Supplementary Material to "Quasi-maximum Likelihood Inference for Linear Double Autoregressive Models"

Hua Liu, Songhua Tan and Qianqian Zhu

School of Statistics and Management

Shanghai University of Finance and Economics

Abstract: This supplementary material includes additional results for the simulation and the empirical analysis, as well as technical details for Theorems 1–5 and Remarks 1–3. To show Theorems 1–3, Theorem 5 and Remark 3, Lemmas 1–8 are also introduced. Throughout the supplement, the notation C is a generic constant which may take different values from lines to lines. The norm of a matrix or column vector is defined as $\|A\| = \sqrt{\mathrm{tr}(AA')} = \sqrt{\sum_{i,j} |a_{ij}|^2}$.

S1 Additional simulation results

This section reports additional simulation results which are useful but not reported in the manuscript. Section S1.1 evaluates the finite-sample performance of all inference tools for additional innovation distributions. Section S1.2 explores the size and power of two portmanteau tests $Q(M)$ and $Q^G(M)$ with respect to the lag order M and the sample size n by simulation. Finally, the simulation results for $Q(M)$ and $Q^G(M)$ with $M = 12$, or 18, are presented in Section S1.3.

S1.1 Simulation results for additional innovation distributions

In the first, third and fourth experiments in Sections 4.1, 4.3 and 4.4, owing to space limitations, we only reported the results with $\{\eta_t\}$ being *i.i.d.* normal, Laplace, or Student's

t_3 distributed random variables. This subsection provides additional results with $\{\eta_t\}$ being *i.i.d.* Student's t_2 or t_5 distributed random variables, or skewed Student's t_3 distributed random variables with the skew parameter being -1.5 (denoted by $st_{3,-1.5}$).

Table S.1 presents the biases, ESDs, and ASDs of $\widehat{\boldsymbol{\theta}}_n$ and $\widetilde{\boldsymbol{\theta}}_n$, when the innovations follow the t_2, t_5, or $st_{3,-1.5}$ distribution. The following findings in Section 4.1 remain unchanged: (1) for the E-QMLE $\widehat{\boldsymbol{\theta}}_n$, the ESDs and ASDs of scale-type estimators $\widehat{\omega}_n$ and $\widehat{\beta}_n$ increase as the distribution of η_t becomes more heavy tailed, while that of location-type estimator $\widehat{\alpha}_n$ are decreasing; (2) for the G-QMLE $\widetilde{\boldsymbol{\theta}}_n$, the ESDs and ASDs increase as the distribution of η_t gets more heavy tailed. If $\eta_t \sim t_2$, then $E(\eta_t^2) = \infty$ such that the G-QMLE is not applicable, which results in the inferior performance of the G-QMLE in this case. Comparing Table S.1 with Table 1 in the manuscript, we further have the following findings: (1) the E-QMLE performs satisfactorily with small biases and ESDs close to ASDs even for t_2 distribution with $E(\eta_t^2) = \infty$, indicating that the E-QMLE is more robust for thick tails than G-QMLE; (2) nonzero skewness will increase the bias of the E-QMLE, but it cannot affect its ESD and ASD. While the G-QMLE still has small biases but large ESDs and ASDs in fitting skewed data.

Table S.2 provides the percentages of underfitting, correct selection, and overfitting cases by the BIC and BIC^G when the innovations follow the t_2, t_5, or $st_{3,-1.5}$ distribution. It can be seen that the conclusions in Section 4.3 of the manuscript are unchanged. Moreover, we find that nonzero skewness could improve the selection accuracy of BIC but worsen that of BIC^G. This is probably because nonzero skewness almost has no effects on ESDs and ASDs of the E-QMLE $\widehat{\boldsymbol{\theta}}_n$, but it increases ESDs and ASDs of the G-QMLE $\widetilde{\boldsymbol{\theta}}_n$.

Table S.3 summarizes the rejection rates of $Q(6)$ and $Q^G(6)$ at the 5% significance level

when the innovations follow the t_2, t_5, or $st_{3,-1.5}$ distribution. The following findings in Section 4.4 remain unchanged: (1) all sizes are close to the nominal rate except for the t_2 case, and the power increases as the sample size n or the departure level increases. This is expected, because the E-QMLE and G-QMLE are not applicable if η_t follows t_2 distribution such that $E(\eta_t^2) = \infty$, making the size inaccurate; (2) for the same level of departures, $Q(6)$ and $Q^G(6)$ are more powerful in terms of detecting the misspecification in the conditional mean ($c_1 \neq 0, c_2 = 0$) than they are in doing so in the conditional standard deviation ($c_1 = 0$, $c_2 > 0$); (3) as the innovation distribution becomes more heavy tailed, $Q(6)$ and $Q^G(6)$ perform worse in detecting misspecifications in the conditional standard deviation.

S1.2 Portmanteau tests with respect to n and M

In this subsection, we conduct simulation studies to examine the size and power of portmanteau tests $Q(M)$ and $Q^G(M)$ with respect to the sample size n and the lag order M.

To evaluate the size with respect to n and M, we generate 1000 replications of sample size $n = 100, 200, \ldots, 2000$, from the following model:

$$y_t = 0.1y_{t-1} + \eta_t(1 + 0.2|y_{t-1}|),$$

where $\{\eta_t\}$ are *i.i.d.* normal or Laplace distributed random variables. Here $\{\eta_t\}$ are standardized with median zero and $E(|\eta_t|) = 1$ to evaluate $Q(M)$ using the E-QMLE, and $\{\eta_t\}$ are standardized with mean zero and $\text{var}(\eta_t) = 1$ to evaluate $Q^G(M)$ using the G-QMLE. We fit a linear DAR model with $p = 1$ using the Gaussian or exponential quasi-maximum likelihood estimation. Figure 1 plots the empirical size of $Q(M)$ and $Q^G(M)$ at the 5% significance level, for different settings of $M = 2, 4, \ldots, 50$ and $n = 100, 200, \ldots, 2000$. We have the following findings: (1) the size with $M > 10$ is decreasing with respect to the lag M, and

most of the sizes with $M > 20$ are less than 0.05; (2) the size is insensitive to the sample size n and the distribution of η_t. Therefore, we suggest $M \leq 20$ to avoid undersize for two portmanteau tests, which is consistent with the guideline given by Box et al. (2015).

To evaluate the power with respect to n and M, we generate 1000 replications of sample size $n = 100, 200, \ldots, 2000$, from the following model:

$$y_t = 0.1y_{t-1} + 0.1y_{t-2} + \eta_t(1 + 0.2|y_{t-1}| + 0.2|y_{t-2}|),$$

where all other settings are the same as in the above experiment for size evaluation. Figure 2 plots the empirical power with different settings of n and M. We have the following findings: (1) the power is linearly increasing with respect to the sample size n, i.e. power $= O(n)$; (2) the ln(power) is linearly decreasing with respect to the lag M, i.e. ln(power) $= O(M)$, or power $= O(e^M)$. As a result, to avoid power loss for tests with a large lag M, sample size $n = O(e^M)$ may be used, or the lag order $M = O(\ln(n))$ may be suggested for a fixed sample size, which is consistent with the suggestion by Tsay (2005). In practice, we can choose several values of M, such as $M_1 = \lfloor \ln(n) \rfloor$, $M_2 = 2\lfloor \ln(n) \rfloor, \ldots$, where $\lfloor x \rfloor$ denotes the largest integer not greater than x.

In summary, based on the above simulation studies, we may suggest that the rule-of-thumb for multiple choices of M could be

$$(M_1, M_2, \ldots, M_K) = (\lfloor \ln(n) \rfloor, 2\lfloor \ln(n) \rfloor, \ldots, K\lfloor \ln(n) \rfloor), \tag{S1.1}$$

where K is the maximum value of k, such that $M_k = k\lfloor \ln(n) \rfloor \leq 20$.

S1.3 Portmanteau tests with $M = 12$ or 18

In the fourth experiment in Section 4.4, owing to space limitations, we only reported the results of $Q(M)$ and $Q^G(M)$ with $M = 6$. This subsection provides the results for $M = 12$,

and 18. Table S.4 reports the rejection rates of $Q(M)$ and $Q^G(M)$ at the 5% significance level for $M = 12$ or 18. The following findings in Section 4.4 in the manuscript remain unchanged: (1) the power increases as the sample size n or the departure level increases; (2) for the same level of departures, $Q(M)$ and $Q^G(M)$ for $M = 12$ or 18 are more powerful in terms of detecting the misspecification in the conditional mean ($c_1 \neq 0, c_2 = 0$) than they are in doing so in the conditional standard deviation ($c_1 = 0$, $c_2 > 0$); (3) as the innovation distribution becomes more heavy tailed, $Q(M)$ and $Q^G(M)$ perform worse in detecting misspecifications in the conditional standard deviation. In addition, most of the sizes of $Q(M)$ and $Q^G(M)$ with $M = 12$ or 18 are less than the nominal level, and the power of $Q(M)$ and $Q^G(M)$ decreases slowly as M increases.

Table S.1: Biases ($\times 10$), ESDs, and ASDs of the E-QMLE $\widehat{\boldsymbol{\theta}}_n$ and G-QMLE $\widetilde{\boldsymbol{\theta}}_n$ when the innovations follow the t_2, t_5, or $st_{3,-1.5}$ distribution.

		t_2			t_5			$st_{3,-1.5}$		
	n	Bias	ESD	ASD	Bias	ESD	ASD	Bias	ESD	ASD
					E-QMLE					
α	500	-0.034	0.045	0.047	-0.040	0.059	0.061	-0.004	0.053	0.056
	1000	-0.022	0.031	0.033	-0.011	0.042	0.043	-0.004	0.037	0.039
ω	500	0.112	0.138	0.144	0.066	0.078	0.083	-0.072	0.090	0.094
	1000	0.057	0.098	0.108	0.031	0.055	0.059	-0.075	0.062	0.068
β	500	-0.096	0.115	0.100	-0.060	0.053	0.053	-0.043	0.070	0.068
	1000	-0.051	0.079	0.076	-0.027	0.038	0.038	-0.038	0.050	0.049
					G-QMLE					
α	500	-0.056	0.073	0.067	-0.037	0.054	0.054	-0.008	0.061	0.058
	1000	-0.023	0.053	0.049	-0.012	0.039	0.038	-0.005	0.043	0.042
ω	500	-0.365	0.249	0.222	0.036	0.091	0.095	-0.016	0.181	0.157
	1000	-0.312	0.213	0.202	0.020	0.068	0.071	-0.008	0.145	0.135
β	500	0.428	0.581	0.298	-0.072	0.090	0.085	0.000	0.239	0.165
	1000	0.479	0.569	0.284	-0.036	0.067	0.063	0.002	0.217	0.145

Table S.2: Percentages of underfitted, correctly selected, and overfitted models by the BIC and BIC^G when the innovations follow the t_2, t_5, or $st_{3,-1.5}$ distribution.

		t_2			t_5			$st_{3,-1.5}$		
	n	Under	Exact	Over	Under	Exact	Over	Under	Exact	Over
BIC	300	7.2	89.7	3.1	30.9	68.9	0.2	14.1	85.0	0.9
	500	0.5	95.8	3.7	7.2	92.7	0.1	1.5	98.1	0.4
	1000	0.0	97.0	3.0	0.2	99.8	0.0	0.0	99.6	0.4
BIC^G	300	29.2	48.6	22.2	36.6	61.2	2.2	36.6	54.4	9.0
	500	10.7	59.5	29.8	11.2	87.5	1.3	14.6	73.4	12.0
	1000	1.3	65.3	33.4	0.4	98.0	1.6	0.7	83.2	16.1

Table S.3: Rejection rates of the tests $Q(6)$ and $Q^G(6)$ at the 5% significance level, when the innovations follow the t_2, t_5, or $st_{3,-1.5}$ distribution.

			t_2		t_5		$st_{3,-1.5}$	
	c_1	c_2	500	1000	500	1000	500	1000
Q	0.0	0.0	0.077	0.090	0.057	0.064	0.056	0.066
	0.1	0.0	0.253	0.478	0.203	0.475	0.278	0.534
	0.3	0.0	0.910	0.985	0.996	1.000	0.991	1.000
	0.0	0.1	0.090	0.152	0.096	0.227	0.088	0.164
	0.0	0.3	0.284	0.593	0.734	0.978	0.409	0.792
Q^G	0.0	0.0	0.151	0.168	0.048	0.062	0.062	0.081
	0.1	0.0	0.367	0.607	0.182	0.460	0.298	0.553
	0.3	0.0	0.937	0.994	0.998	1.000	0.994	1.000
	0.0	0.1	0.143	0.177	0.064	0.126	0.077	0.132
	0.0	0.3	0.231	0.367	0.420	0.871	0.270	0.582

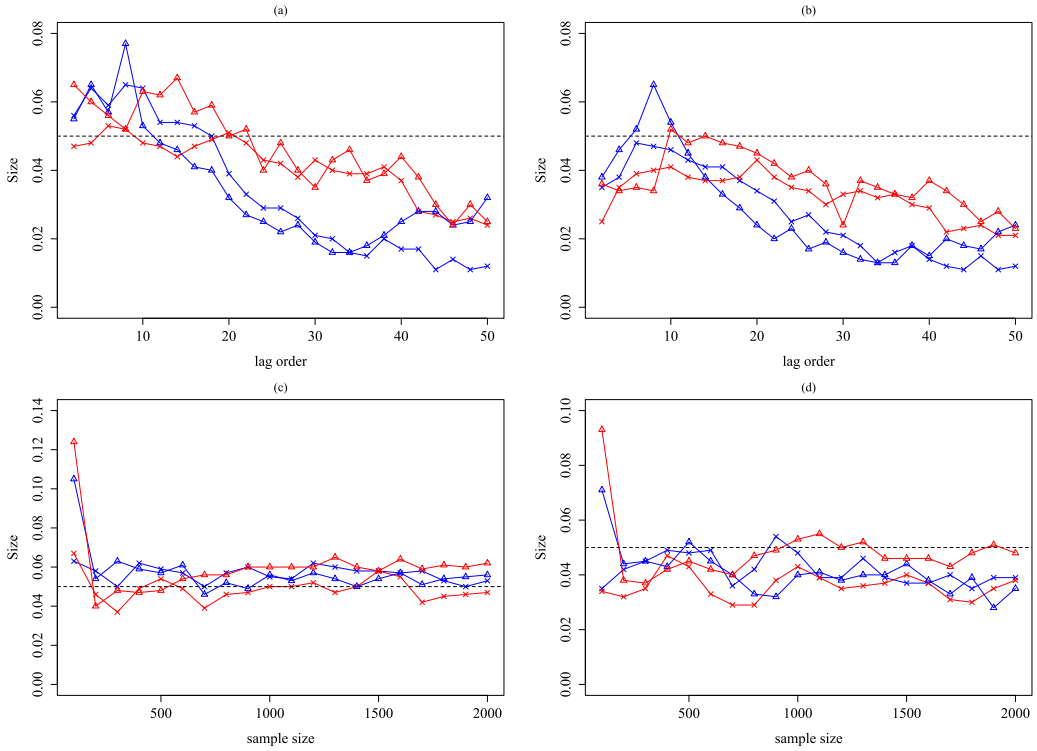

Figure 1: The empirical size of $Q(M)$ (a) and $Q^G(M)$ (b) with respect to the lag order M at the significance level 0.05, where the line with color blue or red presents $N = 500$ or 2000, the symbol $\times$ or $\triangle$ represents η_t being a standard normal or Laplace random variable; the empirical size of $Q(M)$ (c) and $Q^G(M)$ (d) with respect to the sample size n at the significance level 0.05, where the line with color blue or red presents $M = 6$ or 12, the symbol $\times$ or $\triangle$ represents η_t being a standard normal or Laplace random variable.

