Biometrika (2017), **xx**, x, *pp*. 1–42
Printed in Great Britain

Supplementary material for 'A robust goodness-of-fit test for generalized autoregressive conditional heteroscedastic models'

BY YAO ZHENG, WAI KEUNG LI AND GUODONG LI
Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China
yaozheng@connect.hku.hk hrntlwk@hku.hk gdli@hku.hk

SUMMARY

This supplementary material is organized as follows.

In § S1 we present additional results on the noncentrality parameter; in particular, we compare the values of the noncentrality parameter c_Ψ for different Ψ under specific local alternatives.

Section S2 contains further simulation results. Three simulation studies are carried out to verify the asymptotic distribution of $Q(M)$ and to evaluate the performance of the proposed method for selecting M.

Section S3 discusses tail index estimation in the empirical example.

In § S4 we give the proofs of Theorems 1 and 2, Corollary 1 and Lemmas A1–A3 from the main paper, as well as two auxiliary lemmas, Lemmas S1 and S2.

In § S5 we present the proofs of Proposition 2 and Theorems 3 and 4, and also introduce and prove Lemmas S3–S8.

Finally, § S6 contains the proofs of Theorems 5 and 6.

S1. ADDITIONAL RESULTS ON THE NONCENTRALITY PARAMETER

In this section, we calculate the value of the noncentrality parameter c_Ψ for local alternatives of the following null hypotheses: (i) the autoregressive conditional heteroscedastic model of order one, $y_t = \varepsilon_t h_t^{1/2}$, $h_t = \omega_0 + \alpha_0 y_{t-1}^2$, denoted by ARCH$(1)$; and (ii) the GARCH$(1,1)$ model, $y_t = \varepsilon_t h_t^{1/2}$, $h_t = \omega_0 + \alpha_0 y_{t-1}^2 + \beta_0 h_{t-1}$. Three types of departures, $s_{t,n} = G(|y_{t-2,n}|)$, $|y_{t-2,n}|$ and $y_{t-2,n}^2$, are considered, and four transformations, $\Psi(x) = G(x)$, $\mathrm{sgn}(x-1)$, x and x^2, are studied. The innovation distributions that we consider include the zero-mean normal distribution and Student's t_7, t_5, t_3, $t_{2\cdot5}$ and t_1 distributions, which are standardized such that $\mathrm{median}(|\varepsilon_t|) = 1$. In addition, we consider the innovations resampled from the residuals of the fitted GARCH$(1,1)$ model in § S6, and such cases are denoted by $\{\hat\varepsilon_t\}$ in Tables S1–S4. For the sign-based test, $Q_{\mathrm{sgn}}(M)$, although the corresponding transformation $\Psi(x) = \mathrm{sgn}(x-1)$ is not differentiable at $x = 1$, we can verify that the result of Theorem 4 still holds with κ_Ψ replaced by $2g(1)$. We focus on the value of c_Ψ corresponding to the least absolute deviations estimator (Peng & Yao, 2003) and approximate the quantities in Υ_Ψ and Σ_Ψ by sample averages based on a generated sequence $\{y_1, \ldots, y_n\}$ with $n = 100\,000$. We set $M = 6$ and consider the following parameter settings: $\alpha_0 = 0{\cdot}03, 0{\cdot}5$ and $0{\cdot}9$ for the ARCH(1) model, and $(\alpha_0, \beta_0) = (0{\cdot}03, 0{\cdot}2), (0{\cdot}3, 0{\cdot}2)$ and $(0{\cdot}03, 0{\cdot}6)$ for the GARCH$(1,1)$ model; for all these cases, ω_0 is set to 1.

Table S1. *Noncentrality parameter c_Ψ $(\times 10)$ under different local alternatives of the* ARCH(1) *model with $(\omega_0, \alpha_0) = (1, 0{\cdot}5)$, for $\Psi(x) = G(x)$, $\mathrm{sgn}(x-1)$, x and x^2*

	$s_{t,n} = G(\|y_{t-2,n}\|)$				$s_{t,n} = \|y_{t-2,n}\|$				$s_{t,n} = y^2_{t-2,n}$			
	G	sgn	x	x^2	G	sgn	x	x^2	G	sgn	x	x^2
$\{\hat{\varepsilon}_t\}$	0·025	0·013	1E-04	6E-09	1·251	0·573	0·020	5E-05	472	177	73	1
t_1	0·003	0·002			2·117	0·810			2639771	854755		
$t_{2\cdot5}$	0·020	0·010	0·002		1·987	0·889	0·538		224222	91791	94406	
t_3	0·024	0·012	0·005		1·581	0·705	0·699		3038	1168	1794	
t_5	0·032	0·016	0·018	0·002	1·381	0·621	1·193	0·234	444	170	503	162
t_7	0·037	0·017	0·027	0·006	1·322	0·587	1·361	0·519	211	79	271	151
Normal	0·052	0·023	0·053	0·028	1·262	0·550	1·601	1·196	84	33	121	112

Small numbers are written in standard form, e.g., 1E-04 means 1×10^{-4}.

Table S2. *Noncentrality parameter c_Ψ $(\times 10^2)$ under different local alternatives of the* GARCH$(1,1)$ *model with $(\omega_0, \alpha_0, \beta_0) = (1, 0{\cdot}3, 0{\cdot}2)$, for $\Psi(x) = G(x)$, $\mathrm{sgn}(x-1)$, x and x^2*

	$s_{t,n} = G(\|y_{t-2,n}\|)$				$s_{t,n} = \|y_{t-2,n}\|$				$s_{t,n} = y^2_{t-2,n}$			
	G	sgn	x	x^2	G	sgn	x	x^2	G	sgn	x	x^2
$\{\hat{\varepsilon}_t\}$	0·08	0·05	3E-04	2E-06	1·41	0·82	0·01	1E-04	24·90	13·47	0·45	7E-04
t_1	3E-05	2E-05			2E-03	1E-03			99·52	70·96		
$t_{2\cdot5}$	0·05	0·03	3E-03		1·17	0·70	0·13		31·38	17·89	8·45	
t_3	0·07	0·04	0·01		1·32	0·77	0·27		26·10	14·51	12·15	
t_5	0·10	0·05	0·03	3E-03	1·42	0·78	0·72	0·11	17·07	8·90	16·86	3·98
t_7	0·11	0·06	0·05	0·01	1·42	0·77	0·93	0·26	14·35	7·24	16·96	7·91
Normal	0·15	0·07	0·10	0·04	1·36	0·69	1·25	0·74	9·32	4·43	13·62	12·80

Table S3. *The transformation Ψ which results in the largest c_Ψ under different local alternatives of the* ARCH(1) *model with $\alpha_0 = 0{\cdot}03, 0{\cdot}5$ or $0{\cdot}9$ and $\omega_0 = 1$*

	$s_{t,n} = G(\|y_{t-2,n}\|)$			$s_{t,n} = \|y_{t-2,n}\|$			$s_{t,n} = y^2_{t-2,n}$		
	0·03	0·5	0·9	0·03	0·5	0·9	0·03	0·5	0·9
$\{\hat{\varepsilon}_t\}$	G	G	G	G	G	G	x	G	G
t_1	G	G	G	G	G	G	G	G	G
$t_{2\cdot5}$	G	G	G	G	G	G	x	G	G
t_3	G	G	G	x	G	G	x	G	G
t_5	G	G	G	x	G	G	x	x	G
t_7	x	G	G	x	x	G	x	x	x
Normal	x	x	G	x	x	x	x^2	x	x^2

Tables S1 and S2 report the values of c_Ψ for the ARCH(1) model with $\alpha_0 = 0{\cdot}5$ and the GARCH$(1,1)$ model with $(\alpha_0, \beta_0) = (0{\cdot}3, 0{\cdot}2)$, respectively. It can be seen that $G(x)$ dominates $\mathrm{sgn}(x-1)$, and x dominates x^2 in all cases. Moreover, $G(x)$ dominates all of the transformations for heavy-tailed innovations, and even for moderate-tailed innovations, i.e., $E(\varepsilon_t^4) < \infty$, when the departure $s_{t,n}$ is $G(|y_{t-2,n}|)$ or $|y_{t-2,n}|$. Tables S3 and S4 give the transformation that leads to the largest value of c_Ψ among the four transformations $\Psi(x) = G(x)$, $\mathrm{sgn}(x-1)$, x and x^2 in each parameter setting. We summarize the findings as follows. Firstly, $G(x)$ is always the best transformation when $s_{t,n} = G(|y_{t-2,n}|)$, which is probably due to the matching of the transformation and the form of the departure. Secondly, $G(x)$ generally achieves more favourable performance when the value of α_0 or β_0 is larger. Thirdly, for the case where the innovations are resampled from the residuals of the fitted GARCH$(1,1)$ model, $G(x)$ dominates all of the other transformations except for one case.

Table S4. *The transformation Ψ which results in the largest c_Ψ under different local alternatives of the* GARCH$(1,1)$ *model with* $(\alpha_0, \beta_0) = (0{\cdot}03, 0{\cdot}2), (0{\cdot}3, 0{\cdot}2)$ *or* $(0{\cdot}03, 0{\cdot}6)$ *and* $\omega_0 = 1$

	$s_{t,n} = G(\lvert y_{t-2,n}\rvert)$			$s_{t,n} = \lvert y_{t-2,n}\rvert$			$s_{t,n} = y_{t-2,n}^2$		
	(0·03, 0·2)	(0·3, 0·2)	(0·03, 0·6)	(0·03, 0·2)	(0·3, 0·2)	(0·03, 0·6)	(0·03, 0·2)	(0·3, 0·2)	(0·03, 0·6)
$\{\hat{\varepsilon}_t\}$	G	G	G	G	G	G	G	G	G
t_1	G	G	G	G	G	G	G	G	G
$t_{2\cdot5}$	G	G	G	G	G	G	x	G	G
t_3	G	G	G	G	G	G	x	G	x
t_5	G	G	G	G	G	G	x	G	x
t_7	G	G	G	G	G	G	x	x	x
Normal	G	G	G	G	G	x	x^2	x	x^2

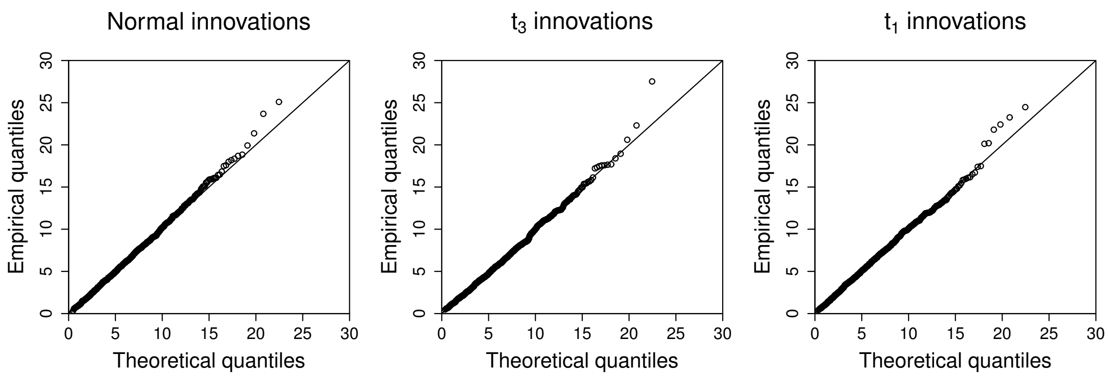

Fig. S1. Q-Q plots for $Q(6)$ under H_0 against the χ^2_6 distribution with 45° reference lines, for sample size $n = 1000$ and $\{\varepsilon_t\}$ following three different distributions.

S2. Additional simulation studies

This section reports on three additional simulation experiments. The first two experiments verify the asymptotic results of $Q(M)$ under the null hypothesis and the local alternatives. The third experiment evaluates the performance of the proposed Bayesian information criterion-type method for selecting the order M for different joint test statistics. All estimation methods are the same as those in § 5 of the main paper, unless specified otherwise.

First, to assess the performance of the chi-squared approximation for the asymptotic null distribution of $Q(M)$, we generate 1000 replications with sample size $n = 1000$ from

$$y_t = \varepsilon_t h_t^{1/2}, \quad h_t = 0{\cdot}01 + 0{\cdot}03 y_{t-1}^2 + 0{\cdot}2 h_{t-1},$$

where $\{\varepsilon_t\}$ follow the normal distribution with mean zero or Student's t_1 or t_3 distribution, standardized such that $\mathrm{median}(|\varepsilon_t|) = 1$. Figure S1 shows that the empirical quantiles of $Q(6)$ well match the quantiles of the chi-squared distribution with six degrees of freedom, i.e., χ^2_6. Particularly, the points in the upper tails lie near the 45° reference lines, indicating close agreement between the empirical and nominal sizes.

Second, to verify the asymptotic results of $Q(M)$ under the local alternatives, we construct $\tilde{Q}(M) = (n^{1/2}\hat{\rho} - \hat{\varUpsilon})^{\mathrm{T}}\hat{\Sigma}^{-1}(n^{1/2}\hat{\rho} - \hat{\varUpsilon})$, where $\hat{\varUpsilon}$ and $\hat{\Sigma}$ are consistent estimators of $\varUpsilon$ and Σ, respectively. By Theorem 3, we have that $\tilde{Q}(M)$ is asymptotically distributed as χ^2_M under H_{1n}. We consider the following local alternatives:

$$y_{t,n} = \varepsilon_t h_{t,n}^{1/2}, \quad h_{t,n} = 0{\cdot}01 + 0{\cdot}03 y_{t-1,n}^2 + 0{\cdot}2 h_{t-1,n} + n^{-1/2} s_{t,n},$$

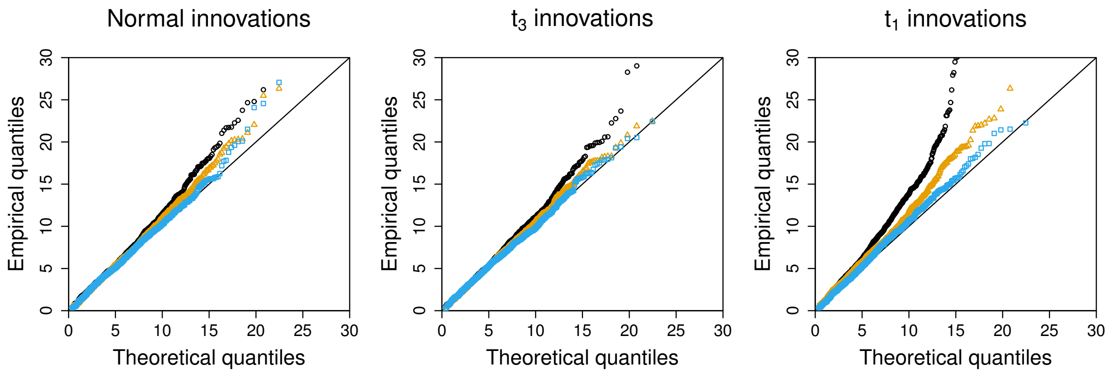

Fig. S2. Q-Q plots for $\tilde{Q}(6)$ under H_{1n} against the χ^2_6 distribution with 45° reference lines, for three sample sizes, $n = 1000$ (circles), $10\,000$ (triangles) and $50\,000$ (squares), and with $\{\varepsilon_t\}$ following three different distributions.

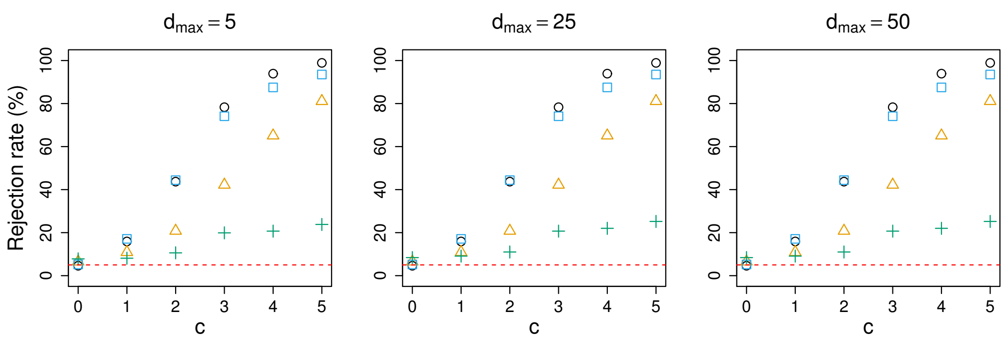

Fig. S3. Rejection rates (%) of four automatic goodness-of-fit tests: Q^{A} (circles), $Q^{\mathrm{A}}_{\mathrm{sgn}}$ (triangles), $Q^{\mathrm{A}}_{\mathrm{abs}}$ (squares) and $Q^{\mathrm{A}}_{\mathrm{sqr}}$ (pluses), for $d_{\max} = 5$, 25 and 50. The horizontal lines indicate the 5% nominal level.

where $s_{t,n} = 2y^2_{t-2,n}$ and $\{\varepsilon_t\}$ are specified as in the previous experiment. As $\varUpsilon = 6\kappa(DJ^{-1}\lambda - V)$, we can estimate it by $\hat{\varUpsilon} = 6\hat{\kappa}(\hat{D}\hat{J}^{-1}\hat{\lambda} - \hat{V})$, where $\hat{\kappa}$, $\hat{D}$ and $\hat{J}$ are the consistent estimators used for constructing $\hat{\Sigma}$ in § 2 of the main paper. In addition, for the aforementioned model, we can show that $r_{t,n} = 2h^{-1}_{t,n}(\theta_0)\partial h_{t-1,n}(\theta_0)/\partial\alpha$. Let $\tilde{r}_{t,n}(\theta) = 2\tilde{h}^{-1}_{t,n}(\theta)\partial\tilde{h}_{t-1,n}(\theta)/\partial\alpha$, and write $\hat{r}_{t,n} = \tilde{r}_{t,n}(\hat{\theta}_n)$. Then $\hat{\lambda} = n^{-1}\sum_{t=1}^{n}\hat{r}_{t,n}\hat{h}^{-1}_{t,n}\partial\tilde{h}_{t,n}(\hat{\theta}_n)/\partial\theta$ and $\hat{V} = (\hat{v}_1, \ldots, \hat{v}_M)^{\mathrm{T}}$, where $\hat{v}_k = n^{-1}\sum_{t=k+1}^{n}\{0{\cdot}5 - \hat{G}_n(|\hat{\varepsilon}_{t-k}|)\}\hat{r}_{t,n}$, are consistent estimators of λ and V, respectively. Thus, $\hat{\varUpsilon} = \varUpsilon + o_{\mathrm{p}}(1)$. We generate 1000 replications with sample sizes $n = 1000$, $10\,000$ and $50\,000$. Figure S2 displays the Q-Q plots of $\tilde{Q}(6)$ against the χ^2_6 distribution. Convergence to the reference lines can be observed as n increases, although the rates are relatively slow. Moreover, the convergence rate in the case of Student's t_1-distributed innovations seems slightly slower than for the other two innovation distributions, probably due to the extreme heavy-tailedness of the Student's t_1 distribution.

Third, to further investigate the performance of the proposed Bayesian information criterion-type order selection method, we apply it to four goodness-of-fit test statistics, namely $Q(M)$, $Q_{\mathrm{sgn}}(M)$, $Q_{\mathrm{abs}}(M)$ and $Q_{\mathrm{sqr}}(M)$, using the data generated in the second simulation experiment in § 5 of the main article. Henceforth we use a superscript A to indicate that M is selected

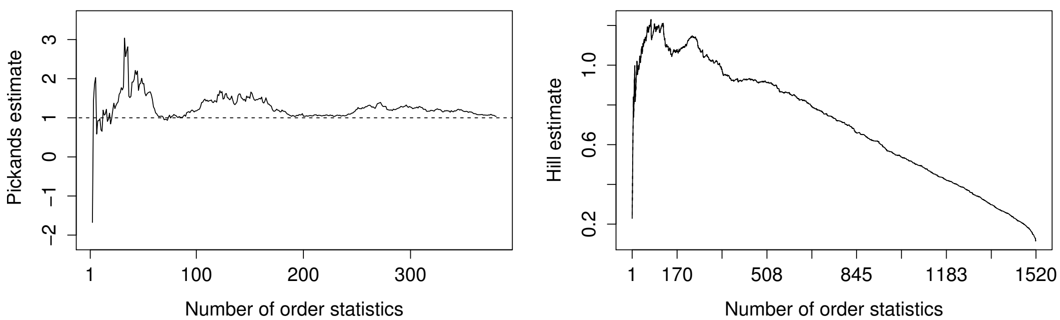

Fig. S4. Pickands plot (left) and Hill plot (right) for the tail index of squared residuals of the fitted GARCH$(1, 1)$ model.

automatically. Figure S3 plots the rejection rates, from which we have the following findings. First, the performance of the proposed method is insensitive to the value of $d_{\max}$; second, the size of the test is fairly accurate except for $Q^{\rm A}_{\rm sqr}$, which is oversized probably because of the infinite fourth-order moment of the Student's t_3 distribution, as $Q_{\rm sqr}(M)$ requires $E(\varepsilon_t^4) < \infty$; third, the power of the four tests can be ordered as $Q^{\rm A} > Q^{\rm A}_{\rm abs} > Q^{\rm A}_{\rm sgn} > Q^{\rm A}_{\rm sgn}$, which is as expected since the innovations follow the heavy-tailed Student's t_3 distribution; fourth, the power increases as c becomes larger, and the power of these tests for $c = 2$ is similar to that exhibited in Fig. 1(a) of the main paper, where a fixed M was employed.

S3. Tail index estimation in the empirical example

Figure S4 presents the Pickands and Hill estimates for the tail index of the squared residuals of the fitted GARCH$(1, 1)$ model in § 6 of the main article. While the Hill estimates fail to converge as the number of order statistics increases, the Pickands plot indicates that the tail index of $\{\hat{\varepsilon}_t^2\}$ is greater than 1 and less than 2, suggesting that $E(\varepsilon_t^2) < \infty$ and $E(\varepsilon_t^4) = \infty$; see Resnick (2007) for a more detailed discussion of tail index estimation.

S4. Proofs of Theorems 1 and 2 and Corollary 1

S4·1. *Proofs of Theorems* 1 *and* 2 *and Corollary* 1

In this section we give the proofs of Theorems 1 and 2, Corollary 1 and Lemmas A1–A3 in the main paper. Two auxiliary lemmas are also presented: Lemma S1 summarizes some existing results that are used repeatedly in our proofs, and Lemma S2 is used to establish Lemma A1.

Throughout the proofs, we let $C > 0$ and $0 < \rho < 1$ be generic constants which may take different values at different occurrences. Denote by $\|\cdot\|$ the Euclidean norm for a vector and the spectral norm for a square matrix. For a random variable X, let $\|X\|_m$ be its L_m-norm, where $m \geqslant 1$, i.e., $\|X\|_m = \{E(|X|^m)\}^{1/m}$.

Proof of Theorem 1. To prove the theorem, we first establish two intermediate results:

$$n^{-1/2} \sum_{t=k+1}^{n} \Big\{ \hat{G}_n(|\hat{\varepsilon}_t|)\hat{G}_n(|\hat{\varepsilon}_{t-k}|) - G_n(|\hat{\varepsilon}_t|)G_n(|\hat{\varepsilon}_{t-k}|) \Big\} - \kappa d_0^{*{\rm T}} n^{1/2}(\hat{\theta}_n - \theta_0) = o_{\rm p}(1) \quad \text{(S1)}$$

and

$$n^{-1/2}\sum_{t=k+1}^{n}\left\{G_n(|\hat\varepsilon_t|)G_n(|\hat\varepsilon_{t-k}|)-G_n(|\varepsilon_t|)G_n(|\varepsilon_{t-k}|)\right\} + 0{\cdot}5\kappa(d_0^*+d_k^*)^{\mathrm T}n^{1/2}(\hat\theta_n-\theta_0)=o_{\mathrm p}(1) \tag{S2}$$

for any positive integer k, where $d_k^*=E\{G(|\varepsilon_{t-k}|)h_t^{-1}\partial h_t(\theta_0)/\partial\theta\}$ for $k\geqslant 1$.

