\Omega_2^T \Omega_1^{-1} \Omega_2)^{-1} \Omega_2^T \Omega_1^{-1/2}) \\ = \; & k+r+s+3 - tr((\Omega_2^T \Omega_1^{-1} \Omega_2)^{-1} \Omega_2^T \Omega_1^{-1/2} \Omega_1^{-1/2} \Omega_2) \\ = \; & k+r+s+3 - tr((\Omega_2^T \Omega_1^{-1} \Omega_2)^{-1} \Omega_2^T \Omega_1^{-1} \Omega_2) \\ = \; & k+r+s+3 - (k+r+s+2) = 1 \end{aligned}$$

and $\Omega_1^{-1/2} \bar{\mathbb{W}}_n \xrightarrow{d} N(0, I_{k+r+s+3})$ by Lemma F.7. □

REFERENCES

ADRIAN, T. AND BRUNNERMEIER, M. (2016), CoVaR, *American Economic Review* **106** 1705–1741.

CHUN, S. Y., SHAPIRO, A. AND URYASEV, S. (2012), Conditional value-at-risk and average value-at-risk: estimation and asymptotics, *Operations Research* **60** 739–756.

DAVIS, R. A. AND WU, W. (1997), M-estimation for linear regression with infinite variance, *Probability and Mathematical Statistics* **17** 1–20.

HALL, P. AND HEYDE, C. C. (1980), *Martingale Limit Theory and Its Applications.* Academic Press.

LING, S. (2007), Self-weighted and local quasi-maximum likelihood estimators for ARMA-GARCH/IGARCH models, *Journal of Econometrics* **140** 849–873.

MAMMEN, E. (1996). Empirical processes of residuals for high-dimensional linear models, *The Annals of Statistics* **24**, 307–335.

OWEN, A. (1990), Empirical Likelihood Ratio Confidence Regions, *The Annals of Statistics* **18** 90–120.

OWEN, A. (2001), *Empirical Likelihood.* Chapman & Hall/CRC, New York.

QIN, J. AND LAWLESS, J. (1994), Empirical likelihood and general estimating equations, *The Annals of Statistics* **22** 300–325.

SHERMAN, R. P. (1993), The limiting distribution of the maximum rank correlation estimator, *Econometrica* **61** 123–137.

GALEN R. SHORACK AND JON A. WELLNER (1986), *Empirical Processes with Applications to Statistics*, SIAM.

VAN DER VAART, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

WHANG, Y. J. (2006), Smoothed empirical likelihood methods for quantile regression models, *Econometric Theory* **22** 173-205.

AMSTERDAM SCHOOL OF ECONOMICS
UNIVERSITY OF AMSTERDAM
AMSTERDAM 1018 WB, THE NETHERLANDS
E-MAIL: y.he2@uva.nl

SCHOOL OF DATA SCIENCE
FUDAN UNIVERSITY
SHANGHAI 200433, CHINA
E-MAIL: oishihou@yahoo.com

DEPARTMENT OF RISK MANAGEMENT AND INSURANCE
GEORGIA STATE UNIVERSITY
ATLANTA, GA 30303, USA
E-MAIL: lpeng@gsu.edu

FACULTY OF BUSINESS AND ECONOMICS
THE UNIVERSITY OF HONG KONG
POK FU LAM ROAD, HONG KONG
E-MAIL: haipeng@hku.hk