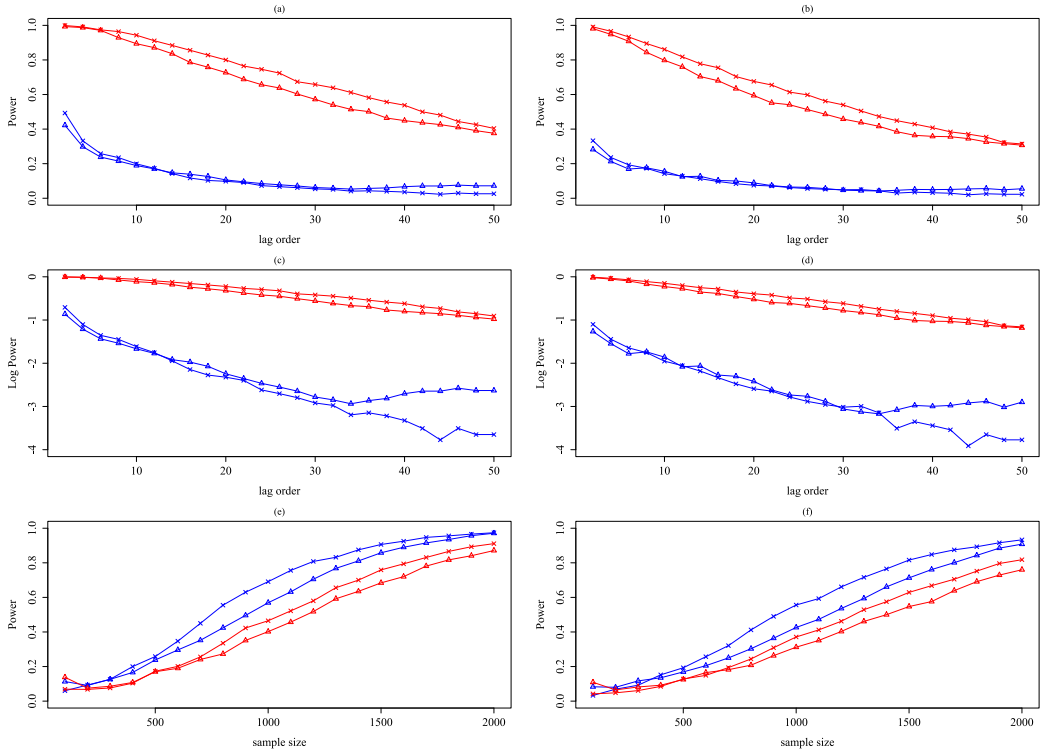

Figure 2: The empirical power of $Q(M)$ (a) and $Q^G(M)$ (b) with respect to the lag order M at the significance level 0.05, where the line with color blue or red presents $N = 500$ or 2000, the symbol $\times$ or $\triangle$ represents η_t being a standard normal or Laplace random variable; (c) and (d) correspond to the logarithmic power version of (a) and (b), respectively; the empirical power of $Q(M)$ (e) and $Q^G(M)$ (f) with respect to the sample size n at the significance level 0.05, where the line with color blue or red presents $M = 6$ or 12, the symbol $\times$ or $\triangle$ represents η_t being a standard normal or Laplace random variable.

Table S.4: Rejection rates of the tests $Q(M)$ and $Q^G(M)$ at the 5% significance level with M=12 or 18, where the innovations follow the normal, Laplace, or Student's t_3 distribution.

			Normal		Laplace		t_3	
	c_1	c_2	500	1000	500	1000	500	1000
					$M = 12$			
Q	0.0	0.0	0.031	0.038	0.042	0.055	0.048	0.043
	0.1	0.0	0.140	0.334	0.154	0.337	0.144	0.361
	0.3	0.0	0.964	1.000	0.973	1.000	0.946	0.999
	0.0	0.1	0.092	0.219	0.073	0.129	0.062	0.125
	0.0	0.3	0.712	0.997	0.426	0.900	0.301	0.742
Q^G	0.0	0.0	0.041	0.032	0.064	0.060	0.130	0.106
	0.1	0.0	0.143	0.299	0.179	0.383	0.275	0.422
	0.3	0.0	0.981	1.000	0.986	1.000	0.951	1.000
	0.0	0.1	0.078	0.119	0.082	0.105	0.130	0.143
	0.0	0.3	0.506	0.946	0.230	0.626	0.241	0.440
					$M = 18$			
Q	0.0	0.0	0.028	0.040	0.031	0.043	0.041	0.047
	0.1	0.0	0.096	0.244	0.117	0.254	0.113	0.283
	0.3	0.0	0.914	1.000	0.907	1.000	0.855	0.999
	0.0	0.1	0.060	0.155	0.063	0.108	0.057	0.120
	0.0	0.3	0.510	0.988	0.268	0.793	0.232	0.634
Q^G	0.0	0.0	0.043	0.029	0.055	0.043	0.124	0.091
	0.1	0.0	0.097	0.211	0.136	0.282	0.213	0.350
	0.3	0.0	0.924	1.000	0.941	1.000	0.891	0.999
	0.0	0.1	0.044	0.084	0.062	0.092	0.129	0.121
	0.0	0.3	0.321	0.874	0.158	0.493	0.185	0.339

S2 Diagnosis for model (5.2)

For model (5.2) fitted using the G-QMLE, Figure 3 illustrates the QQ plots of the fitted residuals $\{\widetilde{\eta}_t\}$ against the Students' t_2, t_3, or t_4 distribution, indicating that probably $E(\eta_t^2) < \infty$ and $E(\eta_t^4) = \infty$, and thus the G-QMLE may be unsuitable for the data $\{y_t\}$. The KSD test is also used to check whether η_t follows the standard normal or Laplace distribution. The p-value of for the normal distribution test is less than 0.01, and for the Laplace distribution test is 0.65. We can conclude that the E-QMLE may be more suitable for the data $\{y_t\}$.

To check ACFs of residuals $\{\widetilde{\eta}_t\}$ jointly, we employ the portmanteau test $Q_1(M)$ with $M = 6, 12$, and 18, and the p-values are $0.71, 0.37$, and 0.11, respectively. Moreover, Figure 4 plots the residual ACFs $\widetilde{\rho}_k$'s and $\widetilde{\gamma}_k$'s up to lag 18, and they all fall within their corresponding 95% pointwise confidence intervals (CIs). Clearly almost all residual ACFs are nonsignificant both individually and jointly at the 5% significance level, and hence the fitted model at (5.2) is adequate.

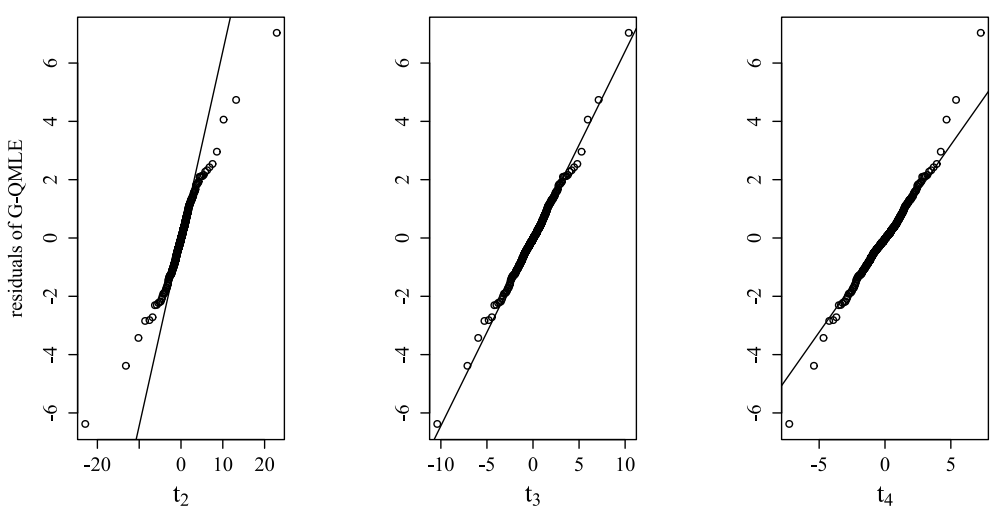

Figure 3: QQ plots of the residuals $\{\widetilde{\eta}_t\}$ against the Student's t_2 (left panel), t_3 (middle panel), and t_4 (right panel) distributions.

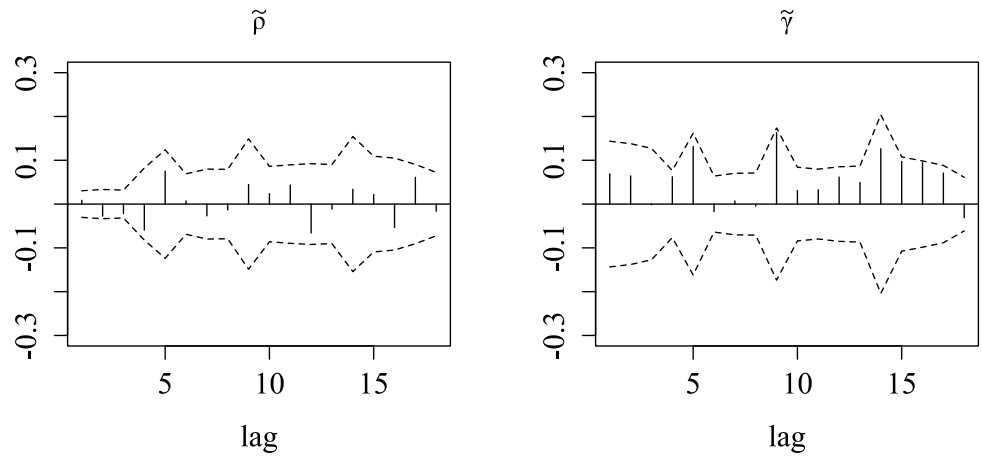

Figure 4: Residual ACF plots for $\widetilde{\rho}_l$ (left panel) and $\widetilde{\gamma}_l$ (right panel), where the dashed lines are the corresponding 95% pointwise confidence intervals.

S3 Technical Proofs

S3.1 Proof of Theorem 1

To show Theorem 1, we introduce the following lemma.

Lemma 1. *For any $\boldsymbol{\theta}^* \in \Theta$, let $B_\eta(\boldsymbol{\theta}^*) = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| < \eta\}$ be an open neighborhood of $\boldsymbol{\theta}^*$ with radius $\eta > 0$. If Assumptions 1, 2 and 3(i) hold, then*

(i) $E \sup_{\boldsymbol{\theta} \in \Theta} \left|\ell_t^E(\boldsymbol{\theta})\right| < \infty$;

(ii) $E\left[\ell_t^E(\boldsymbol{\theta})\right]$ *has a unique minimum at* $\boldsymbol{\theta}_0$;

(iii) $E\left[\sup_{\boldsymbol{\theta} \in B_\eta(\boldsymbol{\theta}^*)} |\ell_t^E(\boldsymbol{\theta}) - \ell_t^E(\boldsymbol{\theta}^*)|\right] \to 0$ *as* $\eta \to 0$.

Proof. Recall that $\ell_t^E(\boldsymbol{\theta}) = \ln h_t(\boldsymbol{\delta}) + h_t^{-1}(\boldsymbol{\delta})|\varepsilon_t(\boldsymbol{\alpha})|$, where $\varepsilon_t(\boldsymbol{\alpha}) = y_t - \sum_{i=1}^p \alpha_i y_{t-i}$ and $h_t(\boldsymbol{\delta}) = \omega + \sum_{i=1}^p \beta_i |y_{t-i}|$. We first show (i). Denote $c_i = \sup_{\boldsymbol{\theta} \in \Theta} |\alpha_{i0} - \alpha_i|$, then by Assumption 1 we have $c_i < \infty$ for $i = 1, \ldots, p$. Note that $\varepsilon_t(\boldsymbol{\alpha}) = \sum_{i=1}^p (\alpha_{i0} - \alpha_i) y_{t-i} + \varepsilon_t(\boldsymbol{\alpha}_0)$ and $\varepsilon_t(\boldsymbol{\alpha}_0) = \eta_t h_t(\boldsymbol{\delta}_0) = \eta_t \left(\omega_0 + \sum_{i=1}^p \beta_{i0} |y_{t-i}|\right)$. Then by Assumption 1 and $E(|\eta_t|) = 1$ by

Assumption 3(i), it follows that

$$
\begin{aligned}
&E \sup_{\boldsymbol{\theta}\in\Theta}\left[\frac{|\varepsilon_t(\boldsymbol{\alpha})|}{h_t(\boldsymbol{\delta})}\right] \\
&\leq E \sup_{\boldsymbol{\theta}\in\Theta}\left[\frac{\sum_{i=1}^p |\alpha_{i0}-\alpha_i||y_{t-i}|}{\omega+\sum_{i=1}^p \beta_i\,|y_{t-i}|}\right] + E(|\eta_t|)E \sup_{\boldsymbol{\theta}\in\Theta}\left[\frac{\omega_0+\sum_{i=1}^p \beta_{i0}|y_{t-i}|}{\omega+\sum_{i=1}^p \beta_i|y_{t-i}|}\right] \\
&\leq \sum_{i=1}^p E\left[\frac{c_i|y_{t-i}|}{\underline{\beta}\,|y_{t-i}|}\right] + E\left[\frac{\overline{\omega}}{\underline{\omega}+\underline{\beta}\sum_{i=1}^p |y_{t-i}|}\right] + \sum_{i=1}^p E\left[\frac{\beta_{i0}|y_{t-i}|}{\underline{\beta}|y_{t-i}|}\right] \\
&\leq \frac{1}{\underline{\beta}}\sum_{i=1}^p c_i + \frac{\overline{\omega}}{\underline{\omega}} + \frac{1}{\underline{\beta}}\sum_{i=1}^p \beta_{i0} < \infty. \qquad \text{(S3.2)}
\end{aligned}
$$

By (S3.44) and (S3.2), we have

$$
E \sup_{\boldsymbol{\theta}\in\Theta}\left|\ell_t^E(\boldsymbol{\theta})\right| \leq E \sup_{\boldsymbol{\theta}\in\Theta}|\ln h_t(\boldsymbol{\delta})| + E \sup_{\boldsymbol{\theta}\in\Theta}\left[\frac{|\varepsilon_t(\boldsymbol{\alpha})|}{h_t(\boldsymbol{\delta})}\right] < \infty.
$$

Thus, (i) holds.