We begin by proving (S1). First notice that Assumption 3 implies

$$L=\sup_{0\leqslant x<\infty} xg(x)<\infty. \tag{S3}$$

Let $W_t=G_n(|\hat\varepsilon_t|)+|\hat\varepsilon_t|g(|\hat\varepsilon_t|)d_0^{*\mathrm T}(\hat\theta_n-\theta_0)$. By (S3) and the fact that $n^{1/2}(\hat\theta_n-\theta_0)=O_{\mathrm p}(1)$, we have $\max_{1\leqslant t\leqslant n}|W_t|=O_{\mathrm p}(1)$. Moreover, applying Lemma A1 with $w_t\equiv 1$, we have $n^{1/2}\max_{1\leqslant t\leqslant n}|\hat G_n(|\hat\varepsilon_t|)-W_t|=o_{\mathrm p}(1)$. As a result,

$$n^{-1/2}\sum_{t=k+1}^{n}\left\{\hat G_n(|\hat\varepsilon_t|)\hat G_n(|\hat\varepsilon_{t-k}|)-W_tW_{t-k}\right\}=o_{\mathrm p}(1).$$

Hence, to prove (S1), it remains to show that

$$n^{-1/2}\sum_{t=k+1}^{n}\{W_tW_{t-k}-G_n(|\hat\varepsilon_t|)G_n(|\hat\varepsilon_{t-k}|)\}-\kappa d_0^{*\mathrm T}n^{1/2}(\hat\theta_n-\theta_0)=o_{\mathrm p}(1). \tag{S4}$$

By the Dvoretzky–Kiefer–Wolfowitz inequality (Dvoretzky et al., 1956; Serfling, 1980; Massart, 1990), we have

$$n^{1/2}\sup_{0\leqslant x<\infty}|G_n(x)-G(x)|=O_{\mathrm p}(1), \tag{S5}$$

which, in conjunction with (S3), implies

$$n^{-1/2}\sum_{t=k+1}^{n}|\hat\varepsilon_{t-k}|g(|\hat\varepsilon_{t-k}|)\left\{G_n(|\hat\varepsilon_t|)-G(|\hat\varepsilon_t|)\right\}=O_{\mathrm p}(1). \tag{S6}$$

For any $A>0$, by (S19), (S21) and Lemma S1 we have

$$\sup_{\|u\|\leqslant A}\sup_{0\leqslant x<\infty}\frac{1}{n}\sum_{t=1}^{n}\left|G\{x\tilde Z_t(u)\}-G(x)\right|\leqslant\frac{C}{n}\sum_{t=1}^{n}\rho^t\zeta_0+\frac{C}{n^{3/2}}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\right\| = O_{\mathrm p}(n^{-1/2}),$$

where $\tilde Z_t(u)$ is defined in (S13). This, together with the fact that $n^{1/2}(\hat\theta_n-\theta_0)=O_{\mathrm p}(1)$, implies

$$\frac{1}{n}\sum_{t=k+1}^{n}\left|G(|\hat\varepsilon_t|)-G(|\varepsilon_t|)\right|=O_{\mathrm p}(n^{-1/2}). \tag{S7}$$

It then follows from (S3) and (S7) that

$$\frac{1}{n}\sum_{t=k+1}^{n}|\hat\varepsilon_{t-k}|g(|\hat\varepsilon_{t-k}|)\left|G(|\hat\varepsilon_t|)-G(|\varepsilon_t|)\right|\leqslant\frac{L}{n}\sum_{t=k+1}^{n}\left|G(|\hat\varepsilon_t|)-G(|\varepsilon_t|)\right|=o_{\mathrm p}(1). \tag{S8}$$

By (S15) and a method similar to that used for (S19), we can show that

$$\max_{\log n \leqslant t \leqslant n} \sup_{\theta \in \Theta} \left| \frac{h_t^{1/2}(\theta)}{\tilde{h}_t^{1/2}(\theta)} - 1 \right| \leqslant \max_{\log n \leqslant t \leqslant n} C\rho^t \zeta_0 \leqslant C n^{-\log(1/\rho)} \zeta_0;$$

then, by arguments similar to those for (S37), we can further obtain that

$$\sup_{0 \leqslant x < \infty} \max_{\log n \leqslant t \leqslant n} \sup_{\theta \in \Theta} \left| \frac{x h_t^{1/2}(\theta)}{\tilde{h}_t^{1/2}(\theta)} g \left\{ \frac{x h_t^{1/2}(\theta)}{\tilde{h}_t^{1/2}(\theta)} \right\} - x g(x) \right| = o_{\mathrm{p}}(1).$$

This, together with (S3), yields

$$\sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \Big| |\tilde{\varepsilon}_t(\theta)| g\{|\tilde{\varepsilon}_t(\theta)|\} - |\varepsilon_t(\theta)| g\{|\varepsilon_t(\theta)|\} \Big| = o_{\mathrm{p}}(1). \tag{S9}$$

In view of (S9) and the result in (S37), we have

$$\frac{1}{n} \sum_{t=k+1}^{n} \big\{ |\hat{\varepsilon}_{t-k}| g(|\hat{\varepsilon}_{t-k}|) - |\varepsilon_{t-k}| g(|\varepsilon_{t-k}|) \big\} G(|\varepsilon_t|) = o_{\mathrm{p}}(1). \tag{S10}$$

By (S6), (S8), (S10) and the ergodic theorem, it can be shown that

$$\begin{aligned}
&\frac{1}{n} \sum_{t=k+1}^{n} |\hat{\varepsilon}_{t-k}| g(|\hat{\varepsilon}_{t-k}|) G_n(|\hat{\varepsilon}_t|) - 0{\cdot}5\kappa \\
&= \frac{1}{n} \sum_{t=k+1}^{n} |\hat{\varepsilon}_{t-k}| g(|\hat{\varepsilon}_{t-k}|) \{G_n(|\hat{\varepsilon}_t|) - G(|\hat{\varepsilon}_t|)\} + \frac{1}{n} \sum_{t=k+1}^{n} |\hat{\varepsilon}_{t-k}| g(|\hat{\varepsilon}_{t-k}|) \{G(|\hat{\varepsilon}_t|) - G(|\varepsilon_t|)\} \\
&\quad + \frac{1}{n} \sum_{t=k+1}^{n} \{ |\hat{\varepsilon}_{t-k}| g(|\hat{\varepsilon}_{t-k}|) - |\varepsilon_{t-k}| g(|\varepsilon_{t-k}|) \} G(|\varepsilon_t|) + \frac{1}{n} \sum_{t=k+1}^{n} |\varepsilon_{t-k}| g(|\varepsilon_{t-k}|) G(|\varepsilon_t|) \\
&\quad - 0{\cdot}5\kappa \\
&= o_{\mathrm{p}}(1).
\end{aligned} \tag{S11}$$

Similarly we can show that $n^{-1} \sum_{t=k+1}^{n} |\hat{\varepsilon}_t| g(|\hat{\varepsilon}_t|) G_n(|\hat{\varepsilon}_{t-k}|) - 0{\cdot}5\kappa = o_{\mathrm{p}}(1)$ and, moreover, it follows immediately from (S3) that $n^{-3/2} \sum_{t=k+1}^{n} |\hat{\varepsilon}_t| g(|\hat{\varepsilon}_t|) |\hat{\varepsilon}_{t-k}| g(|\hat{\varepsilon}_{t-k}|) = o_{\mathrm{p}}(1)$. These, together with (S11), yield (S4), and the proof of (S1) is complete.

Next we prove (S2). Observe that

$$\begin{aligned}
n^{-1/2} &\sum_{t=k+1}^{n} \big\{ G_n(|\hat{\varepsilon}_t|) G_n(|\hat{\varepsilon}_{t-k}|) - G_n(|\varepsilon_t|) G_n(|\varepsilon_{t-k}|) \big\} + 0{\cdot}5\kappa (d_0^* + d_k^*)^{\mathrm{T}} n^{1/2} (\hat{\theta}_n - \theta_0) \\
&= B_{1n} + B_{2n} + B_{3n} + B_{4n},
\end{aligned}$$

where

$$B_{1n} = n^{-1/2} \sum_{t=k+1}^{n} \{G_n(|\hat{\varepsilon}_t|) - G_n(|\varepsilon_t|)\} G(|\varepsilon_{t-k}|) + 0{\cdot}5\kappa d_k^{*\mathrm{T}} n^{1/2}(\hat{\theta}_n - \theta_0),$$

$$B_{2n} = n^{-1/2} \sum_{t=k+1}^{n} \{G_n(|\hat{\varepsilon}_t|) - G_n(|\varepsilon_t|)\} \{G_n(|\hat{\varepsilon}_{t-k}|) - G(|\varepsilon_{t-k}|)\},$$

$$B_{3n} = n^{-1/2} \sum_{t=1}^{n-k} \{G_n(|\hat{\varepsilon}_t|) - G_n(|\varepsilon_t|)\} G(|\varepsilon_{t+k}|) + 0{\cdot}5\kappa d_0^{*\mathrm{T}} n^{1/2}(\hat{\theta}_n - \theta_0),$$

$$B_{4n} = n^{-1/2} \sum_{t=1}^{n-k} \{G_n(|\hat{\varepsilon}_t|) - G_n(|\varepsilon_t|)\} \{G_n(|\varepsilon_{t+k}|) - G(|\varepsilon_{t+k}|)\}.$$

Applying Lemma A1 with $w_t = G(|\varepsilon_{t-k}|)$, we have

$$\sup_{0 \leqslant x < \infty} \left| n^{-1/2} \sum_{t=1}^{n} G(|\varepsilon_{t-k}|)\{I(|\varepsilon_t| < x) - I(|\hat{\varepsilon}_t| < x)\} + 0{\cdot}5xg(x) d_k^{*\mathrm{T}} n^{1/2}(\hat{\theta}_n - \theta_0) \right| = o_{\mathrm{p}}(1).$$

Thus,

$$\begin{aligned} B_{1n} &= \frac{1}{n}\sum_{j=1}^{n} n^{-1/2} \sum_{t=k+1}^{n} G(|\varepsilon_{t-k}|)\{I(|\varepsilon_t| < |\varepsilon_j|) - I(|\hat{\varepsilon}_t| < |\varepsilon_j|)\} + 0{\cdot}5\kappa d_k^{*\mathrm{T}} n^{1/2}(\hat{\theta}_n - \theta_0) \\ &\leqslant \sup_{0 \leqslant x < \infty} \left| n^{-1/2} \sum_{t=1}^{n} G(|\varepsilon_{t-k}|) \{I(|\varepsilon_t| < x) - I(|\hat{\varepsilon}_t| < x)\} + 0{\cdot}5xg(x) d_k^{*\mathrm{T}} n^{1/2}(\hat{\theta}_n - \theta_0) \right| \\ &\quad + 0{\cdot}5 \left| \frac{1}{n}\sum_{j=1}^{n} |\varepsilon_j| g(|\varepsilon_j|) - \kappa \right| \left| d_k^{*\mathrm{T}} n^{1/2}(\hat{\theta}_n - \theta_0) \right| + o_{\mathrm{p}}(1) \\ &= o_{\mathrm{p}}(1). \end{aligned}$$

We decompose B_{2n} into four parts:

$$B_{2n} = B_{21n} + B_{22n} + B_{23n} + B_{24n},$$

where

$$B_{21n} = n^{-1/2} \sum_{t=k+1}^{n} \{G(|\hat{\varepsilon}_t|) - G(|\varepsilon_t|)\}\{G_n(|\hat{\varepsilon}_{t-k}|) - G(|\hat{\varepsilon}_{t-k}|)\},$$

$$B_{22n} = n^{-1/2} \sum_{t=k+1}^{n} \{G(|\hat{\varepsilon}_t|) - G(|\varepsilon_t|)\}\{G(|\hat{\varepsilon}_{t-k}|) - G(|\varepsilon_{t-k}|)\},$$

$$B_{23n} = n^{-1/2} \sum_{t=k+1}^{n} \big[\{G_n(|\hat{\varepsilon}_t|) - G(|\hat{\varepsilon}_t|)\} - \{G_n(|\varepsilon_t|) - G(|\varepsilon_t|)\}\big]\{G_n(|\hat{\varepsilon}_{t-k}|) - G(|\hat{\varepsilon}_{t-k}|)\},$$

$$B_{24n} = n^{-1/2} \sum_{t=k+1}^{n} \big[\{G_n(|\hat{\varepsilon}_t|) - G(|\hat{\varepsilon}_t|)\} - \{G_n(|\varepsilon_t|) - G(|\varepsilon_t|)\}\big]\{G(|\hat{\varepsilon}_{t-k}|) - G(|\varepsilon_{t-k}|)\}.$$

By (S5) and (S7), we have

$$
\begin{aligned}
|B_{21n}| &\leqslant n^{-1/2} \sum_{t=k+1}^{n} \left|G(|\hat{\varepsilon}_t|) - G(|\varepsilon_t|)\right| \sup_{0\leqslant x<\infty} |G_n(x) - G(x)| = o_{\rm p}(1), \\
|B_{23n}| &\leqslant 2n^{1/2} \sup_{0\leqslant x<\infty} |G_n(x) - G(x)| \sup_{0\leqslant x<\infty} |G_n(x) - G(x)| = o_{\rm p}(1)
\end{aligned}
$$

and

$$
|B_{24n}| \leqslant 2n^{-1/2} \sup_{0\leqslant x<\infty} |G_n(x) - G(x)| \sum_{t=k+1}^{n} \left|G(|\hat{\varepsilon}_{t-k}|) - G(|\varepsilon_{t-k}|)\right| = o_{\rm p}(1).
$$

By a method similar to that for (S7), we can further show that $B_{22n} = o_{\rm p}(1)$. Consequently, $B_{2n} = o_{\rm p}(1)$.

Using Lemma A2 with $w_t = G(|\varepsilon_{t+k}|)$ and the fact that $E\{G(|\varepsilon_{t+k}|)\} = 0{\cdot}5$, we have

$$
\sup_{0\leqslant x<\infty} \left| n^{-1/2} \sum_{t=1}^{n} G(|\varepsilon_{t+k}|)\{I(|\varepsilon_t| < x) - I(|\hat{\varepsilon}_t| < x)\} + 0{\cdot}5 x g(x) d_0^{*{\rm T}} n^{1/2}(\hat{\theta}_n - \theta_0) \right| = o_{\rm p}(1).
$$

As a result,

$$
\begin{aligned}
B_{3n} &= \frac{1}{n}\sum_{j=1}^{n} n^{-1/2} \sum_{t=1}^{n-k} G(|\varepsilon_{t+k}|)\{I(|\varepsilon_t| < |\varepsilon_j|) - I(|\hat{\varepsilon}_t| < |\varepsilon_j|)\} + 0{\cdot}5\kappa d_0^{*{\rm T}} n^{1/2}(\hat{\theta}_n - \theta_0) \\
&\leqslant \sup_{0\leqslant x<\infty} \left| n^{-1/2} \sum_{t=1}^{n} G(|\varepsilon_{t+k}|)\{I(|\varepsilon_t| < x) - I(|\hat{\varepsilon}_t| < x)\} + 0{\cdot}5 x g(x) d_0^{*{\rm T}} n^{1/2}(\hat{\theta}_n - \theta_0) \right| \\
&\quad + 0{\cdot}5 \left| \frac{1}{n}\sum_{j=1}^{n} |\varepsilon_j|\, g(|\varepsilon_j|) - \kappa \right| \left| d_0^{*{\rm T}} n^{1/2}(\hat{\theta}_n - \theta_0) \right| + o_{\rm p}(1) \\
&= o_{\rm p}(1).
\end{aligned}
$$

By a method similar to that used for B_{2n}, it can be readily verified that $B_{4n} = o_{\rm p}(1)$. Thus, we complete the proof of (S2).

Finally, observe that $\sum_{t=1}^{n}\{\hat{G}_n(|\hat{\varepsilon}_t|) - 0{\cdot}5\} = O(1)$ and, consequently,

$$
n^{1/2}\hat{\gamma}_k = n^{-1/2} \sum_{t=k+1}^{n} \left\{ \hat{G}_n(|\hat{\varepsilon}_t|)\hat{G}_n(|\hat{\varepsilon}_{t-k}|) - 0{\cdot}25 \right\} + O(n^{-1/2}),
$$

where

$$\begin{aligned}
&n^{-1/2}\sum_{t=k+1}^{n}\Big\{\hat{G}_n(|\hat{\varepsilon}_t|)\hat{G}_n(|\hat{\varepsilon}_{t-k}|)-0{\cdot}25\Big\}\\
&\quad= n^{-1/2}\sum_{t=k+1}^{n}\Big\{\hat{G}_n(|\hat{\varepsilon}_t|)\hat{G}_n(|\hat{\varepsilon}_{t-k}|)-G_n(|\hat{\varepsilon}_t|)G_n(|\hat{\varepsilon}_{t-k}|)\Big\}\\
&\qquad+ n^{-1/2}\sum_{t=k+1}^{n}\{G_n(|\hat{\varepsilon}_t|)G_n(|\hat{\varepsilon}_{t-k}|)-G_n(|\varepsilon_t|)G_n(|\varepsilon_{t-k}|)\}\\
&\qquad+ n^{-1/2}\sum_{t=k+1}^{n}\{G_n(|\varepsilon_t|)G_n(|\varepsilon_{t-k}|)-G(|\varepsilon_t|)G(|\varepsilon_{t-k}|)\}\\
&\qquad+ n^{-1/2}\sum_{t=k+1}^{n}\{G(|\varepsilon_t|)G(|\varepsilon_{t-k}|)-0{\cdot}25\}.
\end{aligned}$$

It follows from (S1), (S2) and Lemma A3 that

$$n^{1/2}\hat{\gamma}_k = 0{\cdot}5\kappa(d_0^*-d_k^*)^{\mathrm{T}}n^{1/2}(\hat{\theta}_n-\theta_0)+n^{1/2}\gamma_k+o_{\mathrm{p}}(1), \tag{S12}$$

where

$$\gamma_k=\frac{1}{n}\sum_{t=k+1}^{n}\{G(|\varepsilon_t|)-0{\cdot}5\}\{G(|\varepsilon_{t-k}|)-0{\cdot}5\}.$$

Using the fact that $n^{1/2}\max_{1\leqslant t\leqslant n}|\hat{G}_n(|\hat{\varepsilon}_t|)-W_t|=o_{\mathrm{p}}(1)$, together with (S5) and (S7), we can show that $\hat{\gamma}_0=\gamma_0+o_{\mathrm{p}}(1)=1/12+o_{\mathrm{p}}(1)$. Thus, we complete the proof by Slutsky's lemma, the martingale central limit theorem and the Cramér–Wold device. □

Proof of Theorem 2. Let $\gamma_k^{\Psi}=n^{-1}\sum_{t=k+1}^{n}\{\Psi(|\varepsilon_t|)-\mu_\Psi\}\{\Psi(|\varepsilon_{t-k}|)-\mu_\Psi\}$ for $k\geqslant 0$. Note that $|\hat{\varepsilon}_t|=|y_t|/\tilde{h}_t^{1/2}(\hat{\theta}_n)$. By Taylor expansions and Lemma S1, we can show that $n^{1/2}\hat{\gamma}_k^{\Psi}=n^{1/2}\gamma_k^{\Psi}+0{\cdot}5\kappa_\Psi d_k^{\Psi}n^{1/2}(\hat{\theta}_n-\theta_0)+o_{\mathrm{p}}(1)$ for $k\geqslant 1$. Similarly, we can verify that $\hat{\gamma}_0^{\Psi}=\gamma_0^{\Psi}+o_{\mathrm{p}}(1)=\sigma_\Psi^2+o_{\mathrm{p}}(1)$. By Slutsky's lemma, the martingale central limit theorem and the Cramér–Wold device, we complete the proof of this theorem. □

Proof of Corollary 1. For any $M=d_{\min}+1,\ldots,d_{\max}$, by Theorem 1,

$$\begin{aligned}
\mathrm{pr}(\tilde{M}=M) &\leqslant \mathrm{pr}\{Q(M)-M\log n>Q(d_{\min})-d_{\min}\log n\}\\
&\leqslant \mathrm{pr}\{Q(M)>(M-d_{\min})\log n\}\\
&\leqslant \mathrm{pr}\{Q(M)>\log n\}\to 0
\end{aligned}$$

as $n\to\infty$. Hence $\mathrm{pr}(\tilde{M}=d_{\min})\to 1$ as $n\to\infty$. This, together with Theorem 1, implies (i). Similarly, (ii) can be proved by applying Theorem 2. □

S4·2. *Two auxiliary lemmas*

For any $u\in\mathbb{R}^{p+q+1}$, let

$$Z_t(u)=h_t^{1/2}(\theta_0+n^{-1/2}u)/h_t^{1/2},\quad \tilde{Z}_t(u)=\tilde{h}_t^{1/2}(\theta_0+n^{-1/2}u)/h_t^{1/2}. \tag{S13}$$

Note that $h_t = h_t(\theta_0)$. For simplicity, without causing confusion we can write, for any $u \in \mathbb{R}^{p+q+1}$,

$$h_t(u) = h_t(\theta_0 + n^{-1/2}u), \quad \tilde{h}_t(u) = \tilde{h}_t(\theta_0 + n^{-1/2}u),$$
$$\varepsilon_t(u) = \varepsilon_t(\theta_0 + n^{-1/2}u), \quad \tilde{\varepsilon}_t(u) = \tilde{\varepsilon}_t(\theta_0 + n^{-1/2}u).$$

LEMMA S1. *Suppose that Assumption* 1 *holds. Then there exists a constant* $\iota_0 > 0$ *such that*

$$E(h_t^{\iota_0}) < \infty, \quad E(|y_t|^{2\iota_0}) < \infty, \tag{S14}$$

and for some random variable ζ_0 *independent of* t *with* $E(|\zeta_0|^{\iota_0}) < \infty$, *we have that*

$$\sup_{\theta\in\Theta}\left|h_t(\theta) - \tilde{h}_t(\theta)\right| \leqslant C\rho^t\zeta_0. \tag{S15}$$

Moreover, for any $m > 0$,

$$E\left\{\sup_{\theta\in\Theta}\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\right\|^m\right\} < \infty, \quad E\left\{\sup_{\theta\in\Theta}\left\|\frac{1}{h_t(\theta)}\frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}}\right\|^m\right\} < \infty, \tag{S16}$$

and there exists a constant $c > 0$ *such that*

$$E\left(\left[\sup\left\{\frac{h_t(\theta_2)}{h_t(\theta_1)} : \|\theta_1 - \theta_2\| \leqslant c,\ \theta_1, \theta_2 \in \Theta\right\}\right]^m\right) < \infty. \tag{S17}$$

Proof of Lemma S1. The statements in (S14) are established in Lemma 2.3 of Berkes et al. (2003), and (S15) and (S16) are respectively intermediate results in the proofs of Theorems 2.1 and 2.2 in Francq & Zakoïan (2004). Assertion (S17) can be proved along the same lines as (S47) in Lemma S5(b), and the detailed proof is provided in Lemma A.1 of Zheng et al. (2016). □

LEMMA S2. *Suppose* $L = \sup_{0<x<\infty} xg(x) < \infty$ *and that* $\{w_t\}$ *is a strictly stationary and ergodic process with* $w_t \in \mathcal{F}_{t-1}$ *and* $0 \leqslant w_t \leqslant 1$ *for all* t. *If Assumptions* 1 *and* 3(i) *hold, then for any* $A > 0$,

$$\sup_{\|u\|\leqslant A}\sup_{0\leqslant x<\infty}\left|n^{-1/2}\sum_{t=1}^{n} w_t\big[I\{|\tilde{\varepsilon}_t(u)| \leqslant x\} - I(|\varepsilon_t| \leqslant x) - G\{x\tilde{Z}_t(u)\} + G(x)\big]\right| = o_{\mathrm{p}}(1).$$

Proof of Lemma S2. For $x \in [0, \infty)$ and $u \in \mathbb{R}^{p+q+1}$, let

$$S_n(x,u) = \sum_{t=1}^{n} w_t\xi_t(x,u), \quad \xi_t(x,u) = \xi_{1t}(x,u) + \xi_{2t}(x,u),$$

where

$$\xi_{1t}(x,u) = \big[I\{|\varepsilon_t| \leqslant x\tilde{Z}_t(u)\} - G\{x\tilde{Z}_t(u)\}\big] - \big[I\{|\varepsilon_t| \leqslant xZ_t(u)\} - G\{xZ_t(u)\}\big],$$
$$\xi_{2t}(x,u) = \big[I\{|\varepsilon_t| \leqslant xZ_t(u)\} - G\{xZ_t(u)\}\big] - \{I(|\varepsilon_t| \leqslant x) - G(x)\}.$$

Note that $I\{|\varepsilon_t| \leqslant x\tilde{Z}_t(u)\} = I\{|\tilde{\varepsilon}_t(u)| \leqslant x\}$ and $I\{|\varepsilon_t| \leqslant xZ_t(u)\} = I\{|\varepsilon_t(u)| \leqslant x\}$.

We prove the lemma in the following three steps:

(i) For any $A > 0$, there is a constant C depending on A such that for any $0 < x < \infty$ and u satisfying $\|u\| \leqslant A$, $\mathrm{pr}\{|S_n(x,u)| \geqslant sn^{1/2}\} \leqslant C/(s^4 n)$ for all $s > 0$.

(ii) For any $\|u\| \leqslant A$ with $A > 0$, $\sup_{0\leqslant x<\infty}|S_n(x,u)| = o_{\mathrm{p}}(n^{1/2})$.

(iii) For any $A > 0$, $\sup_{\|u\|\leqslant A}\sup_{0\leqslant x<\infty}|S_n(x,u)| = o_{\mathrm{p}}(n^{1/2})$.

First we verify (i). Observe that for any $x > 0$ and $u \in \mathbb{R}^{p+q+1}$, $\{S_k(x,u), \mathcal{F}_k,\ k = 1, \ldots, n\}$ is a martingale. Then, applying Theorem 2.11 in Hall & Heyde (1980), we have

$$E\{S_n^4(x,u)\} \leqslant C\left[\left\|\sum_{t=1}^{n} E\{w_t^2\xi_t^2(x,u) \mid \mathcal{F}_{t-1}\}\right\|_2^2 + 1\right]$$
$$\leqslant C\left[\left\|\sum_{t=1}^{n} E\{\xi_t^2(x,u) \mid \mathcal{F}_{t-1}\}\right\|_2^2 + 1\right],$$

where the last inequality is due to the fact that $0 \leqslant w_t \leqslant 1$ with probability 1. Moreover,

$$E\{\xi_t^2(x,u) \mid \mathcal{F}_{t-1}\} \leqslant 2E\{\xi_{1t}^2(x,u) \mid \mathcal{F}_{t-1}\} + 2E\{\xi_{2t}^2(x,u) \mid \mathcal{F}_{t-1}\}$$
$$\leqslant 2\big|G\{x\tilde{Z}_t(u)\} - G\{xZ_t(u)\}\big| + 2\big|G\{xZ_t(u)\} - G(x)\big|.$$

As a result,

$$E\{S_n^4(x,u)\} \leqslant C\left[\left\|\sum_{t=1}^{n}\big|G\{x\tilde{Z}_t(u)\} - G\{xZ_t(u)\}\big|\right\|_2^2 + \left\|\sum_{t=1}^{n}\big|G\{xZ_t(u)\} - G(x)\big|\right\|_2^2 + 1\right]. \tag{S18}$$

By Taylor expansion and (S15), we have

$$\sup_{\|u\|\leqslant A}\sup_{0\leqslant x<\infty}\Big|G\{x\tilde{Z}_t(u)\} - G\{xZ_t(u)\}\Big|$$
$$= \sup_{\|u\|\leqslant A}\sup_{0\leqslant x<\infty}\frac{0{\cdot}5x}{h_t^{1/2}h_t^{*1/2}}\, g\left(\frac{xh_t^{*1/2}}{h_t^{1/2}}\right)\Big|\tilde{h}_t(u) - h_t(u)\Big|$$
$$\leqslant \frac{0{\cdot}5L}{\underline{\omega}}\, C\rho^t\zeta_0, \tag{S19}$$

where h_t^* is between $\tilde{h}_t(u)$ and $h_t(u)$, and $\underline{\omega} = \inf_{\theta\in\Theta}\omega > 0$. Then

$$\Big\|G\{x\tilde{Z}_t(u)\} - G\{xZ_t(u)\}\Big\|_2^2 = E\left[\Big|G\{x\tilde{Z}_t(u)\} - G\{xZ_t(u)\}\Big|^2 I(C\rho^t\zeta \leqslant \rho^{t/2})\right]$$
$$+ E\left[\Big|G\{x\tilde{Z}_t(u)\} - G\{xZ_t(u)\}\Big|^2 I(C\rho^t\zeta > \rho^{t/2})\right]$$
$$\leqslant \frac{L^2}{\underline{\omega}}\rho^t + \mathrm{pr}(C\rho^t\zeta > \rho^{t/2}) \leqslant C(\rho^t + \rho^{\iota_0 t/2}),$$

which, together with Minkowski's inequality, implies that

$$\left\|\sum_{t=1}^{n}\Big|G\{x\tilde{Z}_t(u)\} - G\{xZ_t(u)\}\Big|\right\|_2 \leqslant \sum_{t=1}^{n}\Big\|G\{x\tilde{Z}_t(u)\} - G\{xZ_t(u)\}\Big\|_2 \leqslant C. \tag{S20}$$

By Taylor expansion again, we obtain

$$\sup_{\|u\|\leqslant A}\sup_{0\leqslant x<\infty}\big|G\{xZ_t(u)\} - G(x)\big| = \frac{0{\cdot}5}{n^{1/2}}\sup_{0\leqslant x<\infty}\left|\frac{x}{h_t^{1/2}}\, g\left\{\frac{xh_t^{1/2}(\theta^*)}{h_t^{1/2}}\right\}\frac{u^{\mathrm{T}}}{h_t^{1/2}(\theta^*)}\frac{\partial h_t(\theta^*)}{\partial\theta}\right|$$
$$\leqslant \frac{0{\cdot}5AL}{n^{1/2}}\sup_{\theta\in\Theta}\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\right\|, \tag{S21}$$

where θ^* is between θ_0 and $\theta_0 + n^{-1/2}u$. This, together with Minkowski's inequality and (S16), implies that

$$\left\| \sum_{t=1}^{n} \left| G\{xZ_t(u)\} - G(x) \right| \right\|_2 \leqslant \sum_{t=1}^{n} \left\| G\{xZ_t(u)\} - G(x) \right\|_2 \leqslant Cn^{1/2}. \tag{S22}$$

Upon combining (S18), (S20) and (S22), we have $E\{S_n^4(x,u)\} \leqslant Cn$, which, together with the Markov inequality, implies (i).

Next, we prove (ii). Define a partition of $[0,\infty)$ as $0 = x_1 < x_2 < \cdots < x_N < x_{N+1} = \infty$. Specifically, for $\Delta > 0$, choose $0 < M < \sup\{x : G(x) < 1\}$ such that $\sup_{0 \leqslant x \leqslant M} xg(x) \leqslant \Delta$. Let the integer N_1 be given by $N_1 = \max\{k \geqslant 2 : (k-1)n^{-1/2}\Delta \leqslant G(M/2)\}$, and define x_j for $j = 2, \dots, N_1$ by

$$G(x_{j+1}) = G(x_j) + n^{-1/2}\Delta \quad (j = 1, \dots, N_1 - 1).$$

Then, choose x_{N_1+1} such that $M/2 < x_{N_1+1} < 3M/4$ and $G(x_{N_1+1}) \leqslant N_1 n^{-1/2}\Delta$. To define N, first choose a positive integer K such that $K \geqslant 2/\{\gamma(0{\cdot}25 - v)\}$ with $0 < v < 1/4$ and γ as defined in Assumption 1. Let $N = N_1 + N_2 + \cdots + N_{K+1}$, where $N_2 = N_3 = \cdots = N_{K+1} = \lfloor n^{3/4} \rfloor$, with $\lfloor s \rfloor$ denoting the integer part of a real number s. Then define x_j for $j = N_1 + 2, \dots, N$ by

$$x_{N_1+i} = x_{N_1+1} + (i-1)n^{-1/2-v} \quad (i = 2, \dots, N_2),$$
$$x_{N_1+N_2+\cdots+N_k+i} = x_{N_1+N_2+\cdots+N_k} + in^{-3/4+k(1/4-v)} \quad (i = 1, \dots, N_{k+1};\ k = 2, \dots, K).$$

As a result,

$$N \leqslant Cn^{3/4}, \quad \max_{N_1 < j \leqslant N-1} (x_{j+1} - x_j)/x_j \leqslant Cn^{-1/2-v}, \tag{S23}$$

$$1 - G(x_N/2) \leqslant Cn^{-2}, \quad \max_{N_1 \leqslant j \leqslant N-1} \{G(x_{j+1}) - G(x_j)\} \leqslant Cn^{-1/2-v}, \tag{S24}$$

and

$$G(x_{j+1}) - G(x_j) = n^{-1/2}\Delta \quad (j = 1, \dots, N_1); \tag{S25}$$

see also the proof of Lemma 6.2 in Berkes & Horváth (2003).

We can show that

$$\begin{aligned}
&\sup_{x_j \leqslant x \leqslant x_{j+1}} \left| S_n(x,u) - S_n(x_{j+1},u) \right| \\
&\leqslant \max\Bigg(\sum_{t=1}^{n} w_t \left[I(|\varepsilon_t| \leqslant x_{j+1}) - I(|\varepsilon_t| \leqslant x_j) + G\{x_{j+1}\tilde{Z}_t(u)\} - G\{x_j\tilde{Z}_t(u)\} \right], \\
&\qquad \sum_{t=1}^{n} w_t \left[I\{|\varepsilon_t| \leqslant x_{j+1}\tilde{Z}_t(u)\} - I\{|\varepsilon_t| \leqslant x_j\tilde{Z}_t(u)\} + G(x_{j+1}) - G(x_j) \right] \Bigg) \\
&\leqslant |S_n(x_j,u)| + |S_n(x_{j+1},u)| + \left| \sum_{t=1}^{n} w_t \{ I(x_j < |\varepsilon_t| \leqslant x_{j+1}) - G(x_{j+1}) + G(x_j) \} \right| \\
&\quad + \sum_{t=1}^{n} w_t \left[G\{x_{j+1}\tilde{Z}_t(u)\} - G\{x_j\tilde{Z}_t(u)\} \right] + \sum_{t=1}^{n} w_t \left[G(x_{j+1}) - G(x_j) \right],
\end{aligned}$$

and then

$$\begin{aligned}\sup_{0\leqslant x<\infty}|S_n(x,u)| &\leqslant \max_{1\leqslant j\leqslant N+1}|S_n(x_j,u)| + \max_{1\leqslant j\leqslant N}\sup_{x_j\leqslant x\leqslant x_{j+1}}|S_n(x,u)-S_n(x_{j+1},u)| \\ &\leqslant 3A_{1n}+2A_{2n}+A_{3n}+A_{4n}+A_{5n}, \qquad \text{(S26)}\end{aligned}$$

where

$$A_{1n}=\max_{1\leqslant j\leqslant N}|S_n(x_j,u)|, \quad A_{2n}=\max_{2\leqslant j\leqslant N}\sum_{t=1}^{n} w_t\left|G\{x_j\tilde{Z}_t(u)\}-G\{x_jZ_t(u)\}\right|,$$

$$A_{3n}=\max_{1\leqslant j\leqslant N}\left|\sum_{t=1}^{n} w_t\{I(x_j<|\varepsilon_t|\leqslant x_{j+1})-G(x_{j+1})+G(x_j)\}\right|,$$

$$A_{4n}=\max_{1\leqslant j\leqslant N}\sum_{t=1}^{n} w_t\big[G\{x_{j+1}Z_t(u)\}-G\{x_jZ_t(u)\}\big],$$

$$A_{5n}=\max_{1\leqslant j\leqslant N}\sum_{t=1}^{n} w_t\{G(x_{j+1})-G(x_j)\},$$

and $S_n(x_{N+1},u)=S_n(+\infty,u)=0$.

By the intermediate result (i), (S23) and the Markov inequality, for any $s>0$ we have

$$\mathrm{pr}(A_{1n}\geqslant sn^{1/2})\leqslant \sum_{j=1}^{N}\mathrm{pr}\Big\{|S_n(x_j,u)|\geqslant sn^{1/2}\Big\}\leqslant \frac{Cn^{3/4}}{s^4 n},$$

which implies that

$$A_{1n}=o_{\mathrm{p}}(n^{1/2}). \qquad \text{(S27)}$$

From (S19), we have that

$$A_{2n}\leqslant C\zeta_0=O_{\mathrm{p}}(1). \qquad \text{(S28)}$$

By Theorem 2.11 in Hall & Heyde (1980) and (S24), we can show that

$$E\left[\left|\sum_{t=1}^{n} w_t\{I(x_j<|\varepsilon_t|\leqslant x_{j+1})-G(x_{j+1})+G(x_j)\}\right|^4\right]\leqslant Cn,$$

which, by using a method similar to the proof of (S27), yields

$$A_{3n}=o_{\mathrm{p}}(n^{1/2}). \qquad \text{(S29)}$$

For A_{4n} we can show that, by (S25) and a method similar to that for (S21),

$$\begin{aligned}&\max_{1\leqslant j\leqslant N_1}\sum_{t=1}^{n} w_t\big[G\{x_{j+1}Z_t(u)\}-G\{x_jZ_t(u)\}\big] \\ &\quad\leqslant n\max_{1\leqslant j\leqslant N_1}\{G(x_{j+1})-G(x_j)\}+2\max_{1\leqslant j\leqslant N_1}\sum_{t=1}^{n}\big|G\{x_{j+1}Z_t(u)\}-G(x_{j+1})\big| \\ &\quad\leqslant Cn^{1/2}\Delta\left\{1+\frac{A}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\right\|\right\}.\end{aligned}$$

By Taylor expansion and (S23),

$$
\begin{aligned}
&\max_{N_1<j\leqslant N-1}\sum_{t=1}^{n} w_t\big[G\{x_{j+1}Z_t(u)\}-G\{x_jZ_t(u)\}\big]\\
&\leqslant \max_{N_1<j\leqslant N-1}\sum_{t=1}^{n} g\{x_j^{*}Z_t(u)\}Z_t(u)(x_{j+1}-x_j)\\
&\leqslant nL\max_{N_1<j\leqslant N-1}(x_{j+1}-x_j)/x_j\leqslant Cn^{1/2-v},
\end{aligned}
$$

where x_j^* is between x_j and x_{j+1}; and by (S24),

$$
\begin{aligned}
&\sum_{t=1}^{n} w_t\big[1-G\{x_NZ_t(u)\}\big]\\
&\quad\leqslant \sum_{t=1}^{n}\big[1-G\{x_NZ_t(u)\}\big]I\{Z_t(u)<0{\cdot}5\}+\sum_{t=1}^{n}\big[1-G\{x_NZ_t(u)\}\big]I\{Z_t(u)\geqslant 0{\cdot}5\}\\
&\quad\leqslant \sum_{t=1}^{n} I\{Z_t(u)<0{\cdot}5\}+Cn^{-1}=O_{\mathrm{p}}(1)
\end{aligned}
$$

since

$$
\begin{aligned}
\mathrm{pr}\{Z_t(u)<0{\cdot}5\}&=\mathrm{pr}\left\{-n^{-1/2}\frac{1}{h_t^{1/2}h_t^{1/2}(\theta^*)}\frac{\partial h_t(\theta^*)}{\partial\theta^{\mathrm{T}}}u>1\right\}\\
&\leqslant \frac{A^2}{n}E\left\{\sup_{\|\theta-\theta_0\|\leqslant c}\frac{h_t(\theta)}{h_t}\sup_{\theta\in\Theta}\left\|\frac{1}{h_t(\theta)}\frac{\partial h_t(\theta)}{\partial\theta}\right\|^2\right\},
\end{aligned}
$$

where θ^* is between θ_0 and $\theta_0+n^{-1/2}u$. As a result,

$$A_{4n}=\Delta O_{\mathrm{p}}(n^{1/2}). \tag{S30}$$

Using (S24) and (S25), one can verify that

$$A_{5n}\leqslant C\Delta n^{1/2}. \tag{S31}$$

Note that Δ can be chosen arbitrarily small. Thus, we accomplish the proof of (ii) by combining (S26)–(S31).

Finally, we prove (iii). For any $\|u\|\leqslant A$, define a $(p+q+1)$-dimensional cube $V_\delta(u)$ by $\{u^*: u-0{\cdot}5\delta\iota\leqslant u^*\leqslant u+0{\cdot}5\delta\iota \text{ and } \|u^*\|\leqslant A\}$, where $\delta>0$, ι is a vector with all elements equal to 1 and the inequality is elementwise. Write $u_U=u+0{\cdot}5\delta\iota$ and $u_L=u-0{\cdot}5\delta\iota$. Note that for $\theta_1\leqslant\theta_2$, we have $h_t(\theta_1)\leqslant h_t(\theta_2)$ and $\tilde h_t(\theta_1)\leqslant\tilde h_t(\theta_2)$. We can then verify that

$$
\begin{aligned}
&\sup_{u^*\in V_\delta(u)}\big|S_n(x,u^*)-S_n(x,u_L)\big|\\
&\qquad\leqslant |S_n(x,u_U)|+|S_n(x,u_L)|+\sum_{t=1}^{n} w_t\left[G\{x\tilde Z_t(u_U)\}-G\{x\tilde Z_t(u_L)\}\right]
\end{aligned}
\tag{S32}
$$

and

$$\sum_{t=1}^{n} w_t \left[G\{x\tilde{Z}_t(u_U)\} - G\{x\tilde{Z}_t(u_L)\}\right]$$
$$\leqslant \sum_{t=1}^{n} w_t \left|G\{x\tilde{Z}_t(u_U)\} - G\{xZ_t(u_U)\}\right| + \sum_{t=1}^{n} w_t \left|G\{x\tilde{Z}_t(u_L)\} - G\{xZ_t(u_L)\}\right|$$
$$+ \sum_{t=1}^{n} w_t \big[G\{xZ_t(u_U)\} - G\{xZ_t(u_L)\}\big]$$

By a method similar to that used for (S21), we obtain

$$n^{-1/2} \sup_{0\leqslant x<\infty} \sum_{t=1}^{n} w_t \big[G\{xZ_t(u_U)\} - G\{xZ_t(u_L)\}\big] \leqslant \delta \frac{0{\cdot}5L\|\iota\|}{n} \sum_{t=1}^{n} \sup_{\theta\in\Theta} \left\| \frac{1}{h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} \right\|, \tag{S33}$$

and it is a direct consequence of (S19) that

$$n^{-1/2} \sup_{\|u\|\leqslant A} \sup_{0\leqslant x<\infty} \sum_{t=1}^{n} w_t \left|G\{x\tilde{Z}_t(u)\} - G\{xZ_t(u)\}\right| \leqslant Cn^{-1/2}\zeta_0. \tag{S34}$$

By (S32)–(S34), the intermediate result (ii) and the finite covering theorem, we complete the proof of (iii), and thus the lemma follows. □

S4·3. *Proof of Lemma* A1

We first show that for any $A > 0$,

$$\sup_{0\leqslant x<\infty} \sup_{\|u\|\leqslant A} \left| n^{-1/2} \sum_{t=1}^{n} w_t \big[G\{xZ_t(u)\} - G(x)\big] - 0{\cdot}5xg(x)d_w^{\mathrm{T}}u \right| = o_{\mathrm{p}}(1). \tag{S35}$$

By Assumption 3, for any $\Delta > 0$ we can choose $0 < C_1 < C_2 < \infty$ such that $\sup_{0<x\leqslant 2C_1} xg(x) \leqslant \Delta$ and $\sup_{C_2/2\leqslant x<\infty} xg(x) \leqslant \Delta$. By Taylor expansion, we have

$$\sup_{\|u\|\leqslant A} |Z_t(u) - 1| = \sup_{\|u\|\leqslant A} \frac{0{\cdot}5}{n^{1/2}} \left| \frac{u^{\mathrm{T}}}{h_t^{1/2} h_t^{1/2}(\theta^*)} \frac{\partial h_t(\theta^*)}{\partial \theta} \right|$$
$$\leqslant \frac{0{\cdot}5A}{n^{1/2}} \sup_{\|\theta-\theta_0\|\leqslant c} \frac{h_t^{1/2}(\theta)}{h_t^{1/2}} \sup_{\theta\in\Theta} \left\| \frac{1}{h_t(\theta)} \frac{\partial h_t(\theta)}{\partial \theta} \right\|,$$

which, together with (S16), (S17) and the Markov inequality, implies

$$\mathrm{pr}\left\{ \max_{1\leqslant t\leqslant n} \sup_{\|u\|\leqslant A} |Z_t(u) - 1| > n^{-1/8} \right\} \leqslant Cn^{-2},$$

where θ^* is between θ_0 and $\theta_0 + n^{-1/2}u$. Hence, by the Borel–Cantelli lemma, we have

$$\max_{1\leqslant t\leqslant n} \sup_{\|u\|\leqslant A} |Z_t(u) - 1| \leqslant Cn^{-1/8} \tag{S36}$$

with probability 1, which implies

$$\sup_{C_1\leqslant x\leqslant C_2} \max_{1\leqslant t\leqslant n} \sup_{\|u\|\leqslant A} \left| xZ_t(u)g\{xZ_t(u)\} - xg(x) \right| = o_{\mathrm{p}}(1),$$

since the function $xg(x)$ is uniformly continuous on $[C_1/2, 2C_2]$ by Assumption 3(iii). Moreover, using (S36) we can show that

$$\sup_{x\in[0,C_1]\cup[C_2,\infty)} \max_{1\leqslant t\leqslant n} \sup_{\|u\|\leqslant A} \left| xZ_t(u)g\{xZ_t(u)\} - xg(x) \right| \leqslant 2\Delta,$$

and it then follows that

$$\sup_{0\leqslant x<\infty} \max_{1\leqslant t\leqslant n} \sup_{\|u\|\leqslant A} \left| xZ_t(u)g\{xZ_t(u)\} - xg(x) \right| = o_{\rm p}(1). \tag{S37}$$

On the other hand, by Taylor expansion it can be shown that

$$\begin{aligned}
&\sup_{\|u\|\leqslant A} \left\| \frac{1}{n}\sum_{t=1}^{n} \frac{w_t}{h_t(u)} \frac{\partial h_t(u)}{\partial\theta} - \frac{1}{n}\sum_{t=1}^{n} \frac{w_t}{h_t} \frac{\partial h_t(\theta_0)}{\partial\theta} \right\| \\
&\leqslant n^{-1/2} A \left\{ \frac{1}{n}\sum_{t=1}^{n} \sup_{\theta\in\Theta} \left\| \frac{w_t}{h_t(\theta)} \frac{\partial^2 h_t(\theta)}{\partial\theta\partial\theta^{\rm T}} \right\| + \frac{1}{n}\sup_{\theta\in\Theta} \left\| \sum_{t=1}^{n} \frac{w_t}{h_t^2(\theta)} \frac{\partial h_t(\theta)}{\partial\theta} \frac{\partial h_t(\theta)}{\partial\theta^{\rm T}} \right\| \right\},
\end{aligned}$$

which, together with (S16) and the ergodic theorem, implies that

$$\sup_{\|u\|\leqslant A} \left\| \frac{1}{n}\sum_{t=1}^{n} \frac{w_t}{h_t(u)} \frac{\partial h_t(u)}{\partial\theta} - d_w \right\| = o_{\rm p}(1). \tag{S38}$$

By (S37), (S38) and the Taylor expansion in (S21), we have

$$\begin{aligned}
&\sup_{0\leqslant x<\infty} \sup_{\|u\|\leqslant A} \left| n^{-1/2} \sum_{t=1}^{n} w_t \big[G\{xZ_t(u)\} - G(x) \big] - 0{\cdot}5xg(x)d_w^{\rm T}u \right| \\
&= \sup_{0\leqslant x<\infty} \sup_{\|u\|\leqslant A} \left| \frac{0{\cdot}5}{n} \sum_{t=1}^{n} xZ_t(u^*)g\{xZ_t(u^*)\} \frac{w_t u^{\rm T}}{h_t(u^*)} \frac{\partial h_t(u^*)}{\partial\theta} - 0{\cdot}5xg(x)d_w^{\rm T}u \right| \\
&= o_{\rm p}(1),
\end{aligned}$$

where u^* is between zero and u; hence (S35) holds.