We next show (ii). Note that $\varepsilon_t(\boldsymbol{\alpha}) = \varepsilon_t(\boldsymbol{\alpha}_0) - (\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)'\boldsymbol{Y}_{1t}h_t(\boldsymbol{\delta}_0)$ and $\varepsilon_t(\boldsymbol{\alpha}_0) = \eta_t h_t(\boldsymbol{\delta}_0)$, where $\boldsymbol{Y}_{1t} = h_t^{-1}(\boldsymbol{\delta}_0)(y_{t-1},\ldots,y_{t-p})'$. Denote $\mathcal{F}_t$ as the σ-field generated by $\{y_s, s \leq t\}$. Recall that η_t has zero median and $E(|\eta_t|) = 1$, and thus $\min_a E|\eta_t - a| = E|\eta_t - \text{median}(\eta_t)| = E|\eta_t| = 1$. Then by the law of iterated expectations, we have

$$
\begin{aligned}
E\left[\ell_t^E(\boldsymbol{\theta})\right] =& E\left[\ln h_t(\boldsymbol{\delta}) + \frac{|\varepsilon_t(\boldsymbol{\alpha}_0) - (\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)'\boldsymbol{Y}_{1t}h_t(\boldsymbol{\delta}_0)|}{h_t(\boldsymbol{\delta})}\right] \\
=& E\left[\ln h_t(\boldsymbol{\delta}) + \frac{h_t(\boldsymbol{\delta}_0)}{h_t(\boldsymbol{\delta})}E\left\{|\eta_t - (\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)'\boldsymbol{Y}_{1t}|\,\middle|\mathcal{F}_{t-1}\right\}\right] \\
\geq& E\left[\ln h_t(\boldsymbol{\delta}) + \frac{h_t(\boldsymbol{\delta}_0)}{h_t(\boldsymbol{\delta})}E\left(|\eta_t|\middle|\mathcal{F}_{t-1}\right)\right] \\
=& E\left[\ln h_t(\boldsymbol{\delta}) + \frac{h_t(\boldsymbol{\delta}_0)}{h_t(\boldsymbol{\delta})}\right],
\end{aligned}
$$

where the minimum of the inequality is attained if and only if $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ almost surely as η_t has zero median. Moreover, the function $f(x) = \ln x + a/x$, $a \geq 0$, reaches its minimum at $x = a$. Therefore, $E[\ell_t^E(\boldsymbol{\theta})]$ reaches its minimum if and only if $h_t(\boldsymbol{\delta}) = h_t(\boldsymbol{\delta}_0)$ almost surely, and hence $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. As a result, $E[\ell_t^E(\boldsymbol{\theta})]$ is uniformly minimized at $\boldsymbol{\theta}_0$, that is (ii) holds.

Finally, we show (iii). Let $\boldsymbol{\theta}^* = (\boldsymbol{\alpha}^{*\prime}, \boldsymbol{\delta}^{*\prime})' \in \Theta$. For any $\boldsymbol{\theta} \in B_\eta(\boldsymbol{\theta}^*)$, we can see that

$$\ell_t^E(\boldsymbol{\theta}) - \ell_t^E(\boldsymbol{\theta}^*) = \ln h_t(\boldsymbol{\delta}) - \ln h_t(\boldsymbol{\delta}^*) + \left[\frac{|\varepsilon_t(\boldsymbol{\alpha})|}{h_t(\boldsymbol{\delta})} - \frac{|\varepsilon_t(\boldsymbol{\alpha}^*)|}{h_t(\boldsymbol{\delta})}\right] + \left[\frac{|\varepsilon_t(\boldsymbol{\alpha}^*)|}{h_t(\boldsymbol{\delta})} - \frac{|\varepsilon_t(\boldsymbol{\alpha}^*)|}{h_t(\boldsymbol{\delta}^*)}\right].$$

By Taylor's expansion, we can see that

$$\ln h_t(\boldsymbol{\delta}) - \ln h_t(\boldsymbol{\delta}^*) = \frac{(\boldsymbol{\delta} - \boldsymbol{\delta}^*)'}{h_t(\bar{\boldsymbol{\delta}})}(1, |y_{t-1}|, ..., |y_{t-p}|)',$$

where $\bar{\boldsymbol{\delta}}$ lies between $\boldsymbol{\delta}$ and $\boldsymbol{\delta}^*$. Then, by Assumption 1, we have

$$\begin{aligned}
&E\left[\sup_{\boldsymbol{\theta}\in B_\eta(\boldsymbol{\theta}^*)} |\ln h_t(\boldsymbol{\delta}) - \ln h_t(\boldsymbol{\delta}^*)|\right] \\
\leq &\eta E\left[\frac{1 + \sum_{i=1}^p |y_{t-i}|}{\underline{\omega} + \underline{\beta}\sum_{i=1}^p |y_{t-i}|}\right] \leq \eta\left(\frac{1}{\underline{\omega}} + \frac{p}{\underline{\beta}}\right) \to 0 \text{ as } \eta \to 0.
\end{aligned}$$

Similarly, by Taylor's expansion and (S3.2), it can be verified that

$$\begin{aligned}
&E\left[\sup_{\boldsymbol{\theta}\in B_\eta(\boldsymbol{\theta}^*)} |\varepsilon_t(\boldsymbol{\alpha}^*)| \left|\frac{1}{h_t(\boldsymbol{\delta})} - \frac{1}{h_t(\boldsymbol{\delta}^*)}\right|\right] \\
=&E\left[\sup_{\boldsymbol{\theta}\in B_\eta(\boldsymbol{\theta}^*)} |\varepsilon_t(\boldsymbol{\alpha}^*)| \frac{|(\boldsymbol{\delta} - \boldsymbol{\delta}^*)'(1, |y_{t-1}|, ..., |y_{t-p}|)'|}{h_t^2(\bar{\boldsymbol{\delta}})}\right] \\
\leq &\eta E\left[\sup_{\boldsymbol{\theta}\in B_\eta(\boldsymbol{\theta}^*)} \frac{|\varepsilon_t(\boldsymbol{\alpha}^*)|}{h_t(\bar{\boldsymbol{\delta}})}\left(\frac{1 + \sum_{i=1}^p |y_{t-i}|}{\underline{\omega} + \underline{\beta}\sum_{i=1}^p |y_{t-i}|}\right)\right] \\
\leq &\eta\left(\frac{1}{\underline{\beta}}\sum_{i=1}^p c_i + \frac{\overline{\omega}}{\underline{\omega}} + \frac{1}{\underline{\beta}}\sum_{i=1}^p \beta_{i0}\right)\left(\frac{1}{\underline{\omega}} + \frac{p}{\underline{\beta}}\right) \to 0 \text{ as } \eta \to 0,
\end{aligned}$$

and

$$E\left[\sup_{\boldsymbol{\theta}\in B_\eta(\boldsymbol{\theta}^*)} \frac{||\varepsilon_t(\boldsymbol{\alpha})| - |\varepsilon_t(\boldsymbol{\alpha}^*)||}{h_t(\boldsymbol{\delta})}\right] \leq \eta E\left[\frac{\sum_{i=1}^p |y_{t-i}|}{\underline{\omega} + \underline{\beta}\sum_{i=1}^p |y_{t-i}|}\right] \leq \frac{\eta p}{\underline{\beta}} \to 0 \text{ as } \eta \to 0.$$

Therefore, we have

$$E\left[\sup_{\boldsymbol{\theta}\in B_\eta(\boldsymbol{\theta}^*)} |\ell_t^E(\boldsymbol{\theta}) - \ell_t^E(\boldsymbol{\theta}^*)|\right] \to 0 \text{ as } \eta \to 0.$$

Hence, (iii) is verified. The proof of this lemma is accomplished. □

Proof of Theorem 1. To establish the strong consistency of $\widehat{\boldsymbol{\theta}}_n$, we follow the method in Huber (1967). Let V be any open neighborhood of $\boldsymbol{\theta}_0 \in \Theta$. By Lemma 1 (iii), for any $\boldsymbol{\theta}^* \in V^c =$

Θ/V and $\varepsilon > 0$, there exists an $\eta_0 > 0$ such that

$$E\left[\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}^*)} \ell_t^E(\boldsymbol{\theta})\right] \geq E\left[\ell_t^E(\boldsymbol{\theta}^*)\right] - \varepsilon. \tag{S3.3}$$

From Lemma 1 (i), by the Birkhoff ergodic theorem (Fristedt and Gray, 2013), if n is large enough, then we have

$$\frac{1}{n-p}\sum_{t=p+1}^{n} \inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}^*)} \ell_t^E(\boldsymbol{\theta}) \geq E\left[\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}^*)} \ell_t^E(\boldsymbol{\theta})\right] - \varepsilon \quad \text{a.s..} \tag{S3.4}$$

Since V^c is compact, we can choose $\{B_{\eta_0}(\boldsymbol{\theta}_i) : \boldsymbol{\theta}_i \in V^c, i = 1, 2, ..., k\}$ to be a finite covering of V^c. For each $\boldsymbol{\theta}_i \in V^c$, from Lemma 1 (ii), there exists an $\varepsilon_0 > 0$ such that

$$E\left[\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}_i)} \ell_t^E(\boldsymbol{\theta})\right] \geq E\left[\ell_t^E(\boldsymbol{\theta}_0)\right] + 3\varepsilon_0. \tag{S3.5}$$

Thus by (S3.3)–(S3.5), taking $\varepsilon = \varepsilon_0$, if n is large enough, we have

$$\begin{aligned}
\inf_{\boldsymbol{\theta}\in V^c} L_n^E(\boldsymbol{\theta}) &= \min_{1\leq i\leq k} \inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}_i)} L_n^E(\boldsymbol{\theta}) \\
&\geq \min_{1\leq i\leq k} \frac{1}{n-p}\sum_{t=p+1}^{n} \inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}_i)} \ell_t^E(\boldsymbol{\theta}) \\
&\geq \min_{1\leq i\leq k} E\left[\inf_{\boldsymbol{\theta}\in B_{\eta_0}(\boldsymbol{\theta}_i)} \ell_t^E(\boldsymbol{\theta})\right] - \varepsilon \\
&\geq E\left[\ell_t^E(\boldsymbol{\theta}_0)\right] + 2\varepsilon_0.
\end{aligned} \tag{S3.6}$$

On the other hand, by the Birkhoff ergodic theorem (Fristedt and Gray, 2013), it follows that

$$\inf_{\boldsymbol{\theta}\in V} L_n^E(\boldsymbol{\theta}) \leq L_n^E(\boldsymbol{\theta}_0) = \frac{1}{n-p}\sum_{t=p+1}^{n} \ell_t^E(\boldsymbol{\theta}_0) \leq E\left[\ell_t^E(\boldsymbol{\theta}_0)\right] + \varepsilon_0 \quad \text{a.s..} \tag{S3.7}$$

Hence, combining (S3.6) and (S3.7), it holds that

$$\inf_{\boldsymbol{\theta}\in V^c} L_n^E(\boldsymbol{\theta}) \geq E\left[\ell_t^E(\boldsymbol{\theta}_0)\right] + 2\varepsilon_0 > E\left[\ell_t^E(\boldsymbol{\theta}_0)\right] + \varepsilon_0 \geq \inf_{\boldsymbol{\theta}\in V} L_n^E(\boldsymbol{\theta}) \quad \text{a.s..}$$

Since $\widehat{\boldsymbol{\theta}}_n = \arg\min_{\Theta} L_n^E(\boldsymbol{\theta})$, if n is large enough, then $\widehat{\boldsymbol{\theta}}_n \in V$ almost surely for $\forall V$. By the arbitrariness of V, it follows that $\widehat{\boldsymbol{\theta}}_n \to \boldsymbol{\theta}_0$ almost surely. The proof is accomplished. □

S3.2 Proof of Theorem 2

To show Theorem 2, we introduce Lemmas 2–3 below, where Lemma 2 verifies the stochastic differentiability condition defined by Pollard (1985), and Lemma 3 provides a representation to obtain the $\sqrt{n}$-consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}_n$.

Lemma 2. *If Assumptions 1–3 hold, then for any sequence of random variables $\boldsymbol{u}_n$ such that $\boldsymbol{u}_n = o_p(1)$, it follows that*

$$\Pi_{1n}(\boldsymbol{u}_n) = o_p(\sqrt{n}\|\boldsymbol{u}_n\| + n\|\boldsymbol{u}_n\|^2),$$

where $\Pi_{1n}(\boldsymbol{u}) = 2\sum_{t=p+1}^{n}\int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}}\{X_t(s) - E[X_t(s)|\mathcal{F}_{t-1}]\}\,ds$ *with* $\boldsymbol{K}_{1t} = (\boldsymbol{Y}'_{1t}, \boldsymbol{0}'_{p+1})'$ *and* $X_t(s) = I(\eta_t \le s) - I(\eta_t \le 0)$.

Proof. Note that $\partial\varepsilon_t(\boldsymbol{\alpha}_0)/\partial\boldsymbol{\theta} = (-\boldsymbol{Y}'_{1t}h_t, \boldsymbol{0}'_{p+1})' = -h_t\boldsymbol{K}_{1t}$. Then it follows that

$$\int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}} X_t(s)ds = \boldsymbol{u}'\boldsymbol{K}_{1t}\int_0^1 X_t(\boldsymbol{u}'\boldsymbol{K}_{1t}s)ds = -\frac{\boldsymbol{u}'}{h_t}\frac{\partial\varepsilon_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\theta}}M_t(\boldsymbol{u}),$$

where $M_t(\boldsymbol{u}) = \int_0^1 X_t(\boldsymbol{u}'\boldsymbol{K}_{1t}s)ds$. It can be verified that

$$|\Pi_{1n}(\boldsymbol{u})| \le 2\|\boldsymbol{u}\|\sum_{j=1}^{2p+1}\left|\frac{1}{h_t}\frac{\partial\varepsilon_t(\boldsymbol{\alpha}_0)}{\partial\theta_j}\sum_{t=p+1}^{n}\{M_t(\boldsymbol{u}) - E[M_t(\boldsymbol{u})|\mathcal{F}_{t-1}]\}\right|,$$

where θ_j is the jth element of $\boldsymbol{\theta}$ for $1 \le j \le 2p+1$. Let $m_t = h_t^{-1}\partial\varepsilon_t(\boldsymbol{\alpha}_0)/\partial\theta_j$ and $f_t(\boldsymbol{u}) = m_tM_t(\boldsymbol{u})$. Define

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

Then, to prove Lemma 2, it suffices to show that, for any $\eta > 0$,

$$\sup_{\|\boldsymbol{u}\|\le\eta}\frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} = o_p(1). \tag{S3.8}$$

Note that $m_t = \max\{m_t, 0\} - \max\{-m_t, 0\}$. We first prove the case when $m_t \ge 0$.