We complete the proof of this lemma by combining Lemma S2, (S34), (S35) and the fact that $n^{1/2}(\hat\theta_n - \theta_0) = O_{\rm p}(1)$.

S4·4. *Proof of Lemma* A2

We first show that for any $A > 0$,

$$\sup_{\|u\|\leqslant A} \sup_{0\leqslant x<\infty} \left| n^{-1/2} \sum_{t=1}^{n} \{w_t - E(w_t)\} \big[I\{|\tilde\varepsilon_t(u)| \leqslant x\} - I(|\varepsilon_t| \leqslant x) \big] \right| = o_{\rm p}(1); \tag{S39}$$

its proof is similar to that of Lemma S2.

For $x \in [0,\infty)$ and $u \in \mathbb{R}^{p+q+1}$, let

$$\tilde S_n(x,u) = \sum_{t=1}^{n} \{w_t - E(w_t)\}\tilde\xi_t(x,u), \quad \tilde\xi_t(x,u) = \tilde\xi_{1t}(x,u) + \tilde\xi_{2t}(x,u),$$

where

$$\begin{aligned}
\tilde\xi_{1t}(x,u) &= I\{|\varepsilon_t| \leqslant x\tilde Z_t(u)\} - I\{|\varepsilon_t| \leqslant xZ_t(u)\}, \\
\tilde\xi_{2t}(x,u) &= I\{|\varepsilon_t| \leqslant xZ_t(u)\} - I(|\varepsilon_t| \leqslant x).
\end{aligned}$$

Now we are ready to prove (S39) by following the same steps (i)–(iii) as in the proof of Lemma S2 for $\tilde{S}_n(x,u)$.

We begin with (i). Observe that $\{\tilde{S}_k(x,u), \mathcal{F}_k,\ k=1,\ldots,n\}$ is a martingale for any $0<x<\infty$ and $u\in\mathbb{R}^{p+q+1}$. Similarly to (S19), by Theorem 2.11 in Hall & Heyde (1980) we have

$$\begin{aligned} E\big\{\tilde{S}_n^4(x,u)\big\} &\leqslant C\left(\left\|\sum_{t=1}^n E\big[\{w_t-E(w_t)\}^2\tilde{\xi}_t^2(x,u)\mid\mathcal{F}_{t-1}\big]\right\|_2^2+1\right)\\ &\leqslant C\left[\left\|\sum_{t=1}^n E\big\{\tilde{\xi}_t^2(x,u)\mid\mathcal{F}_{t-1}\big\}\right\|_2^2+1\right], \end{aligned}$$

with

$$\begin{aligned} E\big\{\tilde{\xi}_t^2(x,u)\mid\mathcal{F}_{t-1}\big\} &\leqslant 2E\big\{\tilde{\xi}_{1t}^2(x,u)\mid\mathcal{F}_{t-1}\big\}+2E\big\{\tilde{\xi}_{2t}^2(x,u)\mid\mathcal{F}_{t-1}\big\}\\ &=2\big|G\{x\tilde{Z}_t(u)\}-G\{xZ_t(u)\}\big|+2\big|G\{xZ_t(u)\}-G(x)\big|. \end{aligned}$$

Then the method for establishing (i) in the proof of Lemma S2 can be applied to show that (i) also holds for $\tilde{S}_n(x,u)$.

Next, to show (ii), we employ the partition of $[0,\infty)$ defined in (ii) in the proof of Lemma S2. Let $w_t^+=\max\{0,w_t-E(w_t)\}$ and $w_t^-=-\min\{0,w_t-E(w_t)\}$. We can show that

$$\begin{aligned} &\sup_{x_j\leqslant x\leqslant x_{j+1}}\left|\tilde{S}_n(x,u)-\tilde{S}_n(x_{j+1},u)\right|\\ &\leqslant\max\left(\sum_{t=1}^n\left[w_t^+I(x_j<|\varepsilon_t|\leqslant x_{j+1})+w_t^-I\Big\{x_j\tilde{Z}_t(u)<|\varepsilon_t|\leqslant x_{j+1}\tilde{Z}_t(u)\Big\}\right],\right.\\ &\qquad\left.\sum_{t=1}^n\left[w_t^-I(x_j<|\varepsilon_t|\leqslant x_{j+1})+w_t^+I\Big\{x_j\tilde{Z}_t(u)<|\varepsilon_t|\leqslant x_{j+1}\tilde{Z}_t(u)\Big\}\right]\right), \end{aligned}$$

and, since w_t^+ and w_t^- are both bounded with probability 1, we further have that

$$\begin{aligned} &\sup_{x_j\leqslant x\leqslant x_{j+1}}\left|\tilde{S}_n(x,u)-\tilde{S}_n(x_{j+1},u)\right|\\ &\leqslant C\left[\sum_{t=1}^n I(x_j<|\varepsilon_t|\leqslant x_{j+1})+\sum_{t=1}^n I\Big\{x_j\tilde{Z}_t(u)<|\varepsilon_t|\leqslant x_{j+1}\tilde{Z}_t(u)\Big\}\right]\\ &\leqslant C\left|\sum_{t=1}^n\big\{I(x_j<|\varepsilon_t|\leqslant x_{j+1})-G(x_{j+1})+G(x_j)\big\}\right|\\ &\quad+C\left|\sum_{t=1}^n\Big[I\big\{x_j\tilde{Z}_t(u)<|\varepsilon_t|\leqslant x_{j+1}\tilde{Z}_t(u)\big\}+G(x_{j+1})-G(x_j)\Big]\right|. \end{aligned}$$

Therefore, (ii) can be established following the lines of (ii) in the proof of Lemma S2.

Finally, to prove (iii), we consider again the $(p+q+1)$-dimensional cube $V_\delta(u)$. It can be shown that

$$\sup_{u^*\in V_\delta(u)}\left|\tilde{S}_n(x,u^*)-\tilde{S}_n(x,u_L)\right|\leqslant C\sum_{t=1}^n\Big[I\big\{|\varepsilon_t|\leqslant x\tilde{Z}_t(u_U)\big\}-I\big\{|\varepsilon_t|\leqslant x\tilde{Z}_t(u_L)\big\}\Big].$$

Thus, (iii) can be established in a similar way to its verification in the proof of Lemma S2. Hence (S39) holds.

Applying Lemma A2 with $w_t \equiv 1$, we have

$$\sup_{0\leqslant x<\infty}\left|n^{-1/2}\sum_{t=1}^{n}\big\{I(|\hat{\varepsilon}_t|\leqslant x)-I(|\varepsilon_t|\leqslant x)\big\}-xg(x)d_0^{*\mathrm{T}}n^{1/2}(\hat{\theta}_n-\theta_0)\right|=o_{\mathrm{p}}(1), \quad \text{(S40)}$$

which, together with (S39), completes the proof of this lemma.

S4·5. *Proof of Lemma* A3

The proof of this lemma is based on Hallin et al. (1985). For the sample $X_1,\dots,X_n$, let $X_{(\cdot)}=(X_{(1)},\dots,X_{(n)})$ be the order statistic and R_t the rank of the observation X_t. Given $X_{(\cdot)}$, define, for $i,j\in\{1,2,\dots,n\},\ \alpha(i,j)=ij/n^2-F(X_{(i)})F(X_{(j)})$. Denote $\alpha(R_t,R_{t-k})$ by α_t for simplicity. Then for $t=k+1,\dots,n$ we have

$$F_n(X_t)F_n(X_{t-k})-F(X_t)F(X_{t-k})=R_tR_{t-k}/n^2-F(X_{(R_t)})F(X_{(R_{t-k})})=\alpha_t.$$

Observe that

$$\begin{aligned}
n^{-1/2}\sum_{t=k+1}^{n}E(\alpha_t\mid X_{(\cdot)})&=\frac{n-k}{\sqrt{n}}\big[E(R_tR_{t-k}/n^2)-E\big\{F(X_{(R_t)})F(X_{(R_{t-k})})\mid X_{(\cdot)}\big\}\big]\\
&=\frac{n-k}{\sqrt{n}}\left\{\frac{(n+1)(3n+2)}{12n^2}-\binom{n}{2}^{-1}\sum_{1\leqslant t<s\leqslant n}F(X_t)F(X_s)\right\}\\
&=-n^{1/2}\binom{n}{2}^{-1}\sum_{1\leqslant t<s\leqslant n}\{F(X_t)F(X_s)-0{\cdot}25\}+o(1).
\end{aligned}$$

Moreover, by the projection results on U-statistics due to Hoeffding (1948), we have

$$n^{1/2}\binom{n}{2}^{-1}\sum_{1\leqslant t<s\leqslant n}\{F(X_t)F(X_s)-0{\cdot}25\}=n^{-1/2}\sum_{t=1}^{n}\{F(X_t)-0{\cdot}5\}+o_{\mathrm{p}}(1);$$

see also Theorem 12.3 in van der Vaart (1998), for instance. Thus, it follows that

$$n^{-1/2}\sum_{t=k+1}^{n}E(\alpha_t\mid X_{(\cdot)})=-n^{-1/2}\sum_{t=k+1}^{n}\{F(X_t)-0{\cdot}5\}+o_{\mathrm{p}}(1).$$

Let $\Delta_n=n^{-1/2}\sum_{t=k+1}^{n}\alpha_t-n^{-1/2}\sum_{t=k+1}^{n}E(\alpha_t\mid X_{(\cdot)})$. It then suffices to show that $\Delta_n=o_{\mathrm{p}}(1)$. In the following proof, let (r,s) be a pair of integers satisfying $k+1\leqslant r\neq s\leqslant n$ and

$|r-s| \neq k$. We have that

$$\begin{aligned}
&\operatorname{var}\left(\sum_{t=k+1}^{n} \alpha_t \,\middle|\, X_{(\cdot)}\right) \\
&= E\left(\sum_{t=k+1}^{n} \alpha_t^2 + 2\sum_{t=k+1}^{n-k} \alpha_t\alpha_{t+k} + \sum_{\substack{k+1\leqslant r\neq s\leqslant n\\ |r-s|\neq k}} \alpha_r\alpha_s \,\middle|\, X_{(\cdot)}\right) - (n-k)^2\{E(\alpha_t \mid X_{(\cdot)})\}^2 \\
&= (n-k)E(\alpha_t^2 \mid X_{(\cdot)}) + (2n-4k)E(\alpha_t\alpha_{t+k} \mid X_{(\cdot)}) \\
&\quad + \left\{(n-k)^2 - 3n + 5k\right\} E(\alpha_r\alpha_s \mid X_{(\cdot)}) - (n-k)^2\{E(\alpha_t \mid X_{(\cdot)})\}^2 \\
&\leqslant CnE(\alpha_t^2 \mid X_{(\cdot)}) + n^2\left|E(\alpha_r\alpha_s \mid X_{(\cdot)}) - \{E(\alpha_t \mid X_{(\cdot)})\}^2\right|, \qquad \text{(S41)}
\end{aligned}$$

where the last inequality is due to the fact that, for any $t_1, t_2, t_3, t_4 = k+1, \ldots, n$,

$$\left|E\{\alpha(R_{t_1}, R_{t_2})\alpha(R_{t_3}, R_{t_4}) \mid X_{(\cdot)}\}\right| \leqslant E(\alpha_t^2 \mid X_{(\cdot)}). \qquad \text{(S42)}$$

Observe that

$$\begin{aligned}
\{E(\alpha_t \mid X_{(\cdot)})\}^2 = \left[E\{\alpha(R_t, R_{t-k}) \mid X_{(\cdot)}\}\right]^2 &= \frac{1}{n^2(n-1)^2} \sum_{1\leqslant i\neq j\leqslant n}\ \sum_{1\leqslant k\neq l\leqslant n} \alpha(i,j)\alpha(k,l) \\
&= A_1 + A_2 + A_3,
\end{aligned}$$

where

$$\begin{aligned}
A_1 &= \frac{1}{n^2(n-1)^2} \sum_{1\leqslant i\neq j\leqslant n}\ \sum_{\substack{1\leqslant k\neq l\leqslant n\\ \{k,l\}\cap\{i,j\}=\emptyset}} \alpha(i,j)\alpha(k,l) = \frac{(n-2)(n-3)}{n(n-1)} E(\alpha_r\alpha_s \mid X_{(\cdot)}), \\
A_2 &= \frac{1}{n^2(n-1)^2} \sum_{1\leqslant i\neq j\leqslant n}\ \sum_{\substack{1\leqslant k\neq l\leqslant n\\ \#\{k,l\}\cap\{i,j\}=1}} \alpha(i,j)\alpha(k,l) \\
&= \frac{n-2}{n(n-1)}\big[E\{\alpha(R_r, R_{r-k})\alpha(R_r, R_{s-k}) \mid X_{(\cdot)}\} + E\{\alpha(R_r, R_{r-k})\alpha(R_{r-k}, R_{s-k}) \mid X_{(\cdot)}\} \\
&\qquad + E\{\alpha(R_r, R_{r-k})\alpha(R_s, R_r) \mid X_{(\cdot)}\} + E\{\alpha(R_r, R_{r-k})\alpha(R_s, R_{r-k}) \mid X_{(\cdot)}\}\big]
\end{aligned}$$

and

$$\begin{aligned}
A_3 &= \frac{1}{n^2(n-1)^2} \sum_{1\leqslant i\neq j\leqslant n}\ \sum_{\substack{1\leqslant k\neq l\leqslant n\\ \#\{k,l\}\cap\{i,j\}=2}} \alpha(i,j)\alpha(k,l) \\
&= \frac{1}{n(n-1)}\big[E\{\alpha^2(R_r, R_{r-k}) \mid X_{(\cdot)}\} + E\{\alpha(R_r, R_{r-k})\alpha(R_{r-k}, R_r) \mid X_{(\cdot)}\}\big].
\end{aligned}$$

Then, applying (S42) again, we have that

$$\left|E(\alpha_r\alpha_s \mid X_{(\cdot)}) - \frac{n(n-1)}{(n-2)(n-3)}\{E(\alpha_t \mid X_{(\cdot)})\}^2\right| \leqslant \frac{C}{n} E(\alpha_t^2 \mid X_{(\cdot)}),$$

and so

$$\begin{aligned}
&\left|E(\alpha_r\alpha_s \mid X_{(\cdot)}) - \{E(\alpha_t \mid X_{(\cdot)})\}^2\right| \\
&\quad\leqslant \left|E(\alpha_r\alpha_s \mid X_{(\cdot)}) - \frac{n(n-1)}{(n-2)(n-3)}\{E(\alpha_t \mid X_{(\cdot)})\}^2\right| + \frac{4n-6}{(n-2)(n-3)}\{E(\alpha_t \mid X_{(\cdot)})\}^2 \\
&\quad\leqslant \frac{C}{n}E(\alpha_t^2 \mid X_{(\cdot)}).
\end{aligned}$$

This, together with (S41), yields

$$E(\Delta_n^2) = \frac{1}{n}E\left\{\operatorname{var}\left(\sum_{t=k+1}^{n}\alpha_t \,\middle|\, X_{(\cdot)}\right)\right\} \leqslant CE(\alpha_t^2).$$

Note that

$$\begin{aligned}
|\alpha_t| &= \left|\{F_n(X_t) - F(X_t)\}F_n(X_{t-k}) + F(X_t)\{F_n(X_{t-k}) - F(X_{t-k})\}\right| \\
&\leqslant |F_n(X_t) - F(X_t)| + |F_n(X_{t-k}) - F(X_{t-k})|
\end{aligned}$$

and that for any $s > 0$,

$$E\Big\{n^{1/2}\sup_x |F_n(x) - F(x)|\Big\}^s = O(1),$$

which is a direct consequence of the Dvoretzky–Kiefer–Wolfowitz inequality. As a result, using Minkowski's inequality, we obtain that

$$\{E(\alpha_t^2)\}^{1/2} \leqslant 2\big[E\{F_n(X_t) - F(X_t)\}^2\big]^{1/2} = o(1).$$

Therefore $E(\Delta_n^2) = o(1)$, implying $\Delta_n = o_{\mathrm{p}}(1)$. The proof is thus complete.

S5. Proofs of Proposition 2 and Theorems 3 and 4

S5·1. *Proofs of Proposition* 2 *and Theorems* 3 *and* 4

To establish Theorems 3 and 4, we first state five auxiliary lemmas, Lemmas S3–S7, whose proofs are given in subsequent subsections. In particular, the proofs of Lemmas S4 and S5 are based on the method used to prove Theorem 3 in Francq & Zakoïan (2009). We also introduce an additional lemma, Lemma S8, which plays the same role in the proof of Lemma S6 as Lemma S2 does in the proof of Lemma A1.

LEMMA S3. *Suppose that Assumptions* 1, 5 *and* 6 *hold. Then for all t and all $n \geqslant n_0$, we have that $y_t^2 \leqslant y_{t,n+1}^2 \leqslant y_{t,n}^2$ and $h_t \leqslant h_{t,n+1} \leqslant h_{t,n}$, and* $\lim_{n\to\infty} y_{t,n} = y_t$ *and* $\lim_{n\to\infty} h_{t,n} = h_t$ *with probability* 1. *Moreover, there exists a constant* $0 < \iota_1 < \min(\delta_0, 1)$ *independent of n such that $E(|y_{t,n_0}|^{2\iota_1}) < \infty$ and $E(h_{t,n_0}^{\iota_1}) < \infty$.*

LEMMA S4. *Under Assumptions* 1, 5 *and* 6, *there exist processes* $\{Y_{t,n}^{(1l)}\}$, $\{Y_{t,n}^{(1u)}\}$, $\{Y_{t,n}^{(2l)}\}$ *and* $\{Y_{t,n}^{(2u)}\}$ *such that:*

(a) *the random variables $Y_{t,n}^{(1l)}$, $Y_{t,n}^{(1u)}$, $Y_{t,n}^{(2l)}$ and $Y_{t,n}^{(2u)}$ are $\mathcal{F}_{t-1}$-measurable for all t and all n;*
(b) *for all t and all $n \geqslant n_0$,*

$$Y_{t,n_0}^{(1l)} \leqslant \frac{1}{h_{t,n}(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta} \leqslant Y_{t,n_0}^{(1u)}, \quad Y_{t,n_0}^{(2l)} \leqslant \frac{1}{h_{t,n}(\theta_0)}\frac{\partial^2 h_{t,n}(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}} \leqslant Y_{t,n_0}^{(2u)},$$

and $\{Y^{(1l)}_{t,n_0}\}$, $\{Y^{(1u)}_{t,n_0}\}$, $\{Y^{(2l)}_{t,n_0}\}$ *and* $\{Y^{(2u)}_{t,n_0}\}$ *are strictly stationary and ergodic processes with*

$$E\Big(\|Y^{(1u)}_{t,n_0}\|^m\Big) < \infty, \quad E\Big(\|Y^{(2u)}_{t,n_0}\|^m\Big) < \infty$$

for any $m > 0$*;*

(c) *for each fixed* t*, the sequences* $\{Y^{(1l)}_{t,n}\}$ *and* $\{Y^{(2l)}_{t,n}\}$ *are monotone increasing, the sequences* $\{Y^{(1u)}_{t,n}\}$ *and* $\{Y^{(2u)}_{t,n}\}$ *are monotone decreasing, i.e.,* $Y^{(1l)}_{t,n} \leqslant Y^{(1l)}_{t,n+1} \leqslant Y^{(1u)}_{t,n+1} \leqslant Y^{(1u)}_{t,n}$ *and* $Y^{(2l)}_{t,n} \leqslant Y^{(2l)}_{t,n+1} \leqslant Y^{(2u)}_{t,n+1} \leqslant Y^{(2u)}_{t,n}$ *for all* n*, and*

$$\lim_{n\to\infty} Y^{(1l)}_{t,n} = \lim_{n\to\infty} Y^{(1u)}_{t,n} = \frac{1}{h_t}\frac{\partial h_t(\theta_0)}{\partial\theta}, \quad \lim_{n\to\infty} Y^{(2l)}_{t,n} = \lim_{n\to\infty} Y^{(2u)}_{t,n} = \frac{1}{h_t}\frac{\partial^2 h_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}$$

with probability 1.

LEMMA S5. *Under Assumptions* 1*,* 5 *and* 6*, the following results hold.*

(a) *For any* $n \geqslant n_0$*,*

$$\sup_{\theta\in\Theta}\Big|h_{t,n}(\theta) - \tilde{h}_{t,n}(\theta)\Big| \leqslant C\rho^t\zeta_1, \quad \sup_{\theta\in\Theta}\left\|\frac{\partial h_{t,n}(\theta)}{\partial\theta} - \frac{\partial\tilde{h}_{t,n}(\theta)}{\partial\theta}\right\| \leqslant C\rho^t\zeta_1, \tag{S43}$$

$$\sup_{\theta\in\Theta}\left\|\frac{\partial^2 h_{t,n}(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}} - \frac{\partial^2\tilde{h}_{t,n}(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}}\right\| \leqslant C\rho^t\zeta_1, \tag{S44}$$

where ζ_1 *is a random variable independent of* t *and* n *which satisfies* $E(|\zeta_1|^{\iota_1}) < \infty$ *with* ι_1 *defined as in Lemma* S3*.*

(b) *For any* $m > 0$ *and all* $i, j, k \in \{1, \ldots, p+q+1\}$*,*

$$E\left\{\sup_{n\geqslant n_0}\sup_{\theta\in\Theta}\left\|\frac{1}{h_{t,n}(\theta)}\frac{\partial h_{t,n}(\theta)}{\partial\theta}\right\|^m\right\} < \infty, \quad E\left\{\sup_{n\geqslant n_0}\sup_{\theta\in\Theta}\left\|\frac{1}{h_{t,n}(\theta)}\frac{\partial^2 h_{t,n}(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}}\right\|^m\right\} < \infty, \tag{S45}$$

$$E\left\{\sup_{n\geqslant n_0}\sup_{\theta\in\Theta}\left|\frac{1}{h_{t,n}(\theta)}\frac{\partial^3 h_{t,n}(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right|^m\right\} < \infty, \tag{S46}$$

and there exists a constant $c > 0$ *independent of* n *such that*

$$E\left(\left[\sup_{n\geqslant n_0}\sup\left\{\frac{h_{t,n}(\theta_2)}{h_{t,n}(\theta_1)} : \|\theta_1-\theta_2\| \leqslant c,\ \theta_1,\theta_2\in\Theta\right\}\right]^m\right) < \infty. \tag{S47}$$

LEMMA S6. *Suppose that* H_{1n} *and Assumptions* 1 *and* 3–7 *hold with* $E\{(r^{(u)}_{t,n_0})^{4+\delta_1}\} < \infty$ *for some* $\delta_1 > 0$ *and* $n^{1/2}(\hat{\theta}_n - \theta_0) = O_{\mathrm{p}}(1)$*. If* w_t *is* $\mathcal{F}_{t-1}$*-measurable for all* t*, then*

$$\sup_{0\leqslant x<\infty}\left|n^{-1/2}\sum_{t=1}^{n} w_t\{I(|\hat{\varepsilon}_t| \leqslant x) - I(|\varepsilon_t| \leqslant x)\} - 0{\cdot}5xg(x)\{d_w^{\mathrm{T}} n^{1/2}(\hat{\theta}_n - \theta_0) - v_w\}\right| = o_{\mathrm{p}}(1),$$

where d_w *is defined as in Lemma* A1 *and* $v_w = E(w_t r_t)$*.*

LEMMA S7. *Suppose that H_{1n} and Assumptions* 1 *and* 3–7 *hold with $E\{(r_{t,n_0}^{(u)})^{4+\delta_1}\} < \infty$ for some $\delta_1 > 0$ and $n^{1/2}(\hat\theta_n - \theta_0) = O_{\mathrm{p}}(1)$. If w_t is independent of $\mathcal{F}_t$ for all t, then*

$$\sup_{0\leqslant x<\infty}\left|n^{-1/2}\sum_{t=1}^{n} w_t\big\{I(|\hat\varepsilon_t|\leqslant x)-I(|\varepsilon_t|\leqslant x)\big\} - E(w_t)xg(x)\big\{d_0^{*\mathrm{T}}n^{1/2}(\hat\theta_n-\theta_0)-v_0^*\big\}\right| = o_{\mathrm{p}}(1),$$

where d_0^* is defined as in Lemma A2 *and $v_0^* = 0{\cdot}5E(r_t)$.*

Lemmas S3 and S4 provide lower and upper bounds for certain sequences in our proofs so that the sandwich rule can be applied; see also Francq & Zakoïan (2009). Lemma S5 contains some preliminary results that will be used repeatedly. Lemmas S6 and S7 play the same roles in the proof of Theorem 3 as Lemmas A1 and A2 respectively did in the proof of Theorem 1.