We adopt the method in Lemma 4 of Pollard (1985) to verify (S3.8). Let $\mathfrak{F} = \{f_t(\boldsymbol{u}) : \|\boldsymbol{u}\| \leq \eta\}$ be a collection of functions indexed by $\boldsymbol{u}$. We first verify that $\mathfrak{F}$ statisfies the bracketing condition on page 304 of Pollard (1985). Denote $B_r(\boldsymbol{\zeta})$ as an open neighborhood of $\boldsymbol{\zeta}$ with radius $r > 0$, and define a constant C_1 to be selected later. For any fixed $\varepsilon > 0$ and $0 < \delta \leq \eta$, there exists a sequence of small cubes $\{B_{\varepsilon\delta/C_1}(\boldsymbol{u}_i)\}_{i=1}^{K_\varepsilon}$ to cover $B_\delta(\mathbf{0})$, where K_ε is an integer less than $C_0\varepsilon^{-(2p+1)}$ and C_0 is not depending on ε and δ. Denote $V_i(\delta) = B_{\epsilon\delta/C_0}(\boldsymbol{u}_i) \bigcap B_\delta(\mathbf{0})$, and let $U_1(\delta) = V_1(\delta)$ and $U_i(\delta) = V_i(\delta) - \bigcup_{j=1}^{i-1} V_j(\delta)$ for $i \geq 2$. Note that $\{U_i(\delta)\}_{i=1}^{K_\varepsilon}$ is a partition of $B_\delta(\mathbf{0})$. For each $\boldsymbol{u}_i \in U_i(\delta)$ with $1 \leq i \leq K_\varepsilon$, we define the bracketing functions as follows

$$f_t^{\pm}(\boldsymbol{u}) = m_t \int_0^1 X_t \left(\boldsymbol{u}' \boldsymbol{K}_{1t} s \pm \frac{\varepsilon\delta}{C_1 h_t} \left\| \frac{\partial \varepsilon_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\theta}} \right\| \right) ds.$$

Since the indicator function is nondecreasing and $m_t \geq 0$, for any $\boldsymbol{u} \in U_i(\delta)$, we have

$$f_t^-(\boldsymbol{u}_i) \leq f_t(\boldsymbol{u}) \leq f_t^+(\boldsymbol{u}_i).$$

Moreover, by Taylor's expansion and the fact that $\partial\varepsilon_t(\boldsymbol{\alpha}_0)/\partial\boldsymbol{\theta} = (-\boldsymbol{Y}_{1t}' h_t, \mathbf{0}_{p+1}')'$, it holds that

$$E\left[f_t^+(\boldsymbol{u}_i) - f_t^-(\boldsymbol{u}_i)|\mathcal{F}_{t-1}\right] \leq \frac{2\varepsilon\delta}{C_1 h_t^2} \left\| \frac{\partial \varepsilon_t(\boldsymbol{\alpha}_0)}{\partial \boldsymbol{\theta}} \right\|^2 \sup_x f(x) = \frac{\varepsilon\delta\Delta_t}{C_1}, \tag{S3.9}$$

where $\Delta_t = 2\|\boldsymbol{Y}_{1t}\|^2 \sup_x f(x)$. Since $\sup_x f(x) < \infty$ by Assumption 3(ii), we choose $C_1 = E(\Delta_t)$. Then by iterated-expectation, we have

$$E[f_t^+(\boldsymbol{u}_i) - f_t^-(\boldsymbol{u}_i)] = E\left\{E\left[f_t^+(\boldsymbol{u}_i) - f_t^-(\boldsymbol{u}_i)|\mathcal{F}_{t-1}\right]\right\} \leq \varepsilon\delta.$$

Thus, the family $\mathfrak{F}$ satisfies the bracketing condition.

Put $\delta_k = 2^{-k}\eta$. Define $B(k) = B_{\delta_k}(\mathbf{0})$, and $A(k)$ to be the annulus $B(k) \setminus B(k+1)$. Fix $\varepsilon > 0$, for each $1 \leq i \leq K_\varepsilon$, by the bracketing condition, there exists a partition $\{U_i(\delta_k)\}_{i=1}^{K_\varepsilon}$

of $B(k)$. For $\boldsymbol{u} \in U_i(\delta_k)$, by (S3.9) with $\delta = \delta_k$, we have the upper tail

$$\begin{aligned} D_n(\boldsymbol{u}) \leq & \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \left\{ f_t^+(\boldsymbol{u}_i) - E\left[f_t^-(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \right\} \\ = & D_n^+(\boldsymbol{u}_i) + \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} E\left[f_t^+(\boldsymbol{u}_i) - f_t^-(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \\ \leq & D_n^+(\boldsymbol{u}_i) + \sqrt{n}\varepsilon\delta_k \left(\frac{1}{nC_1} \sum_{t=p+1}^{n} \Delta_t \right), \end{aligned}$$

and the lower tail

$$\begin{aligned} D_n(\boldsymbol{u}) \geq & \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \left\{ f_t^-(\boldsymbol{u}_i) - E\left[f_t^+(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \right\} \\ = & D_n^-(\boldsymbol{u}_i) - \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} E\left[f_t^+(\boldsymbol{u}_i) - f_t^-(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \\ \geq & D_n^-(\boldsymbol{u}_i) - \sqrt{n}\varepsilon\delta_k \left(\frac{1}{nC_1} \sum_{t=p+1}^{n} \Delta_t \right), \end{aligned}$$

where

$$\begin{aligned} D_n^+(\boldsymbol{u}_i) &= \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \left\{ f_t^+(\boldsymbol{u}_i) - E\left[f_t^+(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \right\} \quad \text{and} \\ D_n^-(\boldsymbol{u}_i) &= \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \left\{ f_t^-(\boldsymbol{u}_i) - E\left[f_t^-(\boldsymbol{u}_i) | \mathcal{F}_{t-1} \right] \right\}. \end{aligned}$$

Denote the event

$$E_n = \left\{ \omega : \frac{1}{nC_1} \sum_{t=p+1}^{n} \Delta_t(\omega) < 2 \right\}.$$

On E_n with $\boldsymbol{u} \in U_i(\delta_k)$, it follows that

$$D_n^-(\boldsymbol{u}_i) - 2\sqrt{n}\varepsilon\delta_k \leq D_n(\boldsymbol{u}) \leq D_n^+(\boldsymbol{u}_i) + 2\sqrt{n}\varepsilon\delta_k. \tag{S3.10}$$

For $\boldsymbol{u} \in A(k)$, i.e. $\delta_{k+1} \leq \|\boldsymbol{u}\| \leq \delta_k$, we have $1 + \sqrt{n}\|\boldsymbol{u}\| > \sqrt{n}\delta_{k+1} = \sqrt{n}\delta_k/2$. Thus, by the upper tail in (S3.10) and Chebyshev's inequality, it holds that

$$P\left(\sup_{\boldsymbol{u} \in A(k)} \frac{D_n(\boldsymbol{u})}{1 + \sqrt{n}\|\boldsymbol{u}\|} > 6\varepsilon, E_n \right) \leq P\left(\max_{1 \leq i \leq K_\varepsilon} \sup_{\boldsymbol{u} \in U_i(k) \cap A(k)} D_n(\boldsymbol{u}) > 3\sqrt{n}\varepsilon\delta_k, E_n \right)$$

$$\leq K_\varepsilon \max_{1\leq i\leq K_\varepsilon} P(D_n^+(\boldsymbol{u}_i) > \sqrt{n}\varepsilon\delta_k)$$
$$\leq K_\varepsilon \max_{1\leq i\leq K_\varepsilon} \frac{E[(D_n^+(\boldsymbol{u}_i))^2]}{n\varepsilon^2\delta_k^2}, \tag{S3.11}$$

and by the lower tail in (S3.10) and Chebyshev's inequality, we have

$$P\left(\inf_{\boldsymbol{u}\in A(k)} \frac{D_n(\boldsymbol{u})}{1+\sqrt{n}\|\boldsymbol{u}\|} < -6\varepsilon, E_n\right) \leq P\left(\min_{1\leq i\leq K_\varepsilon} \inf_{\boldsymbol{u}\in U_i(k)\cap A(k)} D_n(\boldsymbol{u}) < -3\sqrt{n}\varepsilon\delta_k, E_n\right)$$
$$\leq K_\varepsilon \max_{1\leq i\leq K_\varepsilon} P(D_n^-(\boldsymbol{u}_i) < -\sqrt{n}\varepsilon\delta_k)$$
$$\leq K_\varepsilon \max_{1\leq i\leq K_\varepsilon} \frac{E[(D_n^-(\boldsymbol{u}_i))^2]}{n\varepsilon^2\delta_k^2}. \tag{S3.12}$$

Recall that $\boldsymbol{K}_{1t} = (\boldsymbol{Y}_{1t}^{'}, \boldsymbol{0}_{p+1}')'$ with $\boldsymbol{Y}_{1t} = h_t^{-1}(\boldsymbol{\delta}_0)(y_{t-1}, \ldots, y_{t-p})'$ and $h_t(\boldsymbol{\delta}) = \omega + \sum_{i=1}^p \beta_i |y_{t-i}|$.

By Assumption 1, it holds that

$$\|\boldsymbol{K}_{1t}\| = \left[\sum_{i=1}^p \frac{y_{t-i}^2}{(\omega_0 + \sum_{i=1}^p \beta_{i0}|y_{t-i}|)^2}\right]^{\frac{1}{2}} \leq \left[\sum_{i=1}^p \frac{y_{t-i}^2}{(\underline{\beta}|y_{t-i}|)^2}\right]^{\frac{1}{2}} \leq \left(\frac{p}{\underline{\beta}^2}\right)^{\frac{1}{2}} \leq \frac{p}{\underline{\beta}}. \tag{S3.13}$$

Then it follows that

$$|\boldsymbol{u}^{'}\boldsymbol{K}_{1t}| \leq \|\boldsymbol{u}\|\|\boldsymbol{K}_{1t}\| \leq \frac{p\delta_k}{\underline{\beta}} \quad \text{and} \quad |m_t| \leq \frac{1}{h_t}\left\|\frac{\partial\varepsilon_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\theta}}\right\| = \|\boldsymbol{K}_{1t}\| \leq \frac{p}{\underline{\beta}}.$$

By Taylor's expansion and iterated-expectation, we have

$$E[f_t^+(\boldsymbol{u}_i)^2] = E\left\{E[f_t^+(\boldsymbol{u}_i)^2|\mathcal{F}_{t-1}]\right\}$$
$$\leq E\left\{m_t^2 \int_0^1 E\left[\left|X_t\left(\boldsymbol{u}^{'}\boldsymbol{K}_{1t}s + \frac{\varepsilon\delta_k}{C_1 h_t}\left\|\frac{\partial\varepsilon_t(\boldsymbol{\alpha}_0)}{\partial\boldsymbol{\theta}}\right\|\right)\right|\mathcal{F}_{t-1}\right]ds\right\}$$
$$\leq \frac{p^2}{\underline{\beta}^2} E\left(\sup_{|x|\leq C\delta_k} |F(x) - F(0)|\right) \leq C\delta_k \sup_x f(x).$$

This together with the fact that $f_t^+(\boldsymbol{u}_i) - E\left[f_t^+(\boldsymbol{u}_i)|\mathcal{F}_{t-1}\right]$ is a martingale difference sequence, and $\sup_x f(x) < \infty$ by Assumption 3(ii), implies that

$$E\left[(D_n^+(\boldsymbol{u}_i))^2\right] = \frac{1}{n}\sum_{t=p+1}^n E\left\{f_t^+(\boldsymbol{u}_i) - E\left[f_t^+(\boldsymbol{u}_i)|\mathcal{F}_{t-1}\right]\right\}^2$$
$$\leq \frac{1}{n}\sum_{t=p+1}^n E\left[(f_t^+(\boldsymbol{u}_i))^2\right] \leq C\delta_k \sup_x f(x) := \pi_n(\delta_k). \tag{S3.14}$$

Similarly, we have

$$E\left[(D_n^-(\boldsymbol{u}_i))^2\right] \leq \pi_n(\delta_k). \tag{S3.15}$$

Combining (S3.11) and (S3.14), we have

$$P\left(\sup_{\boldsymbol{u}\in A(k)} \frac{D_n(\boldsymbol{u})}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\varepsilon, E_n\right) \leq \frac{K_\varepsilon \pi_n(\delta_k)}{n\varepsilon^2\delta_k^2}.$$

Combining (S3.12) and (S3.15), we have

$$P\left(\inf_{\boldsymbol{u}\in A(k)} \frac{D_n(\boldsymbol{u})}{1+\sqrt{n}\|\boldsymbol{u}\|} < -6\varepsilon, E_n\right) \leq \frac{K_\varepsilon \pi_n(\delta_k)}{n\varepsilon^2\delta_k^2}.$$

Thus, we can show that

$$P\left(\sup_{\boldsymbol{u}\in A(k)} \frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\varepsilon, E_n\right) \leq \frac{2K_\varepsilon \pi_n(\delta_k)}{n\varepsilon^2\delta_k^2}. \tag{S3.16}$$

Since $\pi_n(\delta_k) \to 0$ as $k \to \infty$, we can choose k_ε so that $2K_\varepsilon\pi_n(\delta_k)/(\varepsilon^2\eta^2) < \varepsilon$ for $k \geq k_\varepsilon$. Let k_n be an integer so that $n^{-1/2} \leq 2^{-k_n} \leq 2n^{-1/2}$. Split $\{\boldsymbol{u} : \|\boldsymbol{u}\| \leq \eta\}$ into two sets $B := B(k_n+1)$ and $B^c := B(0) - B(k_n+1) = \cup_{k=0}^{k_n} A(k)$. Then by (S3.16), we have

$$\begin{aligned}
P\left(\sup_{\boldsymbol{u}\in B^c} \frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\varepsilon\right) \leq & \sum_{k=0}^{k_n} P\left(\sup_{\boldsymbol{u}\in A(k)} \frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 6\varepsilon, E_n\right) + P(E_n^c) \\
\leq & \frac{1}{n}\sum_{k=0}^{k_\varepsilon-1} \frac{2K_\varepsilon\pi_n(\delta_k)}{\varepsilon^2\eta^2} 2^{2k} + \frac{\varepsilon}{n}\sum_{k=k_\varepsilon}^{k_n} 2^{2k} + P(E_n^c) \\
\leq & O\left(\frac{1}{n}\right) + 4\varepsilon + P(E_n^c).
\end{aligned} \tag{S3.17}$$

Since $1+\sqrt{n}\|\boldsymbol{u}\| > 1$ and $\sqrt{n}\delta_{k_n+1} < 1$, similar to the proof of (S3.11) and (S3.14), we have

$$P\left(\sup_{\boldsymbol{u}\in B} \frac{D_n(\boldsymbol{u})}{1+\sqrt{n}\|\boldsymbol{u}\|} > 3\varepsilon, E_n\right) \leq P\left(\max_{1\leq i\leq K_\varepsilon} D_n^+(\boldsymbol{u}_i) > \varepsilon, E_n\right) \leq \frac{K_\varepsilon\pi_n(\delta_{k_n+1})}{\varepsilon^2}.$$

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

$$P\left(\sup_{\boldsymbol{u}\in B} \frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 3\varepsilon\right) \leq P\left(\sup_{\boldsymbol{u}\in B} \frac{|D_n(\boldsymbol{u})|}{1+\sqrt{n}\|\boldsymbol{u}\|} > 3\varepsilon, E_n\right) + P(E_n^c)$$

$$\leq \frac{2K_\varepsilon \pi_n(\delta_{k_n+1})}{\varepsilon^2} + P(E_n^c). \tag{S3.18}$$

Note that $\pi_n(\delta_{k_n+1}) \to 0$ as $n \to \infty$. Furthermore, $P(E_n) \to 1$ by the ergodic theorem. Hence, $P(E_n^c) \to 0$ as $n \to \infty$. Finally, (S3.8) follows by (S3.17) and (S3.18). We can verify (S3.8) similarly for the case of $m_t < 0$. This completes the proof of this lemma. □