Proof of Proposition 2. Notice that model (6) is a GARCH(p^*, q^*) model with parameters $\omega_n = \omega_0 + n^{-1/2}s_0$, $\alpha_{ni} = I(1 \leqslant i \leqslant p)\alpha_{0i} + n^{-1/2}s_i$ for $1 \leqslant i \leqslant p^*$, and $\beta_{nj} = I(1 \leqslant j \leqslant q)\beta_{0j} + n^{-1/2}s_{p^*+j}$ for $1 \leqslant j \leqslant q^*$, where $I(\cdot)$ is the indicator function. Let

$$B_0 = \begin{pmatrix} \beta_{01} & \cdots & \beta_{0q-1} & \beta_{0q} \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}, \quad B_n = \begin{pmatrix} \beta_{n1} & \cdots & \beta_{nq^*-1} & \beta_{nq^*} \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}. \tag{S48}$$

Note that h_t in model (1) admits the ARCH(∞) representation, $h_t = \phi_{00} + \sum_{\ell=1}^{\infty}\phi_{0\ell}y_{t-\ell}^2$ where $\phi_{00} = \omega_0/(1-\sum_{j=1}^{q}\beta_{0j})$ and $\phi_{0\ell} = \sum_{i=1}^{\min(\ell,p)} e_1^{\mathrm{T}}B_0^{\ell-i}e_1\alpha_{0i}$ for $\ell \geqslant 1$. Similarly, we have $h_{t,n} = \phi_{n0} + \sum_{\ell=1}^{\infty}\phi_{n\ell}y_{t-\ell,n}^2$, where $\phi_{n0} = \omega_n/(1-\sum_{j=1}^{q^*}\beta_{nj})$ and $\phi_{n\ell} = \sum_{i=1}^{\min(\ell,p^*)} e_1^{\mathrm{T}}B_n^{\ell-i}e_1\alpha_{ni}$ for $\ell \geqslant 1$.

For any positive integer k, let ${}_kh_t = \phi_{00} + \phi_{0k}y_{t-k,n_0}^2 + \sum_{\ell=1,\ell\neq k}^{\infty}\phi_{0\ell}y_{t-\ell}^2$ and ${}^kh_t = \phi_{00} + \phi_{0k}y_{t-k}^2 + \sum_{\ell=1,\ell\neq k}^{\infty}\phi_{0\ell}y_{t-\ell,n_0}^2$. Notice that both ${}_kh_t$ and ${}^kh_t$ depend on n_0. By a method similar to the proof of Lemma S4, we can verify that for all t and all $n \geqslant n_0$, $r_{t,n}$ is bounded below and above by, respectively,

$$r_{t,n_0}^{(l)} = \sum_{k=0}^{\infty} e_1^{\mathrm{T}}B_0^k e_1\left\{\frac{s_0+\sum_{j=1}^{q^*}s_{p^*+j}\phi_{00}}{h_{t,n_0}(\theta_0)} + \sum_{i=1}^{p^*}s_i\frac{y_{t-k-i}^2}{{}_{k+i}h_t} + \sum_{j=1}^{q^*}s_{p^*+j}\sum_{\ell=1}^{\infty}\phi_{0\ell}\frac{y_{t-k-j-\ell}^2}{{}_{k+j+\ell}h_t}\right\}$$

and

$$r_{t,n_0}^{(u)} = \sum_{k=0}^{\infty} e_1^{\mathrm{T}}B_0^k e_1\left(\frac{s_0+\sum_{j=1}^{q^*}s_{p^*+j}\phi_{n_00}}{h_t} + \sum_{i=1}^{p^*}s_i\frac{y_{t-k-i,n_0}^2}{{}^{k+i}h_t} + \sum_{j=1}^{q^*}s_{p^*+j}\sum_{\ell=1}^{\infty}\phi_{n_0\ell}\frac{y_{t-k-j-\ell,n_0}^2}{{}^{k+j+\ell}h_t}\right)$$

and that the processes $\{r_{t,n}^{(l)}\}$ and $\{r_{t,n}^{(u)}\}$ satisfy Assumption 7, where $r_{t,n}^{(l)}$ and $r_{t,n}^{(u)}$ are defined by replacing all n with n_0 on the right-hand sides of the above expressions as well as in the ${}_kh_t$ and the ${}^kh_t$ for all k. Moreover, for any $m > 0$, we can show that $E\{(r_{t,n_0}^{(u)})^m\} < \infty$. □

Proof of Theorem 3. As in the proof of Theorem 1, we first establish two intermediate results:

$$n^{-1/2}\sum_{t=k+1}^{n}\left\{\hat{G}_n(|\hat{\varepsilon}_t|)\hat{G}_n(|\hat{\varepsilon}_{t-k}|)-G_n(|\hat{\varepsilon}_t|)G_n(|\hat{\varepsilon}_{t-k}|)\right\} \\ -\kappa\left\{d_0^{*\mathrm{T}}n^{1/2}(\hat{\theta}_n-\theta_0)-v_0^*\right\}=o_{\mathrm{p}}(1), \quad \text{(S49)}$$

$$n^{-1/2}\sum_{t=k+1}^{n}\{G_n(|\hat{\varepsilon}_t|)G_n(|\hat{\varepsilon}_{t-k}|)-G_n(|\varepsilon_t|)G_n(|\varepsilon_{t-k}|)\} \\ +0{\cdot}5\kappa\left\{(d_0^*+d_k^*)^{\mathrm{T}}n^{1/2}(\hat{\theta}_n-\theta_0)-(v_0^*+v_k^*)\right\}=o_{\mathrm{p}}(1) \quad \text{(S50)}$$

for any positive integer k, where, for $k\geqslant 1$, d_k^* is defined as in the proof of Theorem 1 and $v_k^*=E\{G(|\varepsilon_{t-k}|)r_t\}$.

To prove (S49), let $\tilde{W}_t=G_n(|\hat{\varepsilon}_t|)+|\hat{\varepsilon}_t|g(|\hat{\varepsilon}_t|)\{d_0^{*\mathrm{T}}(\hat{\theta}_n-\theta_0)-n^{-1/2}v_0^*\}$. Applying Lemma S6 with $w_t\equiv 1$, we have $n^{1/2}\max_{1\leqslant t\leqslant n}|\hat{G}_n(|\hat{\varepsilon}_t|)-\tilde{W}_t|=o_{\mathrm{p}}(1)$, which implies

$$n^{-1/2}\sum_{t=k+1}^{n}\left\{\hat{G}_n(|\hat{\varepsilon}_t|)\hat{G}_n(|\hat{\varepsilon}_{t-k}|)-\tilde{W}_t\tilde{W}_{t-k}\right\}=o_{\mathrm{p}}(1).$$

Hence, to prove (S49), it remains to show that

$$n^{-1/2}\sum_{t=k+1}^{n}\left\{\tilde{W}_t\tilde{W}_{t-k}-G_n(|\hat{\varepsilon}_t|)G_n(|\hat{\varepsilon}_{t-k}|)\right\}-\kappa\left\{d_0^{*\mathrm{T}}n^{1/2}(\hat{\theta}_n-\theta_0)-v_0^*\right\}=o_{\mathrm{p}}(1). \quad \text{(S51)}$$

Notice first that (S6) holds under H_{1n} by the same arguments as those in the proof of Theorem 1. Moreover, for any $A>0$ and $n\geqslant n_0$, by (S88), (S90), (S91) and Lemma S5 we have

$$\sup_{\|u\|\leqslant A}\sup_{0\leqslant x<\infty}\frac{1}{n}\sum_{t=1}^{n}\left|G\{x\tilde{Z}_{t,n}(u)\}-G(x)\right| \\ \leqslant\frac{C}{n}\sum_{t=1}^{n}\rho^t\zeta_1+\frac{C}{n^{3/2}}\sum_{t=1}^{n}\left\{\sup_{n\geqslant n_0}\sup_{\theta\in\Theta}\left\|\frac{1}{h_{t,n}(\theta)}\frac{\partial h_{t,n}(\theta)}{\partial\theta}\right\|+r_{t,n_0}^{(u)}\right\}=O_{\mathrm{p}}(n^{-1/2}),$$

where $\tilde{Z}_{t,n}(u)$ is defined in (S86). This, together with the fact that $n^{1/2}(\hat{\theta}_n-\theta_0)=O_{\mathrm{p}}(1)$, implies that (S7) also holds under H_{1n}, and hence so does (S8). In addition, by (S43) and a method similar to that for (S9), we can show that

$$\sup_{\theta\in\Theta}\frac{1}{n}\sum_{t=1}^{n}\big||\tilde{\varepsilon}_{t,n}(\theta)|g\{|\tilde{\varepsilon}_{t,n}(\theta)|\}-|\varepsilon_{t,n}(\theta)|g\{|\varepsilon_{t,n}(\theta)|\}\big|=o_{\mathrm{p}}(1), \quad \text{(S52)}$$

which, combined with (S94), establishes (S10) under H_{1n}. As a result, (S11) holds under H_{1n}. Then, by a method similar to that for (S4), we can readily verify (S51) and hence (S49).

Furthermore, (S50) can be proved along the lines of (S2) in the proof of Theorem 1, while here we apply Lemmas S6 and S7 with $w_t=G(|\varepsilon_{t-k}|)$ and $w_t=G(|\varepsilon_{t+k}|)$, respectively. Finally, by a method similar to that in the proof of Theorem 1, we obtain

$$n^{1/2}\hat{\gamma}_k=0{\cdot}5\kappa\left\{(d_0^*-d_k^*)^{\mathrm{T}}n^{1/2}(\hat{\theta}_n-\theta_0)-(v_0^*-v_k^*)\right\}+n^{1/2}\gamma_k+o_{\mathrm{p}}(1),$$

where γ_k is defined as in the proof of Theorem 1, and similarly $\hat\gamma_0 = \gamma_0 + o_{\rm p}(1) = 1/12 + o_{\rm p}(1)$. Applying Slutsky's lemma, the martingale central limit theorem and the Cramér–Wold device, we accomplish the proof of this theorem. □

Proof of Theorem 4. By a method similar to that used for Theorems 2 and 3, we can show that $n^{1/2}\hat\gamma_k^{\Psi} = n^{1/2}\gamma_k^{\Psi} + 0{\cdot}5\kappa_\Psi\{d_k^{\Psi} n^{1/2}(\hat\theta_n - \theta_0) - v_k^{\Psi}\} + o_{\rm p}(1)$ for $k \geqslant 1$ and similarly verify that $\hat\gamma_0^{\Psi} = \gamma_0^{\Psi} + o_{\rm p}(1) = \sigma_\Psi^2 + o_{\rm p}(1)$ under H_{1n}. By Slutsky's lemma, the martingale central limit theorem and the Cramér–Wold device, we complete the proof of this theorem. □

S5·2. *Proof of Lemma* S3

First note that model (1) can be viewed as a GARCH(p^*+1, q^*+1) model with $\alpha_{01} = 0$ for $p < i \leqslant p^*+1$ and $\beta_{0j} = 0$ for $q < j \leqslant q^*+1$. Let $m = p^* + q^* + 1$, and define the $m \times m$ matrix A_{0t}^* written in block form by

$$A_{0t}^* = \begin{pmatrix} \varrho_{0t}^{*\rm T} & \beta_{0q^*+1} & \alpha_{02:p^*}^{\rm T} & \alpha_{0p^*+1} \\ I_{q^*} & 0 & 0 & 0 \\ \varepsilon_t^2 e_1^{\rm T} & 0 & 0 & 0 \\ 0 & 0 & I_{p^*-1} & 0 \end{pmatrix},$$

where $\varrho_{0t}^* = (\beta_{01} + \alpha_{01}\varepsilon_t^2, \beta_{02}, \ldots, \beta_{0q^*})^{\rm T}$, $\alpha_{02:p^*} = (\alpha_{02}, \ldots, \alpha_{0p^*})^{\rm T}$, I_k is the $k \times k$ identity matrix, and 0 denotes a zero vector or matrix with compatible dimensions. By Bougerol & Picard (1992), $\{y_t\}$ is a strictly stationary solution to model (1) if and only if $\gamma(A_0^*) < 0$, where

$$\gamma(A_0^*) = \inf_{0 \leqslant t < \infty} (t+1)^{-1} E(\log \|A_{00}^* \cdots A_{0t}^*\|);$$

see also Berkes et al. (2003). Let $z_{t,n} = (h_{t,n}, \ldots, h_{t-q^*,n}, y_{t-1,n}^2, \ldots, y_{t-p^*,n}^2)^{\rm T}$ and $z_t = (h_t, \ldots, h_{t-q^*}, y_{t-1}^2, \ldots, y_{t-p^*}^2)^{\rm T}$. Then the equations in (1) and (6) can be written equivalently as $z_{t+1} = A_{0t}^* z_t + \omega_0 e_1$ and $z_{t+1,n} = A_{0t}^* z_{t,n} + (\omega_0 + n^{-1/2}s_{t,n})e_1$, respectively. Consequently,

$$z_{t+1,n} - z_{t+1} = A_{0t}^*(z_{t,n} - z_t) + n^{-1/2}s_{t,n}e_1, \qquad \text{(S53)}$$

$$z_{t+1,n} - z_{t+1,n+1} = A_{0t}^*(z_{t,n} - z_{t,n+1}) + n^{-1/2}(s_{t,n} - s_{t,n+1})e_1. \qquad \text{(S54)}$$

For any $x = (x_1, \ldots, x_m)^{\rm T} \in [0, \infty)^m$, define the function $\tilde s$ by $\tilde s(x) = s(x_2, \ldots, x_m)$; then by Assumption 5 we have $\nabla\tilde s(x) = (0, \nabla s(x_2, \ldots, x_m)^{\rm T})^{\rm T} \geqslant 0$, where $\nabla\tilde s$ is the gradient of $\tilde s$. Define the $m \times m$ matrix $D(x) = (\nabla\tilde s(x), 0_{m\times(m-1)})^{\rm T}$, where $0_{m\times(m-1)}$ is an $m \times (m-1)$ zero matrix. Notice that $s_t = \tilde s(z_t)$ and $s_{t,n} = \tilde s(z_{t,n})$. It follows from (S53), (S54) and Taylor expansion that

$$z_{t+1,n} - z_{t+1} = \frac{s_t}{\sqrt{n}}e_1 + \left\{A_{0t}^* + \frac{D(z_{t,n}^*)}{\sqrt{n}}\right\}(z_{t,n} - z_t), \qquad \text{(S55)}$$

$$z_{t+1,n} - z_{t+1,n+1} = \left\{\frac{s_{t,n}}{\sqrt{n}} - \frac{s_{t,n}}{\sqrt{n+1}}\right\}e_1 + \left\{A_{0t}^* + \frac{D(z_{t,n}^{**})}{\sqrt{n+1}}\right\}(z_{t,n} - z_{t,n+1}), \qquad \text{(S56)}$$

where $z_{t,n}^*$ is between $z_{t,n}$ and z_t, and $z_{t,n}^{**}$ is between $z_{t,n}$ and $z_{t,n+1}$. Note that, by Assumptions 1 and 6, z_t and $z_{t,n}$ are almost surely finite for any $n \geqslant n_0$ and for all t. Since s_t, $s_{t,n}$, D and A_{0t}^* are all nonnegative, the recursive equations (S55) and (S56) imply that

$$z_t \leqslant z_{t,n+1} \leqslant z_{t,n}$$

for any $n \geqslant n_0$ and for all t. Moreover, by iterating (S53), we have

$$0 \leqslant \limsup_{n\to\infty} (z_{t+1,n} - z_{t+1}) = \limsup_{n\to\infty} n^{-1/2}\big(s_{t,n} + \sum_{k=0}^{\infty} A_{0t}^* \cdots A_{0t-k}^* s_{t-k-1,n}\big)e_1$$
$$\leqslant \limsup_{n\to\infty} n^{-1/2}\big(s_{t,n_0} + \sum_{k=0}^{\infty} A_{0t}^* \cdots A_{0t-k}^* s_{t-k-1,n_0}\big)e_1 = 0$$

with probability 1, where we have used the facts that $\gamma(A_0^*) < 0$ and $E(s_{t,n_0}^{\delta_0}) < \infty$. Finally, by iterating $z_{t+1,n_0} = A_{0t}^* z_{t,n_0} + (\omega_0 + n_0^{-1/2} s_{t,n_0})e_1$, we have that $z_{t+1,n_0} = (\omega_0 + n_0^{-1/2} s_{t,n_0})e_1 + \sum_{k=0}^{\infty} A_{0t}^* \cdots A_{0t-k}^* (\omega_0 + n_0^{-1/2} s_{t-k-1,n_0})e_1$. Then, along the lines of the proof of Lemma 2.3 in Berkes et al. (2003), we can show that there exists $0 < \iota_1 < \min(\delta_0, 1)$ such that $E(\|z_{t,n_0}\|^{\iota_1}) < \infty$. This completes the proof of the lemma.

S5·3. *Proof of Lemma* S4

One can see from (S70) that $h_{t,n}(\theta)$ can be written in the form

$$h_{t,n}(\theta) = \phi_0 + \sum_{\ell=1}^{\infty} \phi_\ell y_{t-\ell,n}^2,$$

where

$$\phi_0 = \omega \Big/ \Big(1 - \sum_{j=1}^{q} \beta_j\Big), \quad \phi_\ell = \sum_{i=1}^{\min(\ell,p)} e_1^{\mathrm{T}} B^{\ell-i} e_1 \alpha_i \quad (\ell = 1, 2, \ldots). \tag{S57}$$

To prove (b), we first introduce the following notation:

$${}_k h_{t,n}(\theta) = \phi_0 + \phi_k y_{t-k,n_0}^2 + \sum_{\ell=1,\ell\neq k}^{\infty} \phi_\ell y_{t-\ell,n}^2, \quad {}^k h_{t,n}(\theta) = \phi_0 + \phi_k y_{t-k}^2 + \sum_{\ell=1,\ell\neq k}^{\infty} \phi_\ell y_{t-\ell,n}^2,$$
$${}_k h_t(\theta) = \phi_0 + \phi_k y_{t-k,n_0}^2 + \sum_{\ell=1,\ell\neq k}^{\infty} \phi_\ell y_{t-\ell}^2, \quad {}^k h_t(\theta) = \phi_0 + \phi_k y_{t-k}^2 + \sum_{\ell=1,\ell\neq k}^{\infty} \phi_\ell y_{t-\ell,n_0}^2.$$

Consider $y_{t-k,n}^2 / h_{t,n}(\theta)$ as a function of $y_{t-k,n}^2$, which has the form $x \mapsto x/(a + bx)$ for some $a > 0$ and $b > 0$. Since this function is increasing on $(0, \infty)$ and $y_{t-k}^2 \leqslant y_{t-k,n}^2 \leqslant y_{t-k,n_0}^2$ for any $n \geqslant n_0$, we have $y_{t-k}^2 / {}^k h_{t,n}(\theta) \leqslant y_{t-k,n}^2 / h_{t,n}(\theta) \leqslant y_{t-k,n_0}^2 / {}_k h_{t,n}(\theta)$, which, together with the facts that ${}^k h_{t,n}(\theta) \leqslant {}^k h_t(\theta)$ and ${}_k h_{t,n}(\theta) \geqslant {}_k h_t(\theta)$, implies

$$\frac{y_{t-k}^2}{{}^k h_t(\theta)} \leqslant \frac{y_{t-k,n}^2}{h_{t,n}(\theta)} \leqslant \frac{y_{t-k,n_0}^2}{{}_k h_t(\theta)}. \tag{S58}$$

Moreover, for any $n \geqslant n_0$, since $y_t^2 \leqslant y_{t,n}^2 \leqslant y_{t,n_0}^2$,

$$h_t(\theta) \leqslant h_{t,n}(\theta) \leqslant h_{t,n_0}(\theta). \tag{S59}$$

As a result, by (S58), (S59) and (S75), for any $n \geqslant n_0$ we have

$$\frac{1}{h_{t,n}(\theta)} \frac{\partial h_{t,n}(\theta)}{\partial \beta_j} \geqslant \sum_{k=1}^{\infty} e_1^{\mathrm{T}} B_k^{(j)} e_1 \left\{ \frac{\omega}{h_{t,n_0}(\theta)} + \sum_{i=1}^{p} \alpha_i \frac{y_{t-k-i}^2}{{}_{k+i}h_t(\theta)} \right\}, \tag{S60}$$

$$\frac{1}{h_{t,n}(\theta)} \frac{\partial h_{t,n}(\theta)}{\partial \beta_j} \leqslant \sum_{k=1}^{\infty} e_1^{\mathrm{T}} B_k^{(j)} e_1 \left\{ \frac{\omega}{h_t(\theta)} + \sum_{i=1}^{p} \alpha_i \frac{y_{t-k-i,n_0}^2}{{}_{k+i}h_t(\theta)} \right\}. \tag{S61}$$

Similarly, we can obtain lower and upper bounds for the rest of the elements of $h_{t,n}^{-1}(\theta)\partial h_{t,n}(\theta)/\partial\theta$ for $n \geqslant n_0$:

$$\frac{1}{h_{t,n_0}(\theta)} \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B^k e_1 \leqslant \frac{1}{h_{t,n}(\theta)} \frac{\partial h_{t,n}(\theta)}{\partial \omega} \leqslant \frac{1}{h_t(\theta)} \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B^k e_1, \tag{S62}$$

$$\sum_{k=0}^{\infty} e_1^{\mathrm{T}} B^k e_1 \frac{y_{t-k-i}^2}{{}_{k+i}h_t(\theta)} \leqslant \frac{1}{h_{t,n}(\theta)} \frac{\partial h_{t,n}(\theta)}{\partial \alpha_i} \leqslant \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B^k e_1 \frac{y_{t-k-i,n_0}^2}{{}_{k+i}h_t(\theta)}. \tag{S63}$$

Denote by $\{Y_{t,n_0}^{(1l)}\}$ and $\{Y_{t,n_0}^{(1u)}\}$ the lower and upper bounds, respectively, in (S60)–(S63) evaluated at $\theta = \theta_0$. It can be verified that both processes are strictly stationary and ergodic and that for any $n \geqslant n_0$,

$$0 \leqslant Y_{t,n_0}^{(1l)} \leqslant \frac{1}{h_{t,n}(\theta_0)} \frac{\partial h_{t,n}(\theta_0)}{\partial \theta} \leqslant Y_{t,n_0}^{(1u)}.$$

Moreover, by Lemma S3, we have that for each fixed t, $\{Y_{t,n}^{(1l)}\}$ is a monotone increasing sequence, $\{Y_{t,n}^{(1u)}\}$ is a monotone decreasing sequence, and

$$\lim_{n\to\infty} Y_{t,n}^{(1l)} = \lim_{n\to\infty} Y_{t,n}^{(1u)} = \frac{1}{h_t} \frac{\partial h_t(\theta_0)}{\partial \theta}$$

with probability 1.