Lemma 3. *Suppose that Assumptions 1, 2 and 3 hold, then for any sequence of random variables* $\boldsymbol{u}_n$ *such that* $\boldsymbol{u}_n = o_p(1)$*, it follows that*

$$(n-p)[L_n(\boldsymbol{\theta}_0 + \boldsymbol{u}_n) - L_n(\boldsymbol{\theta}_0)] = \sqrt{n}\boldsymbol{u}_n^{'}\boldsymbol{T}_n + (\sqrt{n}\boldsymbol{u}_n)^{'}\Sigma(\sqrt{n}\boldsymbol{u}_n) + o_p(\sqrt{n}\|\boldsymbol{u}_n\| + n\|\boldsymbol{u}_n\|^2),$$

where $\boldsymbol{T}_n = \boldsymbol{T}_{1n} + \boldsymbol{T}_{2n}$ *with* $\boldsymbol{T}_{1n} = n^{-1/2}\sum_{t=p+1}^n \boldsymbol{K}_{1t}[I(\eta_t < 0) - I(\eta_t > 0)]$ *and* $\boldsymbol{T}_{2n} = n^{-1/2}\sum_{t=p+1}^n \boldsymbol{K}_{2t}(1 - |\eta_t|)$*, and* $\Sigma = \mathrm{diag}\,\{f(0)E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}'), E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}')/2\}$*. Moreover,*

$$\boldsymbol{T}_n \to_{\mathcal{L}} N(0, \Omega) \quad as \quad n \to \infty,$$

where

$$\Omega = \begin{pmatrix} E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}') & \kappa_1 E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{2t}') \\ \kappa_1 E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{1t}') & \kappa_2 E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}') \end{pmatrix} \quad with \quad \kappa_1 = E(\eta_t) \;\; and \;\; \kappa_2 = E(\eta_t^2) - 1.$$

Proof. We first re-parameterize the objective function as

$$H_n(\boldsymbol{u}) = (n-p)\left[L_n^E(\boldsymbol{\theta}_0 + \boldsymbol{u}) - L_n^E(\boldsymbol{\theta}_0)\right],$$

where $\boldsymbol{u} \in \Lambda \equiv \{\boldsymbol{u} = (\boldsymbol{u}_1', \boldsymbol{u}_2')' : \boldsymbol{u} + \boldsymbol{\theta}_0 \in \Theta\}$. Let $\widehat{\boldsymbol{u}}_n = \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0$. Then, $\widehat{\boldsymbol{u}}_n$ is the minimizer of $H_n(\boldsymbol{u})$ in Λ. Furthermore, we have

$$H_n(\boldsymbol{u}) = \sum_{t=p+1}^n A_t(\boldsymbol{u}) + \sum_{t=p+1}^n B_t(\boldsymbol{u}) + \sum_{t=p+1}^n C_t(\boldsymbol{u}),$$

where

$$A_t(\boldsymbol{u}) = \frac{1}{h_t(\boldsymbol{\delta}_0)}\left[|\varepsilon_t(\boldsymbol{\alpha}_0 + \boldsymbol{u}_1)| - |\varepsilon_t(\boldsymbol{\alpha}_0)|\right],$$

$$B_t(\boldsymbol{u}) = \ln h_t(\boldsymbol{\delta}_0 + \boldsymbol{u}_2) - \ln h_t(\boldsymbol{\delta}_0) + \frac{|\varepsilon_t(\boldsymbol{\alpha}_0)|}{h_t(\boldsymbol{\delta}_0 + \boldsymbol{u}_2)} - \frac{|\varepsilon_t(\boldsymbol{\alpha}_0)|}{h_t(\boldsymbol{\delta}_0)},$$

$$C_t(\boldsymbol{u}) = \left[\frac{1}{h_t(\boldsymbol{\delta}_0 + \boldsymbol{u}_2)} - \frac{1}{h_t(\boldsymbol{\delta}_0)}\right] \left[|\varepsilon_t(\boldsymbol{\alpha}_0 + \boldsymbol{u}_1)| - |\varepsilon_t(\boldsymbol{\alpha}_0)|\right].$$

Recall that $\boldsymbol{K}_{1t} = (\boldsymbol{Y}'_{1t}, \mathbf{0}'_{p+1})'$. By Taylor's expansion, we can show that

$$\frac{\varepsilon_t(\boldsymbol{\alpha}_0 + \boldsymbol{u}_1)}{h_t(\boldsymbol{\delta}_0)} - \frac{\varepsilon_t(\boldsymbol{\alpha}_0)}{h_t(\boldsymbol{\delta}_0)} = -\frac{\boldsymbol{u}_1'(y_{t-1}, ..., y_{t-p})'}{h_t(\boldsymbol{\delta}_0)} = -\boldsymbol{u}'\boldsymbol{K}_{1t}. \tag{S3.19}$$

Let $I(\cdot)$ be the indicator function. For $x \neq 0$, using the identity

$$|x - y| - |x| = -y[I(x > 0) - I(x < 0)] + 2\int_0^y [I(x \leq s) - I(x \leq 0)]ds, \tag{S3.20}$$

together with (S3.19), we have

$$\begin{aligned} |\varepsilon_t(\boldsymbol{\alpha}_0 + \boldsymbol{u}_1)| - |\varepsilon_t(\boldsymbol{\alpha}_0)| = &- \boldsymbol{u}'\boldsymbol{K}_{1t} h_t[I(\eta_t > 0) - I(\eta_t < 0)] \\ &+ 2h_t \int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}} X_t(s)\, ds. \end{aligned} \tag{S3.21}$$

Therefore, it follows that

$$A_t(\boldsymbol{u}) = \boldsymbol{u}'\boldsymbol{K}_{1t}[I(\eta_t < 0) - I(\eta_t > 0)] + 2\int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}} X_t(s)ds,$$

where $X_t(s) = I(\eta_t \leq s) - I(\eta_t \leq 0)$. Then we have

$$\sum_{t=p+1}^{n} A_t(\boldsymbol{u}) = \sqrt{n}\boldsymbol{u}'\boldsymbol{T}_{1n} + \Pi_{1n}(\boldsymbol{u}) + \Pi_{2n}(\boldsymbol{u}), \tag{S3.22}$$

where

$$\boldsymbol{T}_{1n} = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \boldsymbol{K}_{1t}[I(\eta_t < 0) - I(\eta_t > 0)],$$

$$\Pi_{1n}(\boldsymbol{u}) = 2\sum_{t=p+1}^{n} \int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}} \{X_t(s) - E\left[X_t(s)|\mathcal{F}_{t-1}\right]\} ds,$$

$$\Pi_{2n}(\boldsymbol{u}) = 2\sum_{t=p+1}^{n} \int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}} E\left[X_t(s)|\mathcal{F}_{t-1}\right] ds.$$

For $\Pi_{1n}(\boldsymbol{u}_n)$ with $\boldsymbol{u}_n = o_p(1)$, by Lemma 2, it holds that

$$\Pi_{1n}(\boldsymbol{u}_n) = o_p(\sqrt{n}\|\boldsymbol{u}_n\| + n\|\boldsymbol{u}_n\|^2). \tag{S3.23}$$

For $\Pi_{2n}(\boldsymbol{u})$, by iterated-expectation and Taylor's expansion, we have

$$\begin{aligned}
\Pi_{2n}(\boldsymbol{u}) &= 2\sum_{t=p+1}^{n}\int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}} E\left[I(\eta_t \le s) - I(\eta_t \le 0)|\mathcal{F}_{t-1}\right] ds \\
&= 2\sum_{t=p+1}^{n}\int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}} [F(s) - F(0)]ds \\
&= 2\sum_{t=p+1}^{n}\int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}} sf(0)ds + 2\sum_{t=p+1}^{n}\int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}} s[f(s^*) - f(0)]ds \\
&= (\sqrt{n}\boldsymbol{u})'[K_{1n} + K_{2n}(\boldsymbol{u})](\sqrt{n}\boldsymbol{u}),
\end{aligned} \tag{S3.24}$$

where s^* lies between 0 and s,

$$K_{1n} = \frac{f(0)}{n}\sum_{t=p+1}^{n}\boldsymbol{K}_{1t}\boldsymbol{K}_{1t}' \text{ and } K_{2n}(\boldsymbol{u}) = \frac{2}{n\|\boldsymbol{u}\|^2}\sum_{t=p+1}^{n}\int_0^{\boldsymbol{u}'\boldsymbol{K}_{1t}} s[f(s^*) - f(0)]ds.$$

By the ergodic theorem, it holds that

$$K_{1n} = f(0)E[\boldsymbol{K}_{1t}\boldsymbol{K}_{1t}'] + o_p(1) = \Sigma_{21} + o_p(1), \tag{S3.25}$$

where $\Sigma_{21} = \text{diag}\left\{f(0)E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}'), 0_{(p+1)\times(p+1)}\right\}$.

Since $\|\boldsymbol{K}_{1t}\| \le p/\underline{\beta}$ by (S3.13) and $\|\boldsymbol{u}\| \le \eta$, we have $|\boldsymbol{u}'\boldsymbol{K}_{1t}| \le p\eta/\underline{\beta}$. Furthermore, it follows that $\lim_{\eta\to 0}\sup_{0\le s\le p\eta/\underline{\beta}}|f(s) - f(0)| = 0$ as $\sup_x f(x) < \infty$ by Assumption 3(ii). Then for any $\eta > 0$, it holds that

$$\begin{aligned}
\sup_{\|\boldsymbol{u}\|<\eta}|K_{2n}(\boldsymbol{u})| &\le \sup_{\|\boldsymbol{u}\|<\eta}\frac{2}{n\|\boldsymbol{u}\|^2}\sum_{t=p+1}^{n}\int_0^{|\boldsymbol{u}'\boldsymbol{K}_{1t}|} s|f(s^*) - f(0)|ds \\
&\le \frac{1}{n\|\boldsymbol{u}\|^2}\sum_{t=p+1}^{n}\|\boldsymbol{u}\|^2\|\boldsymbol{K}_{1t}\|^2\left[\sup_{0\le s\le p\eta/\underline{\beta}}|f(s) - f(0)|\right] \\
&\le \frac{p^2}{n\underline{\beta}^2}\sum_{t=p+1}^{n}\left[\sup_{0\le s\le p\eta/\underline{\beta}}|f(s) - f(0)|\right]
\end{aligned}$$

$$<\frac{p^2}{\underline{\beta}^2}\left[\sup_{0\le s\le p\eta/\underline{\beta}}|f(s)-f(0)|\right]\to 0 \quad \text{as } \eta\to 0.$$

Therefore, by Markov's theorem and the stationarity of $\{y_t\}$ by Assumption 2, for $\forall\varepsilon,\delta>0$, there exists $\eta_0=\eta_0(\varepsilon)>0$ such that

$$P\left(\sup_{\|\boldsymbol{u}\|\le\eta_0}|K_{2n}(\boldsymbol{u})|>\delta\right)<\frac{\varepsilon}{2}, \tag{S3.26}$$

for all $n\ge 1$. On the other hand, since $\boldsymbol{u}_n=o_p(1)$, it follows that

$$P(\|\boldsymbol{u}_n\|>\eta_0)<\frac{\varepsilon}{2} \tag{S3.27}$$

as n is large enough. By (S3.26) and (S3.27), for $\forall\varepsilon,\delta>0$, we have

$$\begin{aligned}P\left(|K_{2n}(\boldsymbol{u}_n)|>\delta\right)=&P(|K_{2n}(\boldsymbol{u}_n)|>\delta,\|\boldsymbol{u}_n\|\le\eta_0)+P(|K_{2n}(\boldsymbol{u}_n)|>\delta,\|\boldsymbol{u}_n\|>\eta_0)\\ \le&P\left(\sup_{\|\boldsymbol{u}\|\le\eta_0}|K_{2n}(\boldsymbol{u})|>\delta\right)+P(\|\boldsymbol{u}_n\|>\eta_0)<\varepsilon\end{aligned}$$

as n is large enough. Hence, it holds that $K_{2n}(\boldsymbol{u}_n)=o_p(1)$. Furthermore, combining (S3.24) and (S3.25), we can show that

$$\Pi_{2n}(\boldsymbol{u}_n)=(\sqrt{n}\boldsymbol{u}_n)'\Sigma_{21}(\sqrt{n}\boldsymbol{u}_n)+o_p(n\|\boldsymbol{u}_n\|^2). \tag{S3.28}$$

As a result, combining (S3.22), (S3.23) and (S3.28), we have

$$\sum_{t=p+1}^{n}A_t(\boldsymbol{u}_n)=\sqrt{n}\boldsymbol{u}_n'\boldsymbol{T}_{1n}+(\sqrt{n}\boldsymbol{u}_n)'\Sigma_{21}(\sqrt{n}\boldsymbol{u}_n)+o_p(\sqrt{n}\|\boldsymbol{u}_n\|+n\|\boldsymbol{u}_n\|^2). \tag{S3.29}$$

We next consider $B_t(\boldsymbol{u})$. Note that $\partial h_t(\boldsymbol{\delta}_0)/\partial\boldsymbol{\delta}=h_t\boldsymbol{Y}_{2t}$ and $\varepsilon_t(\boldsymbol{\alpha}_0)=h_t\eta_t$. By Taylor's expansion, we can show that

$$\begin{aligned}B_t(\boldsymbol{u})&=\ln h_t(\boldsymbol{\delta}_0+\boldsymbol{u}_2)+\frac{|\varepsilon_t(\boldsymbol{\alpha}_0)|}{h_t(\boldsymbol{\delta}_0+\boldsymbol{u}_2)}-\left(\ln h_t(\boldsymbol{\delta}_0)+\frac{|\varepsilon_t(\boldsymbol{\alpha}_0)|}{h_t(\boldsymbol{\delta}_0)}\right)\\ &=\boldsymbol{u}_2'\boldsymbol{Y}_{2t}-|\eta_t|\boldsymbol{u}_2'\boldsymbol{Y}_{2t}-\frac{\boldsymbol{u}_2'\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial\boldsymbol{\delta}}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial\boldsymbol{\delta}'}\boldsymbol{u}_2}{2h_t^2(\bar{\boldsymbol{\delta}})}+\frac{|\varepsilon_t(\boldsymbol{\alpha}_0)|\boldsymbol{u}_2'\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial\boldsymbol{\delta}}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial\boldsymbol{\delta}'}\boldsymbol{u}_2}{h_t^3(\bar{\boldsymbol{\delta}})}\end{aligned}$$

$$= \boldsymbol{u}'\boldsymbol{K}_{2t}(1-|\eta_t|) + \boldsymbol{u}'\left(\frac{|\varepsilon_t(\boldsymbol{\alpha}_0)|}{h_t(\bar{\boldsymbol{\delta}})} - \frac{1}{2}\right)\frac{1}{h_t^2(\bar{\boldsymbol{\delta}})}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial \boldsymbol{\theta}}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial \boldsymbol{\theta}'}\boldsymbol{u},$$

where $\bar{\boldsymbol{\delta}}$ lies between $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}_0 + \boldsymbol{u}_2$, and $\boldsymbol{K}_{2t} = (\boldsymbol{0}_p', \boldsymbol{Y}_{2t}')'$. Then, we have

$$\sum_{t=p+1}^{n} B_t(\boldsymbol{u}) = \sqrt{n}\boldsymbol{u}'\boldsymbol{T}_{2n} + \Pi_{3n}(\boldsymbol{u}), \tag{S3.30}$$

where

$$\boldsymbol{T}_{2n} = \frac{1}{\sqrt{n}}\sum_{t=p+1}^{n}\boldsymbol{K}_{2t}(1-|\eta_t|) \text{ and } \Pi_{3n}(\boldsymbol{u}) = \boldsymbol{u}'\sum_{t=p+1}^{n}\left(\frac{|\varepsilon_t(\boldsymbol{\alpha}_0)|}{h_t(\bar{\boldsymbol{\delta}})} - \frac{1}{2}\right)\frac{1}{h_t^2(\bar{\boldsymbol{\delta}})}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial \boldsymbol{\theta}}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial \boldsymbol{\theta}'}\boldsymbol{u}.$$