Turning now to the second-order derivatives, by (S58), (S59), (S79) and (S80), we can similarly show that for any $n \geqslant n_0$,

$$\frac{1}{h_{t,n}(\theta)} \frac{\partial^2 h_{t,n}(\theta)}{\partial \beta_j \partial \beta_{j'}} \geqslant \sum_{k=1}^{\infty} e_1^{\mathrm{T}} B_k^{(j,j')} e_1 \left\{ \frac{\omega}{h_{t,n_0}(\theta)} + \sum_{i=1}^{p} \alpha_i \frac{y_{t-k-i}^2}{{}_{k+i}h_t(\theta)} \right\}, \tag{S64}$$

$$\frac{1}{h_{t,n}(\theta)} \frac{\partial^2 h_{t,n}(\theta)}{\partial \beta_j \partial \beta_{j'}} \leqslant \sum_{k=1}^{\infty} e_1^{\mathrm{T}} B_k^{(j,j')} e_1 \left\{ \frac{\omega}{h_t(\theta)} + \sum_{i=1}^{p} \alpha_i \frac{y_{t-k-i,n_0}^2}{{}_{k+i}h_t(\theta)} \right\} \tag{S65}$$

and

$$\frac{1}{h_{t,n_0}(\theta)} \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B_k^{(j)} e_1 \leqslant \frac{1}{h_{t,n}(\theta)} \frac{\partial^2 h_{t,n}(\theta)}{\partial \omega \partial \beta_j} \leqslant \frac{1}{h_t(\theta)} \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B_k^{(j)} e_1, \tag{S66}$$

$$\sum_{k=0}^{\infty} e_1^{\mathrm{T}} B_k^{(j)} e_1 \frac{y_{t-k-i}^2}{{}_{k+i}h_t(\theta)} \leqslant \frac{1}{h_{t,n}(\theta)} \frac{\partial^2 h_{t,n}(\theta)}{\partial \alpha_i \partial \beta_j} \leqslant \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B_k^{(j)} e_1 \frac{y_{t-k-i,n_0}^2}{{}_{k+i}h_t(\theta)}. \tag{S67}$$

Denote by $\{Y_{t,n_0}^{(2l)}\}$ and $\{Y_{t,n_0}^{(2u)}\}$ the lower and upper bounds, respectively, in (S64)–(S67) evaluated at $\theta = \theta_0$. In a similar way, one can show that both processes are strictly stationary and ergodic and, for any $n \geqslant n_0$,

$$0 \leqslant Y_{t,n_0}^{(2l)} \leqslant \frac{1}{h_{t,n}(\theta_0)} \frac{\partial^2 h_{t,n}(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}} \leqslant Y_{t,n_0}^{(2u)}.$$

Again, it follows from Lemma S3 that for each fixed t, $\{Y_{t,n}^{(2l)}\}$ is monotone increasing, $\{Y_{t,n}^{(2u)}\}$ is monotone decreasing, and

$$\lim_{n\to\infty} Y_{t,n}^{(2l)} = \lim_{n\to\infty} Y_{t,n}^{(2u)} = \frac{1}{h_t} \frac{\partial^2 h_t(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}$$

with probability 1. In addition, the facts that $E(\|Y_{t,n_0}^{(1u)}\|^m) < \infty$ and $E(\|Y_{t,n_0}^{(2u)}\|^m) < \infty$ for any $m > 0$ are implied by the proof of Lemma S5. This completes the proof of Lemma S4.

S5·4. *Proof of Lemma* S5

Proof of (a): For any $\theta \in \Theta$, we can rewrite $h_{t,n}(\theta)$ in vector form as

$$\begin{pmatrix} h_{t,n}(\theta) \\ h_{t-1,n}(\theta) \\ \vdots \\ h_{t-q+1,n}(\theta) \end{pmatrix} = \begin{pmatrix} c_{t,n}(\theta) \\ 0 \\ \vdots \\ 0 \end{pmatrix} + B \begin{pmatrix} h_{t-1,n}(\theta) \\ h_{t-2,n}(\theta) \\ \vdots \\ h_{t-q,n}(\theta) \end{pmatrix} \tag{S68}$$

where $c_{t,n}(\theta) = \omega + \sum_{i=1}^{p} \alpha_i y_{t-i,n}^2$ and

$$B = \begin{pmatrix} \beta_1 & \dots & \beta_{q-1} & \beta_q \\ 1 & \cdots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & 1 & 0 \end{pmatrix}.$$

Let $\rho(B)$ be the spectral radius of the square matrix B. By Assumption 1(iii) we have

$$\sup_{\theta\in\Theta} \rho(B) < 1. \tag{S69}$$

Hence, iterating (S68) yields

$$h_{t,n}(\theta) = \sum_{k=0}^{t-1} e_1^{\mathrm{T}} B^k e_1 c_{t-k,n}(\theta) + e_1^{\mathrm{T}} B^t e_1 h_{0,n}(\theta) = \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B^k e_1 c_{t-k,n}(\theta), \tag{S70}$$

where the last equality holds almost surely for any $n \geqslant n_0$. Similarly, we have

$$\tilde{h}_{t,n}(\theta) = \sum_{k=0}^{t-p-1} e_1^{\mathrm{T}} B^k e_1 c_{t-k,n}(\theta) + \sum_{k=t-p}^{t-1} e_1^{\mathrm{T}} B^k e_1 \tilde{c}_{t-k,n}(\theta) + e_1^{\mathrm{T}} B^t e_1 \tilde{h}_0(\theta), \tag{S71}$$

where $\tilde{c}_{t,n}(\theta)$ is obtained by replacing $y_{0,n}^2, \ldots, y_{1-p,n}^2$ with their initial values in $c_{t,n}(\theta)$. By (S69)–(S71), Lemma S3 and the compactness of Θ, we have that for any $n \geqslant n_0$,

$$\begin{aligned}
&\sup_{\theta\in\Theta}\left|h_{t,n}(\theta)-\tilde{h}_{t,n}(\theta)\right| \\
&\quad\leqslant \sup_{\theta\in\Theta}\left|\sum_{i=1}^{p} e_1^{\mathrm{T}} B^{t-i} e_1\{c_{i,n}(\theta)-\tilde{c}_{i,n}(\theta)\} + e_1^{\mathrm{T}} B^{t} e_1\left\{h_{0,n}(\theta)-\tilde{h}_0(\theta)\right\}\right| \\
&\quad\leqslant \sup_{\theta\in\Theta}\left|\sum_{i=1}^{p} e_1^{\mathrm{T}} B^{t-i} e_1\{c_{i,n_0}(\theta)+\tilde{c}_{i,n_0}(\theta)\} + e_1^{\mathrm{T}} B^{t} e_1\left\{h_{0,n_0}(\theta)+\tilde{h}_0(\theta)\right\}\right| \\
&\quad\leqslant C\rho^t\zeta_1, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(S72)}
\end{aligned}$$

where ζ_1 is a random variable independent of t and n satisfying $E(|\zeta_1|^{\iota_1}) < \infty$ with ι_1 defined as in Lemma S3, whence the first result in (S43).

For any $j = 1, \ldots, q$, let $I^{(j)}$ be the $q \times q$ matrix whose $(1, j)$th element is 1 and other elements are all zero. For any positive integer k, let

$$B_k^{(j)} = \sum_{m=1}^{k} B^{m-1} I^{(j)} B^{k-m} \quad (j = 1, \ldots, q). \tag{S73}$$

Notice that, since $\beta_j I^{(j)} \leqslant B$ and Θ is compact, we have

$$B_k^{(j)} \leqslant \frac{k}{\beta_j} B^k \leqslant \frac{k}{\underline{\beta}} B^k, \tag{S74}$$

where $\underline{\beta} = \inf_{\theta\in\Theta}\min(\beta_1, \ldots, \beta_q) > 0$. By (S70) we have

$$\begin{aligned}
&\frac{\partial h_{t,n}(\theta)}{\partial\omega} = \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B^k e_1, \quad \frac{\partial h_{t,n}(\theta)}{\partial\alpha_i} = \sum_{k=0}^{\infty} e_1^{\mathrm{T}} B^k e_1 y_{t-k-i,n}^2, \\
&\frac{\partial h_{t,n}(\theta)}{\partial\beta_j} = \sum_{k=1}^{\infty} e_1^{\mathrm{T}} B_k^{(j)} e_1 c_{t-k,n}(\theta).
\end{aligned} \tag{S75}$$

Similarly, using (S71), we have

$$\frac{\partial\tilde{h}_{t,n}(\theta)}{\partial\omega} = \sum_{k=0}^{t-p-1} e_1^{\mathrm{T}} B^k e_1 + \sum_{k=t-p}^{t-1} e_1^{\mathrm{T}} B^k e_1 \frac{\partial\tilde{c}_{t-k,n}(\theta)}{\partial\omega} + e_1^{\mathrm{T}} B^t e_1 \frac{\partial\tilde{h}_0(\theta)}{\partial\omega}, \tag{S76}$$

$$\frac{\partial\tilde{h}_{t,n}(\theta)}{\partial\alpha_i} = \sum_{k=0}^{t-p-1} e_1^{\mathrm{T}} B^k e_1 y_{t-k-i,n}^2 + \sum_{k=t-p}^{t-1} e_1^{\mathrm{T}} B^k e_1 \frac{\partial\tilde{c}_{t-k,n}(\theta)}{\partial\alpha_i} + e_1^{\mathrm{T}} B^t e_1 \frac{\partial\tilde{h}_0(\theta)}{\partial\alpha_i}, \tag{S77}$$

$$\begin{aligned}
\frac{\partial\tilde{h}_{t,n}(\theta)}{\partial\beta_j} = {} & \sum_{k=1}^{t-p-1} e_1^{\mathrm{T}} B_k^{(j)} e_1 c_{t-k,n}(\theta) + \sum_{k=t-p}^{t-1} e_1^{\mathrm{T}} B_k^{(j)} e_1 \tilde{c}_{t-k,n}(\theta) + e_1^{\mathrm{T}} B_t^{(j)} e_1 \tilde{h}_0(\theta) \\
& + e_1^{\mathrm{T}} B^t e_1 \frac{\partial\tilde{h}_0(\theta)}{\partial\beta_j}.
\end{aligned} \tag{S78}$$

In view of (S69) and (S74)–(S78), using a method similar to that for (S72), we can prove the second result in (S43).

Furthermore, for any positive integer k, let

$$B_k^{(j,j')} = \sum_{m=2}^{k} B_{m-1}^{(j')} I^{(j)} B^{k-m} + \sum_{m=1}^{k-1} B^{m-1} I^{(j)} B_{k-m}^{(j')} \quad (j = 1, \ldots, q;\ j' = 1, \ldots, q),$$

where $B_k^{(j)}$ is defined in (S73). From (S75) we have

$$\frac{\partial^2 h_{t,n}(\theta)}{\partial \omega^2} = \frac{\partial^2 h_{t,n}(\theta)}{\partial \omega \partial \alpha_i} = \frac{\partial^2 h_{t,n}(\theta)}{\partial \alpha_i \partial \alpha_j} = 0, \quad \frac{\partial^2 h_{t,n}(\theta)}{\partial \omega \partial \beta_j} = \sum_{k=1}^{\infty} e_1^{\mathrm{T}} B_k^{(j)} e_1, \tag{S79}$$

$$\frac{\partial^2 h_{t,n}(\theta)}{\partial \alpha_i \partial \beta_j} = \sum_{k=1}^{\infty} e_1^{\mathrm{T}} B_k^{(j)} e_1 y_{t-k-i,n}^2, \quad \frac{\partial^2 h_{t,n}(\theta)}{\partial \beta_j \partial \beta_{j'}} = \sum_{k=2}^{\infty} e_1^{\mathrm{T}} B_k^{(j,j')} e_1 c_{t-k,n}(\theta), \tag{S80}$$

and the expressions with initial values can be obtained similarly. Note that by (S74) we have

$$B_k^{(j,j')} \leqslant \frac{k(k-1)}{\underline{\beta}^2} B^k. \tag{S81}$$

Then, using a method similar to that for (S72), (S44) can also be verified, and so the proof of (a) is complete.

Proof of (b): First, notice that (S69) implies $\sup_{\theta \in \Theta} e_1^{\mathrm{T}} B^{\ell} e_1 \leqslant C\rho^{\ell}$ for any integer $\ell \geqslant 0$. Then, by (S57) we have $\sup_{\theta \in \Theta} \phi_\ell \leqslant C\rho^{\ell}$ for $\ell \geqslant 0$. As a result, for any $0 < \delta < 1$ and $\ell \geqslant 1$,

$$\sup_{\theta \in \Theta} \frac{\phi_\ell y_{t-\ell,n_0}^2}{{}_\ell h_t(\theta)} \leqslant \frac{\phi_\ell y_{t-\ell,n_0}^2}{\underline{\omega}^{\delta} (\phi_\ell y_{t-\ell,n_0}^2)^{(1-\delta)}} \leqslant \frac{(C\rho^{\ell})^{\delta} y_{t-\ell,n_0}^{2\delta}}{\underline{\omega}^{\delta}}, \tag{S82}$$

where $\underline{\omega} = \inf_{\theta \in \Theta} \omega > 0$. Moreover, it follows from (S57), (S61) and (S74) that

$$\begin{aligned}
\sup_{n \geqslant n_0} \sup_{\theta \in \Theta} \left| \frac{1}{h_{t,n}(\theta)} \frac{\partial h_{t,n}(\theta)}{\partial \beta_j} \right| &\leqslant \sup_{\theta \in \Theta} \frac{\omega}{h_t(\theta)} \sum_{k=1}^{\infty} e_1^{\mathrm{T}} B_k^{(j)} e_1 + \sup_{\theta \in \Theta} \sum_{\ell=2}^{\infty} \sum_{i=1}^{\min(\ell,p)} e_1^{\mathrm{T}} B_{\ell-i}^{(j)} e_1 \alpha_i \frac{y_{t-\ell,n_0}^2}{{}_\ell h_t(\theta)} \\
&\leqslant \frac{C\overline{\omega}}{\underline{\omega}} \sum_{k=1}^{\infty} k\rho^k + \frac{1}{\underline{\beta}} \sum_{\ell=2}^{\infty} \ell \sup_{\theta \in \Theta} \left\{ \frac{1}{{}_\ell h_t(\theta)} \sum_{i=1}^{\min(\ell,p)} e_1^{\mathrm{T}} B^{\ell-i} e_1 \alpha_i y_{t-\ell,n_0}^2 \right\} \\
&\leqslant C + \frac{1}{\underline{\beta}} \sum_{\ell=2}^{\infty} \ell \sup_{\theta \in \Theta} \frac{\phi_\ell y_{t-\ell,n_0}^2}{{}_\ell h_t(\theta)},
\end{aligned} \tag{S83}$$

where $\overline{\omega} = \sup_{\theta \in \Theta} \omega \in (0, \infty)$. For any $m > 0$ and $\delta \in (0, \iota_1/m)$, where ι_1 is defined as in Lemma S3, by Lemma S3 and the Minkowski inequality we obtain

$$\left\| \sum_{\ell=2}^{\infty} \ell \frac{(C\rho^{\ell})^{\delta} y_{t-\ell,n_0}^{2\delta}}{\underline{\omega}^{\delta}} \right\|_m \leqslant \frac{C^{\delta}}{\underline{\omega}^{\delta}} \sum_{\ell=2}^{\infty} \ell \rho^{\delta \ell} \{E(|y_{t-\ell-i,n_0}^2|^{\delta m})\}^{1/m} < \infty,$$

which, together with (S82) and (S83), implies

$$E\left\{ \sup_{n \geqslant n_0} \sup_{\theta \in \Theta} \left| \frac{1}{h_{t,n}(\theta)} \frac{\partial h_{t,n}(\theta)}{\partial \beta_j} \right|^m \right\} < \infty. \tag{S84}$$

Similarly, using the upper bounds in (S62), (S63) and (S65)–(S67), we can establish (S45) for the rest of the quantities. Notice that the foregoing proof implies that $E(\|Y_{t,n_0}^{(1u)}\|^m < \infty$ and $E(\|Y_{t,n_0}^{(2u)}\|^m) < \infty$ for any $m > 0$, since these are the special cases where $\theta = \theta_0$.

For the third-order derivatives, in a similar fashion we can first obtain their upper bounds, which are independent of n as in the proof of Lemma S4, and then verify (S46) along the lines of the proof of (S84).

Finally, we prove (S47). For any $\theta \in \Theta$ and $r > 1$, define the set

$$U(r,\theta) = \left\{\theta^* = (\omega^*, \alpha_1^*, \ldots, \alpha_p^*, \beta_1^*, \ldots, \beta_q^*)' \in \Theta : \max_{1 \leqslant j \leqslant q} \frac{\beta_j^*}{\beta_j} \leqslant r\right\}.$$

To prove (S47), it suffices to verify a more general result: for any $m > 0$, there exists $r > 1$ such that

$$E\left[\left\{\sup_{n \geqslant n_0} \sup_{\theta \in \Theta} \sup_{\theta^* \in U(r,\theta)} \frac{h_{t,n}(\theta^*)}{h_{t,n}(\theta)}\right\}^m\right] < \infty. \tag{S85}$$

Note that for any θ, the set $U(r,\theta)$ imposes an upper bound only on the β_j^*, while the condition $\|\theta_1 - \theta_2\| \leqslant c$ restricts the distance between the parameter vectors θ_1 and θ_2. For any $\theta \in \Theta$, write $\phi_\ell = \phi_\ell(\theta)$ for $\ell \geqslant 0$, where ϕ_ℓ is defined in (S57). By the compactness of Θ, we have

$$\sup_{\theta \in \Theta} \sup_{\theta^* \in U(r,\theta)} \frac{\phi_\ell(\theta^*)}{\phi_\ell(\theta)} \leqslant C r^\ell$$

for any $\ell \geqslant 1$, and $\sup_{\theta \in \Theta} \phi_0(\theta) \leqslant C$. This, together with (S58), implies

$$\sup_{n \geqslant n_0} \sup_{\theta \in \Theta} \sup_{\theta^* \in U(r,\theta)} \frac{h_{t,n}(\theta^*)}{h_{t,n}(\theta)} \leqslant \frac{C}{\underline{\omega}} + C \sum_{\ell=1}^{\infty} r^\ell \sup_{\theta \in \Theta} \frac{\phi_\ell y_{t-\ell,n_0}^2}{\ell h_t(\theta)}.$$

Then, using (S82) and a method similar to that for (S84), we can show that (S85) holds for r close enough to 1. This completes the proof of the lemma.

S5·5. *Proofs of Lemmas* S6 *and* S7

Let $Z_{t,n} = h_{t,n}^{1/2}(\theta_0)/h_{t,n}^{1/2}$ and, for any $u \in \mathbb{R}^{p+q+1}$,

$$Z_{t,n}(u) = h_{t,n}^{1/2}(\theta_0 + n^{-1/2}u)/h_{t,n}^{1/2}, \quad \tilde{Z}_{t,n}(u) = \tilde{h}_{t,n}^{1/2}(\theta_0 + n^{-1/2}u)/h_{t,n}^{1/2}. \tag{S86}$$

Note that $h_{t,n} \geqslant h_{t,n}(\theta_0)$. For simplicity, without causing confusion we shall write, for any $u \in \mathbb{R}^{p+q+1}$,

$$h_{t,n}(u) = h_{t,n}(\theta_0 + n^{-1/2}u), \quad \tilde{h}_{t,n}(u) = \tilde{h}_{t,n}(\theta_0 + n^{-1/2}u),$$
$$\varepsilon_{t,n}(u) = \varepsilon_{t,n}(\theta_0 + n^{-1/2}u), \quad \tilde{\varepsilon}_{t,n}(u) = \tilde{\varepsilon}_{t,n}(\theta_0 + n^{-1/2}u).$$

LEMMA S8. *Suppose that* $L = \sup_{0<x<\infty} xg(x) < \infty$ *and that* $\{w_t\}$ *is a strictly stationary and ergodic process with* $w_t \in \mathcal{F}_{t-1}$ *and* $0 \leqslant w_t \leqslant 1$ *for all* t. *If Assumptions* 1, 3(i) *and* 5–7 *hold with* $E\{(r_{t,n_0}^{(u)})^2\} < \infty$, *then for any* $A > 0$,

$$\sup_{\|u\| \leqslant A} \sup_{0 \leqslant x < \infty} \left| n^{-1/2} \sum_{t=1}^{n} w_t \Big[I\big\{|\tilde{\varepsilon}_{t,n}(u)| \leqslant x\big\} - I(|\varepsilon_t| \leqslant x) - G\{x\tilde{Z}_{t,n}(u)\} + G(x) \Big] \right| = o_{\mathrm{p}}(1).$$

Proof of Lemma S8. For $x \in [0,\infty)$ and $u \in \mathbb{R}^{p+q+1}$, let

$$H_{k,n}(x,u) = \sum_{t=1}^{k} w_t \phi_{t,n}(x,u), \quad \phi_{t,n}(x,u) = \phi_{1t,n}(x,u) + \phi_{2t,n}(x,u),$$

where

$$\phi_{1t,n}(x,u) = [I\{|\varepsilon_t| \leqslant x\tilde{Z}_{t,n}(u)\} - G\{x\tilde{Z}_{t,n}(u)\}] - [I\{|\varepsilon_t| \leqslant xZ_{t,n}(u)\} - G\{xZ_{t,n}(u)\}],$$
$$\phi_{2t,n}(x,u) = [I\{|\varepsilon_t| \leqslant xZ_{t,n}(u)\} - G\{xZ_{t,n}(u)\}] - \{I(|\varepsilon_t| \leqslant x) - G(x)\}.$$

Note that $I\{|\varepsilon_t| \leqslant x\tilde{Z}_{t,n}(u)\} = I\{|\tilde{\varepsilon}_{t,n}(u)| \leqslant x\}$ and $I\{|\varepsilon_t| \leqslant xZ_{t,n}(u)\} = I\{|\varepsilon_{t,n}(u)| \leqslant x\}$.

As in the proof of Lemma S2, we prove this lemma in the following three steps:
(i) For any $A > 0$, there is a constant C depending on A such that for any $0 < x < \infty$ and u satisfying $\|u\| \leqslant A,\ \mathrm{pr}\{|H_{n,n}(x,u)| \geqslant sn^{1/2}\} \leqslant C/(s^4 n)$ for all $s > 0$.
(ii) For any $\|u\| \leqslant A$ with $A > 0,\ \sup_{0\leqslant x<\infty} |H_{n,n}(x,u)| = o_{\mathrm{p}}(n^{1/2})$.
(iii) For any $A > 0,\ \sup_{\|u\|\leqslant A} \sup_{0\leqslant x<\infty} |H_{n,n}(x,u)| = o_{\mathrm{p}}(n^{1/2})$.

We first verify (i). Let n be a fixed positive integer. Then for any $x > 0$ and $u \in \mathbb{R}^{p+q+1}$, $\{H_{k,n}(x,u), \mathcal{F}_k,\ k = 1,\ldots,n\}$ is a martingale. Applying Theorem 2.11 in Hall & Heyde (1980) and arguments similar to those for (S18), we obtain

$$E\{H_{n,n}^4(x,u)\} \leqslant C\left[\left\|\sum_{t=1}^{n}\left|G\{x\tilde{Z}_{t,n}(u)\} - G\{xZ_{t,n}(u)\}\right|\right\|_2^2 + \left\|\sum_{t=1}^{n}\left|G\{xZ_{t,n}(u)\} - G(x)\right|\right\|_2^2 + 1\right]. \quad \text{(S87)}$$

Similarly to (S19), by Taylor expansion and (S43) we can show that for any $n \geqslant n_0$,

$$\sup_{\|u\|\leqslant A} \sup_{0\leqslant x<\infty} \left|G\{x\tilde{Z}_{t,n}(u)\} - G\{xZ_{t,n}(u)\}\right| = \sup_{\|u\|\leqslant A} \sup_{0\leqslant x<\infty} \frac{0{\cdot}5x}{h_{t,n}^{1/2} h_{t,n}^{*1/2}}\, g\left(\frac{xh_{t,n}^{*1/2}}{h_{t,n}^{1/2}}\right)\left|\tilde{h}_{t,n}(u) - h_{t,n}(u)\right| \leqslant \frac{0{\cdot}5L}{\underline{\omega}} C\rho^t\zeta_1, \quad \text{(S88)}$$

where $h_{t,n}^*$ is between $\tilde{h}_{t,n}(u)$ and $h_{t,n}(u)$, and $\underline{\omega} = \inf_{\theta\in\Theta}\omega > 0$. This implies that

$$\left\|\sum_{t=1}^{n}\left|G\{x\tilde{Z}_{t,n}(u)\} - G\{xZ_{t,n}(u)\}\right|\right\|_2 \leqslant \sum_{t=1}^{n}\left\|G\{x\tilde{Z}_{t,n}(u)\} - G\{xZ_{t,n}(u)\}\right\|_2 \leqslant C. \quad \text{(S89)}$$

Similarly to (S21), for any $n \geqslant n_0$ we have

$$\begin{aligned}
&\sup_{\|u\|\leqslant A} \sup_{0\leqslant x<\infty} \left|G\{xZ_{t,n}(u)\} - G(xZ_{t,n})\right| \\
&= \frac{0{\cdot}5}{n^{1/2}} \sup_{0\leqslant x<\infty} \left|\frac{x}{h_{t,n}^{1/2}} g\left\{\frac{xh_{t,n}^{1/2}(\theta^*)}{h_{t,n}^{1/2}}\right\} \frac{u^{\mathrm{T}}}{h_{t,n}^{1/2}(\theta^*)} \frac{\partial h_{t,n}(\theta^*)}{\partial\theta}\right| \\
&\leqslant \frac{0{\cdot}5AL}{n^{1/2}} \sup_{n\geqslant n_0} \sup_{\theta\in\Theta} \left\|\frac{1}{h_{t,n}(\theta)} \frac{\partial h_{t,n}(\theta)}{\partial\theta}\right\|,
\end{aligned} \quad \text{(S90)}$$

where θ^* is between θ_0 and $\theta_0 + n^{-1/2}u$. Moreover, by Taylor expansion and Assumption 7, for any $n \geqslant n_0$ we have

$$\sup_{0\leqslant x<\infty} \left|G(xZ_{t,n}) - G(x)\right| \leqslant 0{\cdot}5 \sup_{0\leqslant x<\infty} \frac{x}{h_{t,n}^{1/2}} g\left(\frac{xh_{t,n}^{*1/2}}{h_{t,n}^{1/2}}\right) \frac{h_{t,n} - h_{t,n}(\theta_0)}{h_{t,n}^{*1/2}} \leqslant \frac{0{\cdot}5L}{n^{1/2}} r_{t,n_0}^{(u)}, \quad \text{(S91)}$$

where $h_{t,n}(\theta_0) \leqslant h^*_{t,n} \leqslant h_{t,n}$. Then, using (S90), (S91), (S45), the fact that $E\{(r^{(u)}_{t,n_0})^2\} < \infty$ and Minkowski's inequality, we have that for n large enough,

$$\left\| \sum_{t=1}^{n} \left| G\{xZ_{t,n}(u)\} - G(x) \right| \right\|_2 \leqslant Cn^{1/2}. \tag{S92}$$

Combining (S87), (S89) and (S92) and applying the Markov inequality, we establish (i).

The proof of (ii) can be accomplished along the lines of (ii) in the proof of Lemma S2. Similarly to (S26), we have

$$\sup_{0\leqslant x<\infty} |H_{n,n}(x,u)| \leqslant 3\tilde{A}_{1n} + 2\tilde{A}_{2n} + A_{3n} + \tilde{A}_{4n} + A_{5n},$$

where A_{3n} and A_{5n} are defined as in (S26) and

$$\tilde{A}_{1n} = \max_{1\leqslant j\leqslant N} |H_{n,n}(x_j,u)|, \quad \tilde{A}_{2n} = \max_{2\leqslant j\leqslant N} \sum_{t=1}^{n} w_t \left| G\{x_j \tilde{Z}_{t,n}(u)\} - G\{x_j Z_{t,n}(u)\} \right|,$$

$$\tilde{A}_{4n} = \max_{1\leqslant j\leqslant N} \sum_{t=1}^{n} w_t \big[G\{x_{j+1} Z_{t,n}(u)\} - G\{x_j Z_{t,n}(u)\} \big].$$

It is implied by the intermediate result (i) that $\tilde{A}_{1n} = o_{\mathrm{p}}(n^{1/2})$, and by (S88) that $\tilde{A}_{2n} = O_{\mathrm{p}}(1)$. Moreover, following arguments similar to those used for A_{4n} in the proof of Lemma S2, together with (S90) and (S91), we can show that $\tilde{A}_{4n} = \Delta O_{\mathrm{p}}(n^{1/2})$. Combining these with the established results for A_{3n} and A_{5n}, we complete the proof of (ii).

Finally, (iii) can be readily verified in a similar way to that in the proof of Lemma S2, with all $Z_t(u)$ and $\tilde{Z}_t(u)$ being replaced by $Z_{t,n}(u)$ and $\tilde{Z}_{t,n}(u)$, respectively; the lemma thus follows.□

Proof of Lemma S6. The proof of this lemma resembles that of Lemma A1. In view of Lemma S8, (S88) and the fact that $n^{1/2}(\hat{\theta}_n - \theta_0) = O_{\mathrm{p}}(1)$, it remains to show that for any $A > 0$,

$$\sup_{0\leqslant x<\infty} \sup_{\|u\|\leqslant A} \left| n^{-1/2} \sum_{t=1}^{n} w_t \big[G\{xZ_{t,n}(u)\} - G(x) \big] - 0{\cdot}5xg(x)(d_w^{\mathrm{T}} u - v_w) \right| = o_{\mathrm{p}}(1). \tag{S93}$$

By Assumption 3, for any $\Delta > 0$ we can choose $0 < C_1 < C_2 < \infty$ such that $\sup_{0<x\leqslant 2C_1} xg(x) \leqslant \Delta$ and $\sup_{C_2/2\leqslant x<\infty} xg(x) \leqslant \Delta$. By Taylor expansion and Assumption 7, for any $n \geqslant n_0$ we have

$$\begin{aligned}
\sup_{\|u\|\leqslant A} |Z_{t,n}(u) - 1| &\leqslant \sup_{\|u\|\leqslant A} |Z_{t,n}(u) - Z_{t,n}| + |Z_{t,n} - 1| \\
&\leqslant \sup_{\|u\|\leqslant A} \frac{0{\cdot}5}{n^{1/2}} \left| \frac{u^{\mathrm{T}}}{h^{1/2}_{t,n} h^{1/2}_{t,n}(\theta^*)} \frac{\partial h_{t,n}(\theta^*)}{\partial\theta} \right| + \frac{0{\cdot}5\{h_{t,n} - h_{t,n}(\theta_0)\}}{h^{1/2}_{t,n} h^{*1/2}_{t,n}} \\
&\leqslant \frac{0{\cdot}5A}{n^{1/2}} \sup_{\|\theta-\theta_0\|\leqslant c} \frac{h^{1/2}_{t,n}(\theta)}{h^{1/2}_{t,n}(\theta_0)} \sup_{n\geqslant n_0} \sup_{\theta\in\Theta} \left\| \frac{1}{h_{t,n}(\theta)} \frac{\partial h_{t,n}(\theta)}{\partial\theta} \right\| + \frac{0{\cdot}5}{n^{1/2}} r^{(u)}_{t,n_0},
\end{aligned}$$

where θ^* is between θ_0 and $\theta_0 + n^{-1/2}u$, and $h_{t,n}(\theta_0) \leqslant h^*_{t,n} \leqslant h_{t,n}$. Then, by (S45), (S47), Assumption 3 and the fact that $E\{(r^{(u)}_{t,n_0})^{4+\delta_1}\} < \infty$ with $\delta_1 > 0$, together with arguments similar to those for (S37) in the proof of Lemma A1, we can show that

$$\sup_{0\leqslant x<\infty}\max_{1\leqslant t\leqslant n}\sup_{\|u\|\leqslant A}\left|xZ_{t,n}(u)g\{xZ_{t,n}(u)\}-xg(x)\right|=o_{\rm p}(1). \tag{S94}$$

On the other hand, by Lemma S5(b), (S45), the ergodic theorem and the monotone convergence theorem, we can show that

$$\sup_{\|u\|\leqslant A}\left\|\frac{1}{n}\sum_{t=1}^{n}\frac{w_t}{h_{t,n}(u)}\frac{\partial h_{t,n}(u)}{\partial\theta}-d_w\right\|=o_{\rm p}(1). \tag{S95}$$

Similarly, it can be verified that

$$\left|\frac{1}{n}\sum_{t=1}^{n}w_t r_{t,n}\frac{h_{t,n}(\theta_0)}{h_{t,n}^{*}}-v_w\right|=o_{\rm p}(1) \tag{S96}$$

for $\{h_{t,n}^{*}\}$ satisfying $h_{t,n}(\theta_0)\leqslant h_{t,n}^{*}\leqslant h_{t,n}$.

Finally, by (S94)–(S96) and the Taylor expansions in (S90) and (S91), we have

$$\begin{aligned}
&\sup_{0\leqslant x<\infty}\sup_{\|u\|\leqslant A}\left|n^{-1/2}\sum_{t=1}^{n}w_t\big[G\{xZ_{t,n}(u)\}-G(x)\big]-0{\cdot}5xg(x)(d_w^{\rm T}u-v_w)\right|\\
&\quad\leqslant\sup_{0\leqslant x<\infty}\sup_{\|u\|\leqslant A}\left|n^{-1/2}\sum_{t=1}^{n}w_t\big[G\{xZ_{t,n}(u)\}-G(xZ_{t,n})\big]-0{\cdot}5xg(x)d_w^{\rm T}u\right|\\
&\qquad+\sup_{0\leqslant x<\infty}\left|n^{-1/2}\sum_{t=1}^{n}w_t\{G(xZ_{t,n})-G(x)\}+0{\cdot}5xg(x)v_w\right|\\
&\quad=\sup_{0\leqslant x<\infty}\sup_{\|u\|\leqslant A}\left|\frac{0{\cdot}5}{n}\sum_{t=1}^{n}xZ_{t,n}(u^{*})g\{xZ_{t,n}(u^{*})\}\frac{w_t u^{\rm T}}{h_{t,n}(u^{*})}\frac{\partial h_{t,n}(u^{*})}{\partial\theta}-0{\cdot}5xg(x)d_w^{\rm T}u\right|\\
&\qquad+\sup_{0\leqslant x<\infty}\left|-\frac{0{\cdot}5}{n}\sum_{t=1}^{n}xZ_{t,n}^{*}g(xZ_{t,n}^{*})w_t r_{t,n}\frac{h_{t,n}(\theta_0)}{h_{t,n}^{*}}+0{\cdot}5xg(x)v_w\right|\\
&\quad=o_{\rm p}(1),
\end{aligned}$$

where u^{*} is between zero and u, $h_{t,n}(\theta_0)\leqslant h_{t,n}^{*}\leqslant h_{t,n}$ and $Z_{t,n}^{*}=h_{t,n}^{*1/2}/h_{t,n}^{1/2}$. This proves (S93) and hence the lemma. □

Proof of Lemma S7. By Lemma S6,

$$\sup_{0\leqslant x<\infty}\left|n^{-1/2}\sum_{t=1}^{n}\left\{I(|\hat{\varepsilon}_t|\leqslant x)-I(|\varepsilon_t|\leqslant x)\right\}-xg(x)\left\{d_0^{*\rm T}n^{1/2}(\hat{\theta}_n-\theta_0)-v_0^{*}\right\}\right|=o_{\rm p}(1),$$

where $\hat{\varepsilon}_t=\tilde{\varepsilon}_{t,n}(\hat{\theta}_n)$. Hence we only need to show that for any $A>0$,

$$\sup_{\|u\|\leqslant A}\sup_{0\leqslant x<\infty}\left|n^{-1/2}\sum_{t=1}^{n}\{w_t-E(w_t)\}\big[I\{|\tilde{\varepsilon}_{t,n}(u)|\leqslant x\}-I(|\varepsilon_t|\leqslant x)\big]\right|=o_{\rm p}(1).$$

This can be accomplished by verifying steps (i)–(iii) as in the proof of Lemma S6 for $\tilde{H}_{n,n}(x,u)$, where

$$\tilde{H}_{k,n}(x,u)=\sum_{t=1}^{k}\{w_t-E(w_t)\}\tilde{\phi}_{t,n}(x,u),\quad \tilde{\phi}_{t,n}(x,u)=\tilde{\phi}_{1t,n}(x,u)+\tilde{\phi}_{2t,n}(x,u),$$

with

$$\tilde{\phi}_{1t,n}(x,u)=I\{|\varepsilon_t|\leqslant x\tilde{Z}_{t,n}(u)\}-I\{|\varepsilon_t|\leqslant xZ_{t,n}(u)\},$$
$$\tilde{\phi}_{2t,n}(x,u)=I\{|\varepsilon_t|\leqslant xZ_{t,n}(u)\}-I(|\varepsilon_t|\leqslant x).$$

Along the lines of the proof of (S39) and using methods similar to those in the proof of Lemma S6, we can readily establish (i)–(iii) and thereby complete the proof of this lemma. □

S6. Proofs of Theorems 5 and 6

S6·1. *Proof of Theorem* 5

Strong consistency: Write

$$\tilde{l}_{t,n}(\theta)=\log\tilde{h}_{t,n}^{1/2}(\theta)+\frac{|y_{t,n}|}{\tilde{h}_{t,n}^{1/2}(\theta)},\quad l_t(\theta)=\log h_t^{1/2}(\theta)+\frac{|y_t|}{h_t^{1/2}(\theta)},$$

where $\{y_{t,n}\}$ is generated by (6) and $\{y_t\}$ is generated by (1). Define $l_{t,n}(\theta)$ by replacing $\tilde{h}_{t,n}(\theta)$ with $h_{t,n}(\theta)$ in $\tilde{l}_{t,n}(\theta)$. Let $\tilde{L}_n(\theta)=n^{-1}\sum_{t=1}^{n}\tilde{l}_{t,n}(\theta)$ and $L_n(\theta)=n^{-1}\sum_{t=1}^{n}l_{t,n}(\theta)$.

To show the strong consistency, as in Huber (1967) and Francq & Zakoïan (2004) it suffices to establish the following intermediate results:

(C-i) $\sup_{\theta\in\Theta}|L_n(\theta)-\tilde{L}_n(\theta)|\to 0$ almost surely as $n\to\infty$.
(C-ii) $L_n(\theta_0)\to E\{l_t(\theta_0)\}$ almost surely as $n\to\infty$.
(C-iii) $E\{|l_t(\theta_0)|\}<\infty$, and if $\theta\neq\theta_0$ then $E\{l_t(\theta)\}>E\{l_t(\theta_0)\}$.
(C-iv) For any $\theta\neq\theta_0$, there exists a neighbourhood $V(\theta)$ such that, with probability 1,

$$\liminf_{n\to\infty}\inf_{\theta^*\in V(\theta)}\tilde{L}_n(\theta^*)>E\{l_t(\theta_0)\}.$$

We first prove (C-i). By Taylor expansion and (S43), we can show that for any $n\geqslant n_0$,

$$\sup_{\theta\in\Theta}\left|L_n(\theta)-\tilde{L}_n(\theta)\right|\leqslant\frac{1}{n}\sum_{t=1}^{n}\left(\frac{1}{2\underline{\omega}}+\frac{|y_{t,n_0}|}{2\underline{\omega}^{3/2}}\right)\sup_{\theta\in\Theta}\left|h_{t,n}(\theta)-\tilde{h}_{t,n}(\theta)\right|$$
$$\leqslant\frac{C}{n}\sum_{t=1}^{n}(1+|y_{t,n_0}|)\rho^t\zeta_1.$$

By the Cesàro lemma as in the proof of Theorem 2.1 in Francq & Zakoïan (2004), to prove (C-i) it suffices to show that $(1+|y_{t,n_0}|)\rho^t\zeta_1\to 0$ almost surely as $t\to\infty$. By the Markov inequality and Lemma S3, we have that for any $\varepsilon>0$,

$$\sum_{t=1}^{\infty}\mathrm{pr}\big\{(1+|y_{t,n_0}|)\rho^t\zeta_1>\varepsilon\big\}\leqslant\sum_{t=1}^{\infty}\frac{E\{(1+|y_{t,n_0}|)^{\iota_1}\rho^{\iota_1 t}\zeta_1^{\iota_1}\}}{\varepsilon^{\iota_1}}<\infty,$$

which, together with the Borel–Cantelli lemma, implies (C-i).

For (C-ii), by the ergodic theorem, it suffices to show that $n^{-1}\sum_{t=1}^{n}|l_{t,n}(\theta_0)-l_t(\theta_0)|\to 0$ almost surely as $n\to\infty$. By Taylor expansion, Lemma S3, (S59) and the ergodic theorem, we have

$$\begin{aligned}
&\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}|l_{t,n}(\theta_0)-l_t(\theta_0)| \\
&\leqslant \limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\log\frac{h_{t,n}^{1/2}(\theta_0)}{h_t^{1/2}}+\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{|\varepsilon_t|}{2}\frac{h_{t,n}-h_{t,n}(\theta_0)}{h_{t,n}^{1/2}(\theta_0)h_{t,n}^{*1/2}} \\
&\leqslant \limsup_{n\to\infty}\frac{1}{2n}\sum_{t=1}^{n}\log\frac{h_{t,n_0}(\theta_0)}{h_t}+\limsup_{n\to\infty}\frac{1}{2n^{3/2}}\sum_{t=1}^{n}|\varepsilon_t|r_{t,n_0}^{(u)} \\
&=\frac{1}{2}E\left\{\log\frac{h_{t,n_0}(\theta_0)}{h_t}\right\}
\end{aligned} \tag{S97}$$

with probability 1, where $h_{t,n}(\theta_0)\leqslant h_{t,n}^*\leqslant h_{t,n}$; in the last equality we have used the facts that $E[\log\{h_{t,n_0}(\theta_0)/h_t\}]\leqslant \iota_1^{-1}\log E\{h_{t,n_0}^{\iota_1}(\theta_0)\}-\log\underline{\omega}<\infty$ and $E\{|\varepsilon_t|r_{t,n_0}^{(u)}\}=E(r_{t,n_0}^{(u)})<\infty$. Applying the monotone convergence theorem, we have that the expectation in (S97) converges to zero almost surely as $n_0\to\infty$. This establishes (C-ii).

Now we prove (C-iii). First note that $E\{|l_t(\theta_0)|\}<\infty$, since $0{\cdot}5\log\underline{\omega}+1\leqslant E\{l_t(\theta_0)\}=0{\cdot}5E(\log h_t)+1<\infty$. In addition, using the fact that $x-1\geqslant\log x$ for any $x>0$, with equality if and only if $x=1$, we have

$$\begin{aligned}
E\{l_t(\theta)\}-E\{l_t(\theta_0)\}&=\frac{1}{2}E\left\{\log\frac{h_t(\theta)}{h_t}\right\}+E\left\{\frac{h_t^{1/2}}{h_t^{1/2}(\theta)}-1\right\}\\
&\geqslant\frac{1}{2}E\left\{\log\frac{h_t(\theta)}{h_t}\right\}+\frac{1}{2}E\left\{\log\frac{h_t}{h_t(\theta)}\right\}=0,
\end{aligned}$$

where equality holds if and only if $h_t(\theta)=h_t$ with probability 1. From the proof of Theorem 2.1 in Francq & Zakoïan (2004), there exists $t\in\mathbb{Z}$ such that $h_t(\theta)=h_t$ with probability 1 if and only if $\theta=\theta_0$. Hence (C-iii) follows.

Next we prove (C-iv). For any $\theta\in\Theta$ and any positive integer k, let $V_k(\theta)$ be the open ball with centre θ and radius $1/k$. It follows from (C-i) that

$$\begin{aligned}
\liminf_{n\to\infty}\inf_{\theta^*\in V_k(\theta)}\tilde{L}_n(\theta)&\geqslant\liminf_{n\to\infty}\inf_{\theta^*\in V_k(\theta)}L_n(\theta)-\limsup_{n\to\infty}\sup_{\theta\in\Theta}\left|L_n(\theta)-\tilde{L}_n(\theta)\right|\\
&\geqslant\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\inf_{\theta^*\in V_k(\theta)}l_{t,n}(\theta^*).
\end{aligned}$$

Moreover, by (S59) and Lemma S3,

$$\begin{aligned}
\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\inf_{\theta^*\in V_k(\theta)}l_{t,n}(\theta^*)&=\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\inf_{\theta^*\in V_k(\theta)}\left\{\frac{1}{2}\log h_{t,n}(\theta^*)+\frac{|y_{t,n}|}{h_{t,n}^{1/2}(\theta^*)}\right\}\\
&\geqslant\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\inf_{\theta^*\in V_k(\theta)}\left\{\frac{1}{2}\log h_t(\theta^*)+\frac{|y_t|}{h_{t,n_0}^{1/2}(\theta^*)}\right\}\\
&=E\left[\inf_{\theta^*\in V_k(\theta)}\left\{\frac{1}{2}\log h_t(\theta^*)+\frac{|y_t|}{h_{t,n_0}^{1/2}(\theta^*)}\right\}\right]
\end{aligned}$$

with probability 1, where we have used the ergodic theorem as in Francq & Zakoïan (2004): if $\{X_t\}$ is a stationary and ergodic process such that $E(X_t) \in \mathbb{R} \cup \{+\infty\}$, then $n^{-1}\sum_{t=1}^{n} X_t \to E(X_t)$ almost surely as $n \to \infty$. By the monotone convergence theorem, the expectation in the last equality increases to $E\{l_t(\theta)\}$ as k and n_0 tend to ∞. In view of (C-iii), (C-vi) holds. Finally, by a standard compactness argument, we establish strong consistency.

Asymptotic normality: In view of the Taylor expansion

$$0 = n^{-1/2}\sum_{t=1}^{n}\frac{\partial \tilde{l}_{t,n}(\hat{\theta}_n)}{\partial\theta} = n^{-1/2}\sum_{t=1}^{n}\frac{\partial \tilde{l}_{t,n}(\theta_0)}{\partial\theta} + \left\{\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 \tilde{l}_{t,n}(\theta^*)}{\partial\theta\partial\theta^{\mathrm{T}}}\right\} n^{1/2}(\hat{\theta}_n - \theta_0),$$

where θ^* is between $\hat{\theta}_n$ and θ_0, we first establish the following intermediate results:

(AN-i) $\|n^{-1/2}\sum_{t=1}^{n}\{\partial l_{t,n}(\theta_0)/\partial\theta - \partial\tilde{l}_{t,n}(\theta_0)/\partial\theta\}\| \to 0$ in probability as $n \to \infty$, and there exists a neighbourhood $V(\theta_0)$ of θ_0 such that $\sup_{\theta\in V(\theta_0)}\|n^{-1}\sum_{t=1}^{n}\{\partial^2 l_{t,n}(\theta)/(\partial\theta\partial\theta^{\mathrm{T}}) - \partial^2\tilde{l}_{t,n}(\theta)/(\partial\theta\partial\theta^{\mathrm{T}})\}\| \to 0$ in probability as $n \to \infty$.

(AN-ii) $n^{-1/2}\sum_{t=1}^{n}\partial l_{t,n}(\theta_0)/\partial\theta \to N[-\lambda/4,\ \{E(\varepsilon_t^2) - 1\}J/4]$ in distribution as $n \to \infty$.

(AN-iii) $n^{-1}\sum_{t=1}^{n}\partial^2 l_{t,n}(\theta^*)/(\partial\theta\partial\theta^{\mathrm{T}}) \to J/4$ in probability as $n \to \infty$.