Rewrite $\Pi_{3n}(\boldsymbol{u})$ as $\Pi_{3n}(\boldsymbol{u}) = (\sqrt{n}\boldsymbol{u})'\left[n^{-1}\sum_{t=p+1}^{n} J_t(\bar{\boldsymbol{\delta}})\right](\sqrt{n}\boldsymbol{u})$, where

$$J_t(\bar{\boldsymbol{\delta}}) = \left(\frac{|\varepsilon_t(\boldsymbol{\alpha}_0)|}{h_t(\bar{\boldsymbol{\delta}})} - \frac{1}{2}\right)\frac{1}{h_t^2(\bar{\boldsymbol{\delta}})}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial \boldsymbol{\theta}}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial \boldsymbol{\theta}'}.$$

By the ergodic theorem, it is easy to see that

$$\frac{1}{n}\sum_{t=p+1}^{n} J_t(\bar{\boldsymbol{\delta}}) = E[J_t(\bar{\boldsymbol{\delta}})] + o_p(1).$$

Since $\bar{\boldsymbol{\delta}} \to_p \boldsymbol{\delta}_0$, by the dominated convergence theorem, we have

$$\lim_{n\to\infty} E[J_t(\bar{\boldsymbol{\delta}})] = E[J_t(\boldsymbol{\theta}_0)] = \Sigma_{22},$$

where $\Sigma_{22} = \operatorname{diag}\{0_{p\times p}, E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}')/2\}$. Thus, for $\boldsymbol{u}_n = o_p(1)$, we can get that

$$\Pi_{3n}(\boldsymbol{u}_n) = (\sqrt{n}\boldsymbol{u}_n)'\Sigma(\sqrt{n}\boldsymbol{u}_n) + o_p(n\|\boldsymbol{u}_n\|^2). \tag{S3.31}$$

Therefore, (S3.30) and (S3.31) imply that

$$\sum_{t=p+1}^{n} B_t(\boldsymbol{u}_n) = \sqrt{n}\boldsymbol{u}_n'\boldsymbol{T}_{2n} + (\sqrt{n}\boldsymbol{u}_n)'\Sigma_{22}(\sqrt{n}\boldsymbol{u}_n) + o_p(n\|\boldsymbol{u}_n\|^2). \tag{S3.32}$$

Finally, we consider $C_t(\boldsymbol{u})$. By Taylor's expansion, we have

$$\frac{1}{h_t(\boldsymbol{\delta}_0 + \boldsymbol{u}_2)} - \frac{1}{h_t(\boldsymbol{\delta}_0)} = -\frac{\boldsymbol{u}'}{h_t^2(\boldsymbol{\delta}_0)}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial \boldsymbol{\theta}},$$

where $\bar{\boldsymbol{\delta}}$ lies between $\boldsymbol{\delta}_0$ and $\boldsymbol{\delta}_0 + \boldsymbol{u}_2$. This together with (S3.21), implies that

$$\sum_{t=p+1}^{n} C_t(\boldsymbol{u}) = (\sqrt{n}\boldsymbol{u})'[K_{3n}(\boldsymbol{u}) + K_{4n}(\boldsymbol{u})](\sqrt{n}\boldsymbol{u}),$$

where

$$K_{3n}(\boldsymbol{u}) = \frac{1}{n}\sum_{t=p+1}^{n} \frac{1}{h_t}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial \boldsymbol{\theta}} \boldsymbol{K}'_{1t}[I(\eta_t < 0) - I(\eta_t > 0)],$$

$$K_{4n}(\boldsymbol{u}) = -\frac{1}{n}\sum_{t=p+1}^{n} \frac{2}{h_t}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial \boldsymbol{\theta}} \boldsymbol{K}'_{1t}\int_0^1 X_t(\boldsymbol{u}'\boldsymbol{K}_{1t}s)ds.$$

Since η_t has zero median by Assumption 3(i), then by the ergodic theorem, we have

$$K_{3n}(\boldsymbol{u}) = E\left\{\frac{1}{h_t}\frac{\partial h_t(\bar{\boldsymbol{\delta}})}{\partial \boldsymbol{\theta}} \boldsymbol{K}'_{1t}[I(\eta_t < 0) - I(\eta_t > 0)]\right\} + o_p(1) = o_p(1).$$

Since $\boldsymbol{u}_n = o_p(1)$, similar to the proof of $K_{2n}(\boldsymbol{u})$, we can show that $K_{4n}(\boldsymbol{u}_n) = o_p(1)$. Therefore, it follows that

$$\sum_{t=p+1}^{n} C_t(\boldsymbol{u}_n) = o_p(n\|\boldsymbol{u}_n\|^2) \tag{S3.33}$$

Combining (S3.29), (S3.32) and (S3.33), we have

$$(n-p)[L_n(\boldsymbol{\theta}_0 + \boldsymbol{u}_n) - L_n(\boldsymbol{\theta}_0)] = \sqrt{n}\boldsymbol{u}_n'\boldsymbol{T}_n + (\sqrt{n}\boldsymbol{u}_n)'\Sigma(\sqrt{n}\boldsymbol{u}_n) + o_p(\sqrt{n}\|\boldsymbol{u}_n\| + n\|\boldsymbol{u}_n\|^2),$$

where $\boldsymbol{T}_n = \boldsymbol{T}_{1n} + \boldsymbol{T}_{2n}$ and $\Sigma = \Sigma_{21} + \Sigma_{22} = \text{diag}\,\{f(0)E(\boldsymbol{Y}_{1t}\boldsymbol{Y}'_{1t}), E(\boldsymbol{Y}_{2t}\boldsymbol{Y}'_{2t})/2\}$.

Moreover, let $\boldsymbol{G}_t = (\boldsymbol{Y}'_{1t}[I(\eta_t < 0) - I(\eta_t > 0)], \boldsymbol{Y}'_{2t}(1 - |\eta_t|))'$, we have $\boldsymbol{T}_n = \boldsymbol{T}_{1n} + \boldsymbol{T}_{2n} = n^{-1/2}\sum_{t=p+1}^{n}\boldsymbol{G}_t$. If Assumptions 1, 2 and 3 hold, by the Central Limit Theorem, we have

$$\boldsymbol{T}_n \to_{\mathcal{L}} N(0, \Omega) \quad as \quad n \to \infty, \tag{S3.34}$$

where Ω is defined as in Lemma 3. We accomplish the proof of this lemma. □

Proof of Theorem 2. We have $\widehat{\boldsymbol{u}}_n = \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_p(1)$ by Theorem 1. Furthermore, by Lemma 3, we have

$$H_n(\widehat{\boldsymbol{u}}_n) = \sqrt{n}\widehat{\boldsymbol{u}}_n'\boldsymbol{T}_n + (\sqrt{n}\widehat{\boldsymbol{u}}_n)'\Sigma(\sqrt{n}\widehat{\boldsymbol{u}}_n) + o_p(\sqrt{n}\|\widehat{\boldsymbol{u}}_n\| + n\|\widehat{\boldsymbol{u}}_n\|^2) \tag{S3.35}$$

$$\geq -\sqrt{n}\|\widehat{\boldsymbol{u}}_n\| \left[\|\boldsymbol{T}_n\| + o_p(1)\right] + n\|\widehat{\boldsymbol{u}}_n\|^2[\lambda_{\min} + o_p(1)],$$

where $\lambda_{\min} > 0$ is the minimum eigenvalue of Σ. Note that $H_n(\widehat{\boldsymbol{u}}_n) = (n-p)[L_n^E(\widehat{\boldsymbol{\theta}}_n) - L_n^E(\boldsymbol{\theta}_0)] \leq 0$. Then it follows that

$$n\|\widehat{\boldsymbol{u}}_n\| \leq [\lambda_{\min} + o_p(1)]^{-1} \left[\|\boldsymbol{T}_n\| + o_p(1)\right] = O_p(1). \tag{S3.36}$$

This together with Theorem 1, verifies the $\sqrt{n}$-consistency. Hence, Statement (i) holds.

Next, let $\sqrt{n}\boldsymbol{u}_n^* = -\Sigma^{-1}\boldsymbol{T}_n/2$, then, by (S3.34) from Lemma 3, we have

$$\sqrt{n}\boldsymbol{u}_n^* \to_{\mathcal{L}} N\left(0, \frac{1}{4}\Sigma^{-1}\Omega\Sigma^{-1}\right) \text{ as } n \to \infty.$$

As a result, it is sufficient to show that $\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n^* = o_p(1)$. By (S3.35) and (S3.36), we have

$$\begin{aligned} H_n(\widehat{\boldsymbol{u}}_n) =& (\sqrt{n}\widehat{\boldsymbol{u}}_n)'\boldsymbol{T}_n + (\sqrt{n}\widehat{\boldsymbol{u}}_n)'\Sigma(\sqrt{n}\widehat{\boldsymbol{u}}_n) + o_p(1) \\ =& -2(\sqrt{n}\widehat{\boldsymbol{u}}_n)'\Sigma(\sqrt{n}\boldsymbol{u}_n^*) + (\sqrt{n}\widehat{\boldsymbol{u}}_n)'\Sigma(\sqrt{n}\widehat{\boldsymbol{u}}_n) + o_p(1). \end{aligned}$$

Note that (S3.35) still holds when $\widehat{\boldsymbol{u}}_n$ is replaced by $\boldsymbol{u}_n^*$. Thus,

$$\begin{aligned} H_n(\boldsymbol{u}_n^*) =& (\sqrt{n}\boldsymbol{u}_n^*)'\boldsymbol{T}_n + (\sqrt{n}\boldsymbol{u}_n^*)'\Sigma(\sqrt{n}\boldsymbol{u}_n^*) + o_p(1) \\ =& -(\sqrt{n}\boldsymbol{u}_n^*)'\Sigma(\sqrt{n}\boldsymbol{u}_n^*) + o_p(1). \end{aligned}$$

By the previous two equations, it follows that

$$\begin{aligned} H_n(\widehat{\boldsymbol{u}}_n) - H_n(\boldsymbol{u}_n^*) =& (\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n^*)'\Sigma(\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n^*) + o_p(1) \\ \geq& \lambda_{\min}\|\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n^*\|^2 + o_p(1). \end{aligned} \tag{S3.37}$$

Since $H_n(\widehat{\boldsymbol{u}}_n) - H_n(\boldsymbol{u}_n^*) = (n-p)\left[L_n^E(\boldsymbol{\theta}_0 + \widehat{\boldsymbol{u}}_n) - L_n^E(\boldsymbol{\theta}_0 + \boldsymbol{u}_n^*)\right] \leq 0$ almost surely, by (S3.37) we have $\|\sqrt{n}\widehat{\boldsymbol{u}}_n - \sqrt{n}\boldsymbol{u}_n^*\| = o_p(1)$. Therefore, we have

$$\sqrt{n}\widehat{\boldsymbol{u}}_n \to_{\mathcal{L}} N\left(0, \frac{1}{4}\Sigma^{-1}\Omega\Sigma^{-1}\right) \text{ as } n \to \infty.$$

Therefore, Statement (ii) holds. The proof is accomplished. □

S3.3 Proof of Remark 1

Proof. Recall that the error function is $\eta_t(\boldsymbol{\theta}) = \varepsilon_t(\boldsymbol{\alpha})/h_t(\boldsymbol{\delta})$, where $\varepsilon_t(\boldsymbol{\alpha}) = y_t - \sum_{i=1}^p \alpha_i y_{t-i}$ and $h_t(\boldsymbol{\delta}) = \omega + \sum_{i=1}^p \beta_i |y_{t-i}|$, and the residuals are defined as $\widehat{\eta}_t = \eta_t(\widehat{\boldsymbol{\theta}}_n) = \varepsilon_t(\widehat{\boldsymbol{\alpha}}_n)/h_t(\widehat{\boldsymbol{\delta}}_n)$. Note that $\eta_t = \eta_t(\boldsymbol{\theta}_0)$ and $\widehat{f}(0) = (nb_n)^{-1}\sum_{t=p+1}^n K(\widehat{\eta}_t/b_n)$. Since $|K(x) - K(y)| \le L|x-y|$ for some $L > 0$, by Taylor's expansion, it holds that

$$\begin{aligned}\left|\widehat{f}(0) - \frac{1}{nb_n}\sum_{t=p+1}^n K\left(\frac{\eta_t}{b_n}\right)\right| &= \frac{1}{nb_n}\sum_{t=p+1}^n \left|K\left(\frac{\widehat{\eta}_t}{b_n}\right) - K\left(\frac{\eta_t}{b_n}\right)\right| \\ &\le \frac{L}{nb_n^2}\sum_{t=p+1}^n |\widehat{\eta}_t - \eta_t| \\ &\le \frac{L\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|}{nb_n^2}\sum_{t=p+1}^n \left\|\frac{\partial \eta_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}\right\|, \end{aligned} \tag{S3.38}$$

where $\boldsymbol{\theta}^*$ lies between $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}_n$. It can be verified that

$$\frac{\partial \eta_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{h_t(\boldsymbol{\delta})}\frac{\partial \varepsilon_t(\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}} - \frac{\varepsilon_t(\boldsymbol{\alpha})}{h_t^2(\boldsymbol{\delta})}\frac{\partial h_t(\boldsymbol{\delta})}{\partial \boldsymbol{\theta}} = -\frac{h_t}{h_t(\boldsymbol{\delta})}\boldsymbol{K}_{1t} - \frac{h_t\varepsilon_t(\boldsymbol{\alpha})}{h_t^2(\boldsymbol{\delta})}\boldsymbol{K}_{2t}, \tag{S3.39}$$

where $\boldsymbol{K}_{1t} = (\boldsymbol{Y}_{1t}', \boldsymbol{0}_{p+1}')'$ and $\boldsymbol{K}_{2t} = (\boldsymbol{0}_p', \boldsymbol{Y}_{2t}')'$. Since $\|\boldsymbol{K}_{1t}\| \le p/\underline{\beta}$ by (S3.13), we can show that $\|\boldsymbol{K}_{2t}\| \le \underline{\omega}^{-1} + p/\underline{\beta}$, and $\sup_{\boldsymbol{\theta}\in\Theta} h_t/h_t(\boldsymbol{\delta}) \le \omega_0/\underline{\omega} + \sum_{i=1}^p \beta_{i0}/\underline{\beta}$ by Assumption 1. These together with (S3.2), imply that

$$E\left[\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial \eta_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right\|\right] \le E\left[\sup_{\boldsymbol{\theta}\in\Theta}\frac{h_t}{h_t(\boldsymbol{\delta})}\|\boldsymbol{K}_{1t}\|\right] + E\left[\sup_{\boldsymbol{\theta}\in\Theta}\frac{h_t}{h_t(\boldsymbol{\delta})}\frac{|\varepsilon_t(\boldsymbol{\alpha})|}{h_t(\boldsymbol{\delta})}\|\boldsymbol{K}_{2t}\|\right] < \infty.$$

Then by Theorem 3.1 in Ling and McAleer (2003) and the Dominated Convergence Theorem, we have

$$\frac{1}{n}\sum_{t=p+1}^n \left\|\frac{\partial \eta_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}\right\| = E\left\|\frac{\partial \eta_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}\right\| + o_p(1) = E\left\|\frac{\partial \eta_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right\| + o_p(1) = O_p(1). \tag{S3.40}$$