Note that the matrix J is positive definite (Francq & Zakoïan, 2004). In addition, the derivatives of $l_{t,n}(\theta)$ are as follows:

$$\frac{\partial l_{t,n}(\theta)}{\partial\theta} = \frac{1}{2}\left\{1 - \frac{|y_{t,n}|}{h_{t,n}^{1/2}(\theta)}\right\}\frac{1}{h_{t,n}(\theta)}\frac{\partial h_{t,n}(\theta)}{\partial\theta},$$

$$\frac{\partial^2 l_{t,n}(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}} = \frac{1}{2}\left\{1 - \frac{|y_{t,n}|}{h_{t,n}^{1/2}(\theta)}\right\}\frac{1}{h_{t,n}(\theta)}\frac{\partial^2 h_{t,n}(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}} + \left\{\frac{3}{4}\frac{|y_{t,n}|}{h_{t,n}^{1/2}(\theta)} - \frac{1}{2}\right\}\frac{1}{h_{t,n}^2(\theta)}\frac{\partial h_{t,n}(\theta)}{\partial\theta}\frac{\partial h_{t,n}(\theta)}{\partial\theta^{\mathrm{T}}}.$$

By a method similar to that for verifying (C-i) above, we can show that for any $n \geqslant n_0$, $\|n^{-1/2}\sum_{t=1}^{n}\{\partial l_{t,n}(\theta_0)/\partial\theta - \partial\tilde{l}_{t,n}(\theta_0)/\partial\theta\}\|$ is bounded above by

$$Cn^{-1/2}\sum_{t=1}^{n}(1 + |y_{t,n_0}|)\left(1 + \|Y_{t,n_0}^{(1u)}\|\right)\rho^t\zeta_1$$

and $\sup_{\theta\in V(\theta_0)}\|n^{-1}\sum_{t=1}^{n}\{\partial^2 l_{t,n}(\theta)/(\partial\theta\partial\theta^{\mathrm{T}}) - \partial^2\tilde{l}_{t,n}(\theta)/(\partial\theta\partial\theta^{\mathrm{T}})\}\|$ is bounded above by

$$\frac{C}{n}\sum_{t=1}^{n}(1 + |y_{t,n_0}|)\left(1 + \sup_{n\geqslant n_0}\sup_{\theta\in\Theta}\left\|\frac{1}{h_{t,n}(\theta)}\frac{\partial^2 h_{t,n}(\theta)}{\partial\theta\partial\theta^{\mathrm{T}}}\right\| + \sup_{n\geqslant n_0}\sup_{\theta\in\Theta}\left\|\frac{1}{h_{t,n}^2(\theta)}\frac{\partial h_{t,n}(\theta)}{\partial\theta}\frac{\partial h_{t,n}(\theta)}{\partial\theta^{\mathrm{T}}}\right\|\right)\rho^t\zeta_1.$$

As a result, (AN-i) follows from the Markov inequality.

Next we verify (AN-ii). By Taylor expansion and an elementary calculation, we can show that

$$n^{-1/2}\sum_{t=1}^{n}\frac{\partial l_{t,n}(\theta_0)}{\partial\theta} = n^{-1/2}\sum_{t=1}^{n}X_{t,n} - \frac{1}{4n}\sum_{t=1}^{n}|\varepsilon_t|\frac{r_{t,n}}{h_{t,n}(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta} + R_n, \tag{S98}$$

where

$$X_{t,n} = \frac{1 - |\varepsilon_t|}{2}\frac{1}{h_{t,n}(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta}, \quad R_n = \frac{1}{4n^{3/2}}\sum_{t=1}^{n}|\varepsilon_t|\frac{r_{t,n}^2}{h_{t,n}(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta}\left\{\frac{h_{t,n}(\theta_0)}{h_{t,n}^*}\right\}^{3/2},$$

with $h_{t,n}(\theta_0) \leqslant h^*_{t,n} \leqslant h_{t,n}$. Then, by the ergodic theorem we have that for any $n \geqslant n_0$,

$$|R_n| \leqslant \frac{1}{n^{3/2}} \sum_{t=1}^{n} |\varepsilon_t| \left\| \frac{r^2_{t,n}}{h_{t,n}(\theta_0)} \frac{\partial h_{t,n}(\theta_0)}{\partial \theta} \right\| \leqslant \frac{1}{n^{3/2}} \sum_{t=1}^{n} |\varepsilon_t| \big(r^{(u)}_{t,n_0}\big)^2 \big\|Y^{(1u)}_{t,n_0}\big\| = o_{\mathrm{p}}(1), \quad \text{(S99)}$$

where we have used (S45) and the fact that $E\{(r^{(u)}_{t,n_0})^{2+\delta_1}\} < \infty$ for some $\delta_1 > 0$.

Notice that $\{X_{t,n}, \mathcal{F}_{t-1}\}_t$ is a strictly stationary martingale difference with $E(X_{t,n}X^{\mathrm{T}}_{t,n}) < \infty$ for each $n \geqslant n_0$. We will next use the Lindeberg central limit theorem for triangular arrays of martingale differences and the Cramér–Wold device to show that

$$n^{-1/2} \sum_{t=1}^{n} X_{t,n} \to N\left[0,\ \frac{1}{4}\{E(\varepsilon_t^2) - 1\} J\right] \quad \text{(S100)}$$

in distribution as $n \to \infty$. For $c \in \mathbb{R}^{p+q+1}$, let $x_{t,n} = c^{\mathrm{T}} X_{t,n}$. By the ergodic theorem,

$$\begin{aligned} \limsup_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E(x^2_{t,n} \mid \mathcal{F}_{t-1}) &\leqslant \frac{1}{4}\{E(\varepsilon_t^2) - 1\} c^{\mathrm{T}} \limsup_{n\to\infty} \left\{ \frac{1}{n} \sum_{t=1}^{n} Y^{(1u)}_{t,n_0} \big(Y^{(1u)}_{t,n_0}\big)^{\mathrm{T}} \right\} c \\ &= \frac{1}{4}\{E(\varepsilon_t^2) - 1\} c^{\mathrm{T}} E\Big\{ Y^{(1u)}_{t,n_0} \big(Y^{(1u)}_{t,n_0}\big)^{\mathrm{T}} \Big\} c \end{aligned}$$

with probability 1. Similarly, we can show that, with probability 1,

$$\liminf_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} E(x^2_{t,n} \mid \mathcal{F}_{t-1}) \geqslant \frac{1}{4}\{E(\varepsilon_t^2) - 1\} c^{\mathrm{T}} E\Big\{ Y^{(1l)}_{t,n_0} \big(Y^{(1l)}_{t,n_0}\big)^{\mathrm{T}} \Big\} c.$$

Then it follows from the monotone convergence theorem that

$$\frac{1}{n} \sum_{t=1}^{n} E(x^2_{t,n} \mid \mathcal{F}_{t-1}) \to \frac{1}{4}\{E(\varepsilon_t^2) - 1\} c^{\mathrm{T}} J c \quad \text{(S101)}$$

almost surely as $n \to \infty$. Moreover, by Hölder's inequality and the Markov inequality, we can show that for any $\varepsilon > 0$,

$$\frac{1}{n} \sum_{t=1}^{n} E\Big\{ x^2_{t,n} I\big(|x_{t,n}| \geqslant n^{1/2}\varepsilon\big) \Big\} \to 0$$

as $n \to \infty$, where $I(\cdot)$ is the indicator function. Combining this with (S101), by the Lindeberg central limit theorem and the Cramér–Wold device we obtain (S100).

In addition, similarly to (S101), we can verify that

$$-\frac{1}{4n} \sum_{t=1}^{n} |\varepsilon_t| \frac{r_{t,n}}{h_{t,n}(\theta_0)} \frac{\partial h_{t,n}(\theta_0)}{\partial \theta} \to -\frac{1}{4}\lambda$$

in probability as $n \to \infty$, which, in conjunction with (S98)–(S100), implies (AN-ii).

Now we prove (AN-iii). It is implied by (S46) and the strong consistency of $\hat{\theta}^{\mathrm{LQML}}_n$ that

$$\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_{t,n}(\theta^*)}{\partial \theta \partial \theta^{\mathrm{T}}} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2 l_{t,n}(\theta_0)}{\partial \theta \partial \theta^{T}} + o_{\mathrm{p}}(1).$$

Furthermore, by methods similar to those for (S98) and (S99), we can show that

$$\frac{1}{n}\sum_{t=1}^{n}\frac{\partial^2 l_{t,n}(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}=\frac{1}{n}\sum_{t=1}^{n}\frac{3|\varepsilon_t|-2}{4}\frac{1}{h_{t,n}^2(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta^{\mathrm{T}}}+\frac{1}{n}\sum_{t=1}^{n}\frac{1-|\varepsilon_t|}{2}\frac{1}{h_{t,n}(\theta_0)}\frac{\partial^2 h_{t,n}(\theta_0)}{\partial\theta\partial\theta^{\mathrm{T}}}+o_{\mathrm{p}}(1). \tag{S102}$$

For the first term on the right-hand side of (S102), by the ergodic theorem we have

$$\begin{aligned}&\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{3|\varepsilon_t|-2}{4}\frac{1}{h_{t,n}^2(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta^{\mathrm{T}}}\\&\leqslant\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{3|\varepsilon_t|-2}{4}\left\{I\left(|\varepsilon_t|\geqslant\frac{2}{3}\right)Y_{t,n_0}^{(1u)}\big(Y_{t,n_0}^{(1u)}\big)^{\mathrm{T}}+I\left(|\varepsilon_t|<\frac{2}{3}\right)Y_{t,n_0}^{(1l)}\big(Y_{t,n_0}^{(1l)}\big)^{\mathrm{T}}\right\}\\&=E\left\{\frac{3|\varepsilon_t|-2}{4}I\left(|\varepsilon_t|\geqslant\frac{2}{3}\right)Y_{t,n_0}^{(1u)}\big(Y_{t,n_0}^{(1u)}\big)^{\mathrm{T}}\right\}+E\left\{\frac{3|\varepsilon_t|-2}{4}I\left(|\varepsilon_t|<\frac{2}{3}\right)Y_{t,n_0}^{(1l)}\big(Y_{t,n_0}^{(1l)}\big)^{\mathrm{T}}\right\}\end{aligned}$$

with probability 1. Then, by the monotone convergence theorem and the fact that ε_t is independent of both $Y_{t,n_0}^{(1l)}$ and $Y_{t,n_0}^{(1l)}$, we have that the sum of the two expectations converges to $J/4$ as $n_0\to\infty$. Similarly, we can show that

$$\liminf_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{3|\varepsilon_t|-2}{4}\frac{1}{h_{t,n}^2(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta^{\mathrm{T}}}\geqslant J/4$$

with probability 1, and hence the first term on the right-hand side of (S102) converges to $J/4$ almost surely as $n\to\infty$. Along the same lines, we can show that the second term on the right-hand side of (S102) converges to zero almost surely as $n\to\infty$. Thus, (AN-iii) holds. Applying (AN-i)–(AN-iii) and Slutsky's lemma, we accomplish the proof of the theorem.

S6·2. *Proof of Theorem* 6

Strong consistency: Write

$$\tilde{l}_{t,n}(\theta)=|\log y_{t,n}^2-\log\tilde{h}_{t,n}(\theta)|,\quad l_t(\theta)=|\log y_t^2-\log h_t(\theta)|,$$

and let $l_{t,n}(\theta)$, $\tilde{L}_n(\theta)$ and $L_n(\theta)$ be defined in the same way as in the proof of Theorem 5.

The strong consistency can be proved in a similar way to Theorem 5, but unlike the proof of (C-ii) therein, no moment condition on $r_{t,n_0}^{(u)}$ will be required. Indeed, for $\hat{\theta}_n^{\mathrm{LAD}}$, (S97) will be replaced by

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}|l_{t,n}(\theta_0)-l_t(\theta_0)|\leqslant\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\log\frac{h_{t,n_0}}{h_t}+\limsup_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\log\frac{h_{t,n_0}(\theta_0)}{h_t}.$$

Then, by arguments similar to those following (S97), we can show that the right-hand side converges to zero almost surely as n and n_0 tend to ∞, without imposing any moment condition on $r_{t,n_0}^{(u)}$. The rest of the proof is standard and proceeds along the same lines as the proof of Theorem 5.

Asymptotic normality: The proof of asymptotic normality for $\hat{\theta}_n^{\mathrm{LAD}}$ under H_{1n} mimics that under H_0 accomplished by Chen & Zhu (2015). For any $u\in\Lambda=\{u:\theta_0+u\in\Theta\}$, let

$\tilde{D}_n(u) = \sum_{t=1}^{n}\{\tilde{l}_{t,n}(\theta_0+u) - \tilde{l}_{t,n}(\theta_0)\}$. Notice that for $x \neq 0$,

$$|x-y| - |x| = -y\,\mathrm{sgn}(x) + 2\int_0^y \{I(x \leqslant s) - I(x \leqslant 0)\}\,\mathrm{d}s,$$

where $\mathrm{sgn}(x) = I(x>0) - I(x<0)$; see Knight (1998). Let $\epsilon_t = \log \varepsilon_t^2$. Then, by an elementary calculation, we have

$$\tilde{D}_n(u) = -\sum_{t=1}^{n} \tilde{q}_{t,n}(u)\,\mathrm{sgn}(\epsilon_t - \tilde{m}_{t,n}) + 2\sum_{t=1}^{n}\int_0^{\tilde{q}_{t,n}(u)} \tilde{I}_{t,n}(s)\,\mathrm{d}s,$$

where $\tilde{q}_{t,n}(u) = \log \tilde{h}_{t,n}(\theta_0+u) - \log \tilde{h}_{t,n}(\theta_0)$, $\tilde{m}_{t,n} = \log \tilde{h}_{t,n}(\theta_0) - \log h_{t,n}$, and $\tilde{I}_{t,n}(s) = I(\epsilon_t \leqslant s + \tilde{m}_{t,n}) - I(\epsilon_t \leqslant \tilde{m}_{t,n})$.

We first show that

$$\tilde{D}_n(u) = D_n(u) + O_{\mathrm{p}}(\|u\|) \tag{S103}$$

holds uniformly in $u \in \Lambda$, where

$$D_n(u) = -\sum_{t=1}^{n} q_{t,n}(u)\,\mathrm{sgn}(\epsilon_t - m_{t,n}) + 2\sum_{t=1}^{n}\int_0^{q_{t,n}(u)} I_{t,n}(s)\,\mathrm{d}s,$$

with $q_{t,n}(u) = \log h_{t,n}(\theta_0+u) - \log h_{t,n}(\theta_0)$, $m_{t,n} = \log h_{t,n}(\theta_0) - \log h_{t,n}$, and $I_{t,n}(s) = I(\epsilon_t \leqslant s + m_{t,n}) - I(\epsilon_t \leqslant m_{t,n})$.

Note that

$$\tilde{D}_n(u) - D_n(u) = R_{1n}(u) + R_{2n}(u) + R_{3n}(u), \tag{S104}$$

where

$$\begin{aligned}
R_{1n}(u) &= \sum_{t=1}^{n}\{q_{t,n}(u) - \tilde{q}_{t,n}(u)\}\,\mathrm{sgn}(\epsilon_t - \tilde{m}_{t,n}) + 2\sum_{t=1}^{n}\int_{q_{t,n}(u)}^{\tilde{q}_{t,n}(u)} \tilde{I}_{t,n}(s)\,\mathrm{d}s,\\
R_{2n}(u) &= \sum_{t=1}^{n} q_{t,n}(u)\{\mathrm{sgn}(\epsilon_t - m_{t,n}) - \mathrm{sgn}(\epsilon_t - \tilde{m}_{t,n})\},\\
R_{3n}(u) &= 2\sum_{t=1}^{n}\int_0^{q_{t,n}(u)} \{\tilde{I}_{t,n}(s) - I_{t,n}(s)\}\,\mathrm{d}s.
\end{aligned}$$

By (S43), it is straightforward to show that for any $n \geqslant n_0$,

$$\sup_{u\in\Lambda}\frac{1}{\|u\|}|R_{1n}(u)| \leqslant 3\sup_{u\in\Lambda}\frac{1}{\|u\|}\sum_{t=1}^{n}|q_{t,n}(u) - \tilde{q}_{t,n}(u)| \leqslant C\sum_{t=1}^{n}\rho^t\zeta_1 = O_{\mathrm{p}}(1). \tag{S105}$$

Denote by $G_\epsilon(\cdot)$ and $g_\epsilon(\cdot)$ the cumulative distribution function and the density function of ϵ_t, respectively. Notice that $g_\epsilon(x) = 0{\cdot}5\exp(0{\cdot}5x)g\{\exp(0{\cdot}5x)\}$ for any $-\infty < x < \infty$. By Assumption 3, we have that g_ϵ is continuous on $(-\infty,\infty)$ with $\lim_{x\to-\infty} g_\epsilon(x) = 0$ and $\lim_{x\to\infty} g_\epsilon(x) = 0$, which implies $\sup_{-\infty<x<\infty} g_\epsilon(x) < \infty$. Then, by Lemma S4, (S43), Jensen's inequality and Hölder's inequality, for any $n \geqslant n_0$ and the constant $\iota_1 \in (0,1)$ defined in Lemma S3 we can show that

$$\begin{aligned}
&E\left[\left\{\sup_{u\in\Lambda}\frac{1}{\|u\|}|R_{2n}(u)|\right\}^{\iota_1/2}\right]\\
&\leqslant E\left(\left[\sum_{t=1}^{n}\left\|Y_{t,n_0}^{(1u)}\right\|\{\operatorname{sgn}(\epsilon_t-m_{t,n})-\operatorname{sgn}(\epsilon_t-\tilde{m}_{t,n})\}\right]^{\iota_1/2}\right)\\
&\leqslant \sum_{t=1}^{n}E\Big(\left\|Y_{t,n_0}^{(1u)}\right\|^{\iota_1/2}E\big[\{2I(\epsilon_t<\tilde{m}_{t,n})-2I(\epsilon_t<m_{t,n})\}^{\iota_1/2}\mid\mathcal{F}_{t-1}\big]\Big)\\
&\leqslant \sum_{t=1}^{n}E\Big[\left\|Y_{t,n_0}^{(1u)}\right\|^{\iota_1/2}\{2G_\epsilon(\tilde{m}_{t,n})-2G_\epsilon(m_{t,n})\}^{\iota_1/2}\Big]\\
&\leqslant \left\{2\sup_{-\infty<x<\infty}g_\epsilon(x)\right\}^{\iota_1/2}\sum_{t=1}^{n}\left\{E\Big(\left\|Y_{t,n_0}^{(1u)}\right\|^{\iota_1}\Big)\right\}^{1/2}\{E(|\tilde{m}_{t,n}-m_{t,n}|^{\iota_1})\}^{1/2}\\
&\leqslant C\sum_{t=1}^{n}\rho^{\iota_1 t/2}<\infty.
\end{aligned}$$

As a result,

$$\sup_{u\in\Lambda}\frac{1}{\|u\|}|R_{2n}(u)|=O_{\mathrm{p}}(1). \tag{S106}$$

Similarly, we can show that $\sup_{u\in\Lambda}|R_{3n}(u)|/\|u\|=O_{\mathrm{p}}(1)$, which, in conjunction with (S104)–(S106), implies (S103).

Since $q_{t,n}(u)=q_{1t,n}(u)+q_{2t,n}(u)$, where

$$\begin{aligned}
q_{1t,n}(u)&=\frac{u^{\mathrm{T}}}{h_{t,n}(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta},\\
q_{2t,n}(u)&=\frac{u^{\mathrm{T}}}{2}\left\{\frac{1}{h_{t,n}(\theta^*)}\frac{\partial^2 h_{t,n}(\theta^*)}{\partial\theta\partial\theta^{\mathrm{T}}}-\frac{1}{h_{t,n}^2(\theta^*)}\frac{\partial h_{t,n}(\theta^*)}{\partial\theta}\frac{\partial h_{t,n}(\theta^*)}{\partial\theta^{\mathrm{T}}}\right\}u
\end{aligned}$$

with θ^* lying between θ_0 and θ_0+u^*, we can decompose $D_n(u)$ as

$$D_n(u)=(n^{1/2}u)^{\mathrm{T}}T_n+\Pi_{1n}(u)+\Pi_{2n}(u)+\Pi_{3n}(u),$$

where

$$\begin{aligned}
T_n&=-\frac{1}{n^{1/2}}\sum_{t=1}^{n}\frac{\operatorname{sgn}(\epsilon_t-m_{t,n})}{h_{t,n}(\theta_0)}\frac{\partial h_{t,n}(\theta_0)}{\partial\theta},\\
\Pi_{1n}(u)&=\sum_{t=1}^{n}E\big[M_{t,n}(u)-E\{M_{t,n}(u)\mid\mathcal{F}_{t-1}\}\big],\quad \Pi_{2n}(u)=\sum_{t=1}^{n}E\{M_{t,n}(u)\mid\mathcal{F}_{t-1}\},\\
\Pi_{3n}(u)&=-\sum_{t=1}^{n}q_{2t,n}(u)\operatorname{sgn}(\epsilon_t-m_{t,n})+2\sum_{t=1}^{n}\int_{q_{1t,n}(u)}^{q_{t,n}(u)}I_{t,n}(s)\,\mathrm{d}s,
\end{aligned}$$

with $M_{t,n}(u)=2\int_0^{q_{1t,n}(u)}I_{t,n}(s)\,\mathrm{d}s$. Let $u_n=\hat{\theta}_n^{\mathrm{LAD}}-\theta_0$. By arguments similar to those used for Lemmas 2.2 and 2.3 in Zhu & Ling (2011), we can show that $\Pi_{1n}(u_n)=o_{\mathrm{p}}(n^{1/2}\|u_n\|+$

$n\|u_n\|^2)$, $\Pi_{2n}(u_n) = (n^{1/2}u_n)^{\mathrm{T}}\{g_\epsilon(0)J\}(n^{1/2}u_n)$, and $\Pi_{3n}(u_n) = o_{\mathrm{p}}(n\|u_n\|^2)$. Thus, by (S103), we have

$$\tilde{D}_n(u_n) = (n^{1/2}u)^{\mathrm{T}}T_n + (n^{1/2}u_n)^{\mathrm{T}}\{g_\epsilon(0)J\}(n^{1/2}u_n) + o_{\mathrm{p}}(n^{1/2}\|u_n\| + n\|u_n\|^2).$$

Moreover, by methods similar to those used to show (AN-ii) in the proof of Theorem 5 and the techniques for proving Lemma 2.1 in Zhu & Ling (2011), we can show that $T_n \to N[-2g_\epsilon(0)\lambda, J]$ in distribution as $n \to \infty$.

Finally, since $g_\epsilon(0) = g(1)/2$, by applying the arguments for Theorem 2.2 in Zhu & Ling (2011), we accomplish the proof of this theorem.

REFERENCES

BERKES, I. & HORVÁTH, L. (2003). Limit results for the empirical process of squared residuals in GARCH models. *Stoch. Proces. Appl.* **105**, 271–98.

BERKES, I., HORVÁTH, L. & KOKOSZKA, P. (2003). GARCH processes: Structure and estimation. *Bernoulli* **9**, 201–27.

BOUGEROL, P. & PICARD, N. (1992). Stationarity of GARCH processes and of some nonnegative time series. *J. Economet.* **52**, 115–27.

CHEN, M. & ZHU, K. (2015). Sign-based portmanteau test for ARCH-type models with heavy-tailed innovations. *J. Economet.* **189**, 313–20.

DVORETZKY, A., KIEFER, J. & WOLFOWITZ, J. (1956). Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *Ann. Math. Statist.* **27**, 642–69.

FRANCQ, C. & ZAKOÏAN, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* **10**, 605–37.

FRANCQ, C. & ZAKOÏAN, J.-M. (2009). Testing the nullity of GARCH coefficients: Correction of the standard tests and relative efficiency comparisons. *J. Am. Statist. Assoc.* **104**, 313–24.

HALL, P. & HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press.

HALLIN, M., INGENBLEEK, J.-F. & PURI, M. L. (1985). Linear serial rank tests for randomness against ARMA alternatives. *Ann. Statist.* **13**, 1156–81.

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Ann. Math. Statist.* **19**, 293–325.

HUBER, P. J. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1. Berkeley, California: University of California Press, pp. 221–33.

KNIGHT, K. (1998). Limiting distributions for L_1 regression estimators under general conditions. *Ann. Statist.* **26**, 755–70.

MASSART, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Prob.* **18**, 1269–83.

PENG, L. & YAO, Q. (2003). Least absolute deviation estimation for ARCH and GARCH models. *Biometrika* **90**, 967–75.

RESNICK, S. I. (2007). *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. New York: Springer.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. New York: John Wiley & Sons.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. New York: Cambridge University Press.

ZHENG, Y., ZHU, Q., LI, G. & XIAO, Z. (2016). Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *arXiv:* 1610.07453 [stat.ME].

ZHU, K. & LING, S. (2011). Global self-weighted and local quasi-maximum exponential likelihood estimators for ARMA-GARCH/IGARCH models. *Ann. Statist.* **39**, 2131–63.