Since $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_p(1)$ by Theorem 2 and $nb_n^4 \to \infty$ as $n \to \infty$, by (S3.38) and (S3.40), we have

$$\left| \widehat{f}(0) - \frac{1}{nb_n} \sum_{t=p+1}^{n} K\left(\frac{\eta_t}{b_n}\right) \right| \leq O_p\left(\frac{1}{\sqrt{n}b_n^2}\right) = o_p(1). \tag{S3.41}$$

Moreover, since $\sup_x f(x) < \infty$ by Assumption 3(ii) and $\int_{-\infty}^{\infty} K(x)dx = 1$, it holds that

$$E\left[\frac{1}{b_n} K\left(\frac{\eta_t}{b_n}\right)\right] = \int_{-\infty}^{\infty} K(x) f(b_n x) dx < \infty.$$

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

$$\frac{1}{nb_n} \sum_{t=p+1}^{n} K\left(\frac{\eta_t}{b_n}\right) = E\left[\frac{1}{b_n} K\left(\frac{\eta_t}{b_n}\right)\right] + o_p(1). \tag{S3.42}$$

Furthermore, since $\int_{-\infty}^{\infty} |x| K(x)dx < \infty$, $\sup_x |f'(x)| < \infty$ and $b_n \to 0$ as $n \to \infty$, it can be verified that

$$\begin{aligned} \left| E\left[\frac{1}{b_n} K\left(\frac{\eta_t}{b_n}\right)\right] - f(0) \right| &= \left| \int_{-\infty}^{\infty} K(x)[f(b_n x) - f(0)] dx \right| \\ &\leq b_n \sup_x |f'(x)| \int_{-\infty}^{\infty} |x| K(x) dx \to 0 \ \text{ as } \ n \to \infty. \end{aligned} \tag{S3.43}$$

By (S3.41)–(S3.43), we have $\widehat{f}(0) = f(0) + o_p(1)$. Finally, by a similar argument as for (S3.40), we can show that $(n-p)^{-1} \sum_{t=p+1}^{n} \boldsymbol{Y}_{1t} \boldsymbol{Y}_{1t}' = E(\boldsymbol{Y}_{1t} \boldsymbol{Y}_{1t}') + o_p(1)$, $(n-p)^{-1} \sum_{t=p+1}^{n} \boldsymbol{Y}_{2t} \boldsymbol{Y}_{2t}' = E(\boldsymbol{Y}_{2t} \boldsymbol{Y}_{2t}') + o_p(1)$, $(n-p)^{-1} \sum_{t=p+1}^{n} \boldsymbol{Y}_{1t} \boldsymbol{Y}_{2t}' = E(\boldsymbol{Y}_{1t} \boldsymbol{Y}_{2t}') + o_p(1)$, $\widehat{\kappa}_1 = (n-p)^{-1} \sum_{t=p+1}^{n} \widehat{\eta}_t = E(\eta_t) + o_p(1)$ and $\widehat{\kappa}_2 = (n-p)^{-1} \sum_{t=p+1}^{n} \widehat{\eta}_t^2 - 1 = E(\eta_t^2) - 1 + o_p(1)$. As a result, $\widehat{\Sigma}_n \to_p \Sigma$ and $\widehat{\Omega}_n \to_p \Omega$ as $n \to \infty$. The proof is accomplished. □

S3.4 Proof of Theorem 3

To show Theorem 3, we introduce the following Lemmas 4–7.

Lemma 4. *If Assumptions 1–2 hold, then it holds that*

$$\textit{(i)}\ E \sup_{\boldsymbol{\theta} \in \Theta} \left| \ell_t^G(\boldsymbol{\theta}) \right| < \infty; \qquad \textit{(ii)}\ E \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial \ell_t^G(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| < \infty; \qquad \textit{(iii)}\ E \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial^2 \ell_t^G(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| < \infty.$$

Proof. Recall that $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \boldsymbol{\delta}')'$ and $\ell_t^G(\boldsymbol{\theta}) = \ln h_t(\boldsymbol{\delta}) + 0.5\varepsilon_t^2(\boldsymbol{\alpha})/h_t^2(\boldsymbol{\delta})$, where $\varepsilon_t(\boldsymbol{\alpha}) = y_t - \sum_{i=1}^p \alpha_i y_{t-i}$ and $h_t(\boldsymbol{\delta}) = \omega + \sum_{i=1}^p \beta_i |y_{t-i}|$. It can be derived that

$$\frac{\partial \ell_t^G(\boldsymbol{\theta})}{\partial \alpha_i} = -\frac{\varepsilon_t(\boldsymbol{\alpha}) y_{t-i}}{h_t^2(\boldsymbol{\delta})}, \ \frac{\partial \ell_t^G(\boldsymbol{\theta})}{\partial \omega} = \frac{1}{h_t(\boldsymbol{\delta})}\left(1 - \frac{\varepsilon_t^2(\boldsymbol{\alpha})}{h_t^2(\boldsymbol{\delta})}\right), \ \frac{\partial \ell_t^G(\boldsymbol{\theta})}{\partial \beta_i} = \frac{|y_{t-i}|}{h_t(\boldsymbol{\delta})}\left(1 - \frac{\varepsilon_t^2(\boldsymbol{\alpha})}{h_t^2(\boldsymbol{\delta})}\right),$$
$$\frac{\partial^2 \ell_t^G(\boldsymbol{\theta})}{\partial \alpha_i \partial \alpha_j} = \frac{y_{t-i} y_{t-j}}{h_t^2(\boldsymbol{\delta})}, \ \frac{\partial^2 \ell_t^G(\boldsymbol{\theta})}{\partial \alpha_i \partial \omega} = \frac{2\varepsilon_t(\boldsymbol{\alpha}) y_{t-i}}{h_t^3(\boldsymbol{\delta})}, \ \frac{\partial^2 \ell_t^G(\boldsymbol{\theta})}{\partial \alpha_i \partial \beta_j} = \frac{2\varepsilon_t(\boldsymbol{\alpha}) y_{t-i} |y_{t-j}|}{h_t^3(\boldsymbol{\delta})},$$
$$\frac{\partial^2 \ell_t^G(\boldsymbol{\theta})}{\partial^2 \omega} = -\frac{1}{h_t^2(\boldsymbol{\delta})}\left(1 - \frac{3\varepsilon_t^2(\boldsymbol{\alpha})}{h_t^2(\boldsymbol{\delta})}\right), \ \frac{\partial^2 \ell_t^G(\boldsymbol{\theta})}{\partial \omega \partial \beta_j} = -\frac{|y_{t-j}|}{h_t^2(\boldsymbol{\delta})}\left(1 - \frac{3\varepsilon_t^2(\boldsymbol{\alpha})}{h_t^2(\boldsymbol{\delta})}\right) \text{ and}$$
$$\frac{\partial^2 \ell_t^G(\boldsymbol{\theta})}{\partial \beta_i \partial \beta_j} = -\frac{|y_{t-i} y_{t-j}|}{h_t^2(\boldsymbol{\delta})}\left(1 - \frac{3\varepsilon_t^2(\boldsymbol{\alpha})}{h_t^2(\boldsymbol{\delta})}\right).$$

We first show (i). By Assumption 2, there exists a constant $0 < \kappa \leq 1$ such that $E(|y_t|^\kappa) < \infty$. Let $\overline{\omega}^\star = \max\{1, \overline{\omega}\}$. By Jensen's inequality, we have

$$\begin{aligned} E \ln\left(\overline{\omega}^\star + \overline{\beta}\sum_{i=1}^p |y_{t-i}|\right) &= \frac{1}{\kappa} E \ln\left(\overline{\omega}^\star + \overline{\beta}\sum_{i=1}^p |y_{t-i}|\right)^\kappa \\ &\leq \frac{1}{\kappa} E \ln\left(\overline{\omega}^{\star\kappa} + \overline{\beta}^\kappa \sum_{i=1}^p |y_{t-i}|^\kappa\right) \\ &\leq \frac{1}{\kappa} \ln\left(\overline{\omega}^{\star\kappa} + \overline{\beta}^\kappa \sum_{i=1}^p E(|y_{t-i}|^\kappa)\right) < \infty, \end{aligned}$$

where the following elementary relation is used: $(\sum_{i=1}^p \alpha_i)^s \leq \sum_{i=1}^p \alpha_i^s$ for all $\alpha_i > 0$ and $s \in (0, 1]$. This together with Assumption 2, implies that

$$\begin{aligned} E \sup_{\boldsymbol{\theta} \in \Theta} |\ln h_t(\boldsymbol{\delta})| \leq & E \sup_{\boldsymbol{\theta} \in \Theta} \left[I\left(\omega + \overline{\beta}\sum_{i=1}^p |y_{t-i}| \geq 1\right) \ln\left(\omega + \overline{\beta}\sum_{i=1}^p |y_{t-i}|\right)\right] \\ & + E \sup_{\boldsymbol{\theta} \in \Theta} \left[-I\left(\omega + \overline{\beta}\sum_{i=1}^p |y_{t-i}| \leq 1\right) \ln\left(\omega + \overline{\beta}\sum_{i=1}^p |y_{t-i}|\right)\right] \\ \leq & E \ln\left(\overline{\omega}^\star + \overline{\beta}\sum_{i=1}^p |y_{t-i}|\right) - I\{\underline{\omega} < 1\} \ln \underline{\omega} < \infty. \end{aligned} \tag{S3.44}$$

By Assumption 4(i), we have $E(\eta_t) = 0$ and $E(\eta_t^2) = 1$. Moreover, note that η_t is independent of $\mathcal{F}_{t-1}$, and $y_t - \sum_{i=1}^p \alpha_i y_{t-i} = \sum_{i=1}^p (\alpha_{i0} - \alpha_i) y_{t-i} + \eta_t \left(\omega_0 + \sum_{i=1}^p \beta_{i0} |y_{t-i}|\right)$, then by

Assumption 1 and c_r inequality, it can be verified that

$$
\begin{aligned}
&E\sup_{\boldsymbol{\theta}\in\Theta}\left[\frac{\left(y_t-\sum_{i=1}^p\alpha_i y_{t-i}\right)^2}{\left(\omega+\sum_{i=1}^p\beta_i|y_{t-i}|\right)^2}\right]\\
\leq& E\sup_{\boldsymbol{\theta}\in\Theta}\left[\left(\sum_{i=1}^p\frac{(\alpha_{i0}-\alpha_i)\,y_{t-i}}{\underline{\omega}+\underline{\beta}\sum_{i=1}^p|y_{t-i}|}\right)^2\right]+E\left[\left(\frac{\overline{\omega}+\bar{\beta}\sum_{i=1}^p|y_{t-i}|}{\underline{\omega}+\underline{\beta}\sum_{i=1}^p|y_{t-i}|}\right)^2\right]\\
\leq& E\sup_{\boldsymbol{\theta}\in\Theta}\left[\left(\sum_{i=1}^p\frac{(\alpha_{i0}-\alpha_i)\,y_{t-i}}{\underline{\beta}|y_{t-i}|}\right)^2\right]\\
&+2E\left[\left(\frac{\overline{\omega}}{\underline{\omega}+\underline{\beta}\sum_{i=1}^p|y_{t-i}|}\right)^2\right]+2E\left[\left(\sum_{i=1}^p\frac{\bar{\beta}|y_{t-i}|}{\underline{\omega}+\underline{\beta}\sum_{i=1}^p|y_{t-i}|}\right)^2\right]\\
\leq&\frac{p}{\underline{\beta}^2}\sum_{i=1}^p\sup_{\boldsymbol{\theta}\in\Theta}(\alpha_{i0}-\alpha_i)^2+\frac{2\overline{\omega}^2}{\underline{\omega}^2}+\frac{2p^2\bar{\beta}^2}{\underline{\beta}^2}<\infty.
\end{aligned}
\tag{S3.45}
$$

By (S3.44), (S3.45) and the triangle inequality, we have

$$
E\sup_{\boldsymbol{\theta}\in\Theta}\left|\ell_t^G(\boldsymbol{\theta})\right|\leq E\sup_{\boldsymbol{\theta}\in\Theta}|\ln h_t(\boldsymbol{\delta})|+\frac{1}{2}E\sup_{\boldsymbol{\theta}\in\Theta}\frac{\varepsilon_t^2(\boldsymbol{\alpha})}{h_t^2(\boldsymbol{\delta})}<\infty.
$$

Thus, (i) is verified. Similarly, we can show that (ii) and (iii) hold. □

Lemma 5. *If Assumptions 1 and 2 hold, then*

$$
\begin{aligned}
&(i)\ \sup_{\boldsymbol{\theta}\in\Theta}\left|L_n^G(\boldsymbol{\theta})-E\left[\ell_t^G(\boldsymbol{\theta})\right]\right|=o_p(1);\\
&(ii)\ \sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial L_n^G(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}-E\left[\frac{\partial\ell_t^G(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\right\|=o_p(1);\\
&(iii)\ \sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{\partial^2 L_n^G(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}-E\left[\frac{\partial^2\ell_t^G(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right\|=o_p(1).
\end{aligned}
$$

Proof. These follow from Lemma 4 and Theorem 3.1 in Ling and McAleer (2003). □

Lemma 6. *If Assumptions 1, 2 and 4(i) hold, then $E[\ell_t^G(\boldsymbol{\theta})]$ has a unique minimum at $\boldsymbol{\theta}_0$.*

Proof. We first prove that

$$
\boldsymbol{c}_1=\mathbf{0}_p\quad\text{if}\quad\boldsymbol{c}_1'\boldsymbol{Y}_{1t}=0\ a.s.\quad\text{and}\quad\boldsymbol{c}_2=\mathbf{0}_{p+1}\quad\text{if}\quad\boldsymbol{c}_2'\boldsymbol{Y}_{2t}=0\ a.s.,
\tag{S3.46}
$$

where $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ are $p \times 1$ and $(p+1) \times 1$ constant vectors, respectively. If $\boldsymbol{c}_1' \boldsymbol{Y}_{1t} = 0$ $a.s.$ and $\boldsymbol{c}_1 = (c_1, \cdots, c_p)' \neq \mathbf{0}$, without loss of generality, we can assume $c_1 = 1$, thus $y_t = -\sum_{i=2}^{p} c_i y_{t-i+1}$ $a.s.$. Recall that $\eta_t = \varepsilon_t(\boldsymbol{\alpha}_0)/h_t(\boldsymbol{\delta}_0)$ with $\varepsilon_t(\boldsymbol{\alpha}_0), h_t(\boldsymbol{\delta}_0) \in \mathcal{F}_{t-1}$, and η_t is independent of $\mathcal{F}_{t-1}$, we have

$$E(\eta_t^2) = E(\eta_t) E\left(\frac{-\sum_{i=2}^{p} c_{1i} y_{t-i+1} - \varepsilon_t(\boldsymbol{\alpha}_0)}{h_t(\boldsymbol{\delta}_0)} \right) = 0, \tag{S3.47}$$

which is a contradiction with $E(\eta_t^2) = 1$ by Assumption 4(i), thus $\boldsymbol{c}_1 = \mathbf{0}_p$. Similarly, we can show that $\boldsymbol{c}_2 = \mathbf{0}_{p+1}$.

Note that $\varepsilon_t(\boldsymbol{\alpha}) = \varepsilon_t(\boldsymbol{\alpha}_0) - (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)' \boldsymbol{Y}_{1t} h_t(\boldsymbol{\delta}_0)$ and $\varepsilon_t(\boldsymbol{\alpha}_0) = \eta_t h_t(\boldsymbol{\delta}_0)$. As for (S3.45), then by the law of iterated expectations and Assumption 4(i), we can show that

$$\begin{aligned} E[\ell_t^G(\boldsymbol{\theta})] =& E\left\{ \ln h_t(\boldsymbol{\delta}) + \frac{[\varepsilon_t(\boldsymbol{\alpha}_0) - (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)' \boldsymbol{Y}_{1t} h_t(\boldsymbol{\delta}_0)]^2}{2h_t^2(\boldsymbol{\delta})} \right\} \\ =& E\left[\ln h_t(\boldsymbol{\delta}) + \frac{1}{2}\left(\frac{h_t(\boldsymbol{\delta}_0)}{h_t(\boldsymbol{\delta})} \right)^2 \right] - \frac{1}{2} E\left[\frac{(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)' \boldsymbol{Y}_{1t} h_t(\boldsymbol{\delta}_0)}{h_t(\boldsymbol{\delta})} \right]^2. \end{aligned} \tag{S3.48}$$

The second term in (S3.48) reaches its minimum at zero, and this happens if and only if $(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)' \boldsymbol{Y}_{1t} = 0$ $a.s.$, which holds if and only if $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ by (S3.46). For the first term in (S3.48), denote $f(x) = -\ln(x) - 0.5a^2/x^2$, where $x = h_t(\boldsymbol{\delta})$ and $a = h_t(\boldsymbol{\delta}_0)$. We can prove that $f(x)$ reaches its minimum at $x = a$, i.e. $h_t(\boldsymbol{\delta}) = h_t(\boldsymbol{\delta}_0)$, which holds if and only if $\boldsymbol{\delta} = \boldsymbol{\delta}_0$ by (S3.46). Therefore, $E[\ell_t^G(\boldsymbol{\theta})]$ is uniquely minimized at $\boldsymbol{\theta}_0$. □

Lemma 7. *Suppose Assumptions 1, 2 and 4 hold, then*

(i) Ω_1 *and* Σ_1 *are finite and positive definite;*

(ii) $\sqrt{n} \partial L_n^G(\boldsymbol{\theta}_0) / \partial \boldsymbol{\theta} \to_{\mathcal{L}} N(0, \Omega_1)$ *as* $n \to \infty$.

Proof. We first show (i). Recall that $\kappa_3 = E\eta_t^3$, $\kappa_4 = E\eta_t^4 - 1$,

$$\Sigma_1 = \text{diag}\left\{E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}'), 2E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}')\right\} \text{ and } \Omega_1 = \begin{pmatrix} E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}') & \kappa_3 E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{2t}') \\ \kappa_3 E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{1t}') & \kappa_4 E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}') \end{pmatrix}.$$

By Assumptions 1–2, for some constant C, we have

$$||E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{1t}')|| < C,\ ||E(\boldsymbol{Y}_{1t}\boldsymbol{Y}_{2t}')|| < C \text{ and } ||E(\boldsymbol{Y}_{2t}\boldsymbol{Y}_{2t}')|| < C.$$

Thus, Σ_1 is finite. Since $E(\eta_t^4) < \infty$ by Assumption 4(ii), we have $\kappa_1, \kappa_2 < \infty$, then Ω_1 is also finite.

Let $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$, where $\boldsymbol{x}_1 \in \mathbb{R}^p$ and $\boldsymbol{x}_2 \in \mathbb{R}^{p+1}$ are arbitrary non-zero constant vectors. It follows that

$$\begin{aligned} \boldsymbol{x}'\Omega_1\boldsymbol{x} &= E\left\{(\boldsymbol{x}_1'\boldsymbol{Y}_{1t})^2 + \kappa_4(\boldsymbol{x}_2'\boldsymbol{Y}_{2t})^2 + 2\kappa_3\boldsymbol{x}_2'\boldsymbol{Y}_{2t}\boldsymbol{Y}_{1t}'\boldsymbol{x}_1\right\} \\ &= E\left\{(\boldsymbol{x}_1'\boldsymbol{Y}_{1t} + \kappa_3\boldsymbol{x}_2'\boldsymbol{Y}_{2t})^2 + (\kappa_4 - \kappa_3^2)(\boldsymbol{x}_2'\boldsymbol{Y}_{2t})^2\right\}. \end{aligned} \tag{S3.49}$$

By Cauchy-Schwarz inequality and Assumption 4(i), we have $\kappa_3^2 = \left[\text{cov}(\eta_t, \eta_t^2)\right]^2 \leq \text{var}(\eta_t) \times \text{var}(\eta_t^2) = \kappa_4$, and the equality holds when $P(\eta_t^2 - c\eta_t = 1) = 1$ for any $c \in \mathbb{R}$, which is equivalent to $\det(D) = 0$. Since D is positive definite, we have $\kappa_4 - \kappa_3^2 > 0$ and thus $\boldsymbol{x}'\Omega_1\boldsymbol{x} > 0$, i.e. Ω_1 is positive definite. Moreover, by (S3.46), it can be verified that

$$\boldsymbol{x}'\Sigma_1\boldsymbol{x} = E\left\{(\boldsymbol{x}_1'\boldsymbol{Y}_{1t})^2 + 2(\boldsymbol{x}_2'\boldsymbol{Y}_{2t})^2\right\} > 0.$$

As a result, Σ_1 is positive definite. Hence, (i) holds.

Note that $\Omega_1 = E\left[\partial\ell_t\left(\boldsymbol{\theta}_0\right)/\partial\boldsymbol{\theta}\partial\ell_t\left(\boldsymbol{\theta}_0\right)/\partial\boldsymbol{\theta}'\right]$. By the Martingale Central Limit Theorem and the Cramér–Wold device, we can show that (ii) holds. □

Proof of Theorem 3. By Lemma 5(i) and Lemma 6, we have established all the conditions for consistency in Theorem 4.1.1 in Amemiya (1985), and hence $\widetilde{\boldsymbol{\theta}}_n \to_p \boldsymbol{\theta}_0$ as $n \to \infty$.

By Lemma 5(iii), for any $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + o_p(1)$, we have $\partial^2 L_n^G(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}' = \Sigma_1 + o_p(1)$. By Taylor's expansion and the consistency of $\widetilde{\boldsymbol{\theta}}_n$, then we have

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n}\frac{\partial \ell_t^G(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + o_p(1). \tag{S3.50}$$

This together with Lemma 7, we have established all the conditions of Theorem 4.1.3 in Amemiya (1985), and hence the asymptotic normality follows. □

S3.5 Proof of Theorem 4

Proof. In the following proof, notations Θ^p, $\boldsymbol{\theta}_0^p$, $\widehat{\boldsymbol{\theta}}_n^p$, $\boldsymbol{T}_n^p$, and Σ^p are employed to emphasize their dependence on the order p. Let p_0 be the true order of model (1.1). It is sufficient to show that, for any $p \neq p_0$,

$$\lim_{n\to\infty} P(\mathrm{BIC}^E(p) - \mathrm{BIC}^E(p_0) > 0) = 1. \tag{S3.51}$$

We first consider the case that the model is overfitted, i.e. $p > p_0$. Let $\boldsymbol{\theta}_0^p = \mathrm{argmin}_{\boldsymbol{\theta}\in\Theta^p} E\ell_t^E(\boldsymbol{\theta})$ be the unique minimizer, and denote $\mathring{\boldsymbol{\theta}}_0^p = (\boldsymbol{\alpha}_0^{p_0\prime}, \mathbf{0}_{p-p_0}', \boldsymbol{\delta}_0^{p_0\prime}, \mathbf{0}_{p-p_0}')'$, where $\mathbf{0}_m$ is the $m\times 1$ vector of zeros. Let $\mathring{\Theta}^p$ be the parameter space corresponding to Θ^p with $\underline{\beta} = 0$, which allows the scale parameters β_i's to be zero. Note that $E\ell_t^E(\boldsymbol{\theta}_0^p) = \min_{\boldsymbol{\theta}\in\Theta^p} E\ell_t^E(\boldsymbol{\theta}) > \min_{\boldsymbol{\theta}\in\mathring{\Theta}^p} E\ell_t^E(\boldsymbol{\theta}) = E\ell_t^E(\mathring{\boldsymbol{\theta}}_0^p)$, where the inequality holds since $\Theta^p \subset \mathring{\Theta}^p$ and $\mathring{\boldsymbol{\theta}}_0^p \notin \Theta^p$. This together with $E\ell_t^E(\mathring{\boldsymbol{\theta}}_0^p) = E\ell_t^E(\boldsymbol{\theta}_0^{p_0})$, implies that $E\ell_t^E(\boldsymbol{\theta}_0^p) = E\ell_t^E(\boldsymbol{\theta}_0^{p_0}) + c$ for some $c > 0$. Then by ergodic theorem, we have

$$L_n^E(\boldsymbol{\theta}_0^p) = L_n^E(\boldsymbol{\theta}_0^{p_0}) + c + o_p(1).$$

Denote $\widehat{\boldsymbol{u}}_n^{p_0} = \widehat{\boldsymbol{\theta}}_n^{p_0} - \boldsymbol{\theta}_0^{p_0}$. Similar to the proof of Theorems 1 and 2, we can show that $\widehat{\boldsymbol{\theta}}_n^{p_0} \to \boldsymbol{\theta}_0^{p_0}$ almost surely and $\sqrt{n}\widehat{\boldsymbol{u}}_n^{p_0} = O_p(1)$ as $n \to \infty$. Then it follows that

$$(n - p_{\max})[L_n^E(\widehat{\boldsymbol{\theta}}_n^{p_0}) - L_n^E(\boldsymbol{\theta}_0^{p_0})]$$

$$=(\sqrt{n}\widehat{\boldsymbol{u}}_n^{p_0})'\boldsymbol{T}_n^{p_0}+(\sqrt{n}\widehat{\boldsymbol{u}}_n^{p_0})'\Sigma^{p_0}(\sqrt{n}\widehat{\boldsymbol{u}}_n^{p_0})+o_p(1)=O_p(1), \tag{S3.52}$$

Let $\boldsymbol{\theta}_0^p=\operatorname{argmin}_{\boldsymbol{\theta}\in\Theta^p}E\ell_t^E(\boldsymbol{\theta})$ be the unique minimizer. Similar to the proof of Theorems 1–2 and (S3.52), it can be verified that $\sqrt{n}\widehat{\boldsymbol{u}}_n^p=O_p(1)$ and $(n-p_{\max})[L_n^E(\widehat{\boldsymbol{\theta}}_n^p)-L_n^E(\boldsymbol{\theta}_0^p)]=O_p(1)$. Therefore, we have

$$\begin{aligned}&L_n^E(\widehat{\boldsymbol{\theta}}_n^p)-L_n^E(\widehat{\boldsymbol{\theta}}_n^{p_0})\\=&[L_n^E(\widehat{\boldsymbol{\theta}}_n^p)-L_n^E(\boldsymbol{\theta}_0^p)]-[L_n^E(\widehat{\boldsymbol{\theta}}_n^{p_0})-L_n^E(\boldsymbol{\theta}_0^{p_0})]+[L_n^E(\boldsymbol{\theta}_0^p)-L_n^E(\boldsymbol{\theta}_0^{p_0})]=O_p\left(\frac{1}{n}\right)+c+o_p(1).\end{aligned}$$

Hence, it follows that

$$\begin{aligned}&\mathrm{BIC}^E(p)-\mathrm{BIC}^E(p_0)\\=&2(n-p_{\max})[L_n^E(\widehat{\boldsymbol{\theta}}_n^p)-L_n^E(\widehat{\boldsymbol{\theta}}_n^{p_0})]+[(2p+1)\ln(n-p_{\max})-(2p_0+1)\ln(n-p_{\max})]\\=&O_p(1)+2(n-p_{\max})(c+o_p(1))+2(p-p_0)\ln(n-p_{\max})\to\infty\ \text{ as }\ n\to\infty.\end{aligned} \tag{S3.53}$$

We next consider the case that the model is underfitted, i.e. $p<p_0$. Let $\boldsymbol{\theta}_{0,E}^p=\arg\min_{\boldsymbol{\theta}\in\Theta^p}E[\ell_t^E(\boldsymbol{\theta}^p)]$. Similar to the proof of Theorems 1–2 and (S3.52), we can verify that $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n^p-\boldsymbol{\theta}_0^p)=O_p(1)$ and

$$(n-p_{\max})[L_n^E(\widehat{\boldsymbol{\theta}}_n^p)-L_n^E(\boldsymbol{\theta}_0^p)]=O_p(1).$$

Note that $L_n^E(\cdot)$ is the negative likelihood. Since the model with order p corresponds to a smaller model than the true model, we have $E[\ell_t^E(\boldsymbol{\theta}_0^p)]=E[\ell_t^E(\boldsymbol{\theta}_0^{p_0})]+\varepsilon$ for some positive constant ε. By ergodic theorem, we have $L_n^E(\boldsymbol{\theta}_0^p)=E[\ell_t^E(\boldsymbol{\theta}_0^p)]+o_p(1)$. Thus it holds that

$$L_n^E(\boldsymbol{\theta}_0^p)-L_n^E(\boldsymbol{\theta}_0^{p_0})=E[\ell_t^E(\boldsymbol{\theta}_0^p)]-E[\ell_t^E(\boldsymbol{\theta}_0^{p_0})]+o_p(1)=\varepsilon+o_p(1).$$

Therefore, we have

$$L_n^E(\widehat{\boldsymbol{\theta}}_n^p)-L_n^E(\widehat{\boldsymbol{\theta}}_n^{p_0})=[L_n^E(\widehat{\boldsymbol{\theta}}_n^p)-L_n^E(\boldsymbol{\theta}_0^p)]-[L_n^G(\widehat{\boldsymbol{\theta}}_n^{p_0})-L_n^E(\boldsymbol{\theta}_0^{p_0})]$$

$$+ [L_n^E(\boldsymbol{\theta}_0^p) - L_n^E(\boldsymbol{\theta}_0^{p_0})] = O_p\left(\frac{1}{n}\right) + \varepsilon + o_p(1).$$

This together with $(2p+1)\ln(n-p_{\max}) - (2p_0+1)\ln(n-p_{\max}) = O(\ln n)$, implies that

$$\begin{aligned}
&\mathrm{BIC}^E(p) - \mathrm{BIC}^E(p_0) \\
=&2(n-p_{\max})[L_n^E(\widehat{\boldsymbol{\theta}}_n^p) - L_n^E(\widehat{\boldsymbol{\theta}}_n^{p_0})] + [(2p+1)\ln(n-p_{\max}) - (2p_0+1)\ln(n-p_{\max})] \\
=&2(n-p_{\max})\varepsilon + o_p(n-p_{\max}) + O_p(1) + O(\ln n) \to \infty \ \text{ as } \ n \to \infty. \qquad \text{(S3.54)}
\end{aligned}$$

Hence, combining (S3.53) and (S3.54) implies that (S3.51) holds. The proof is accomplished.

□

S3.6 Proof of Remark 2

Proof. In the following proof, notations Θ^p, $\boldsymbol{\theta}_0^p$, $\widetilde{\boldsymbol{\theta}}_n^p$ and Σ_1^p are employed to emphasize their dependence on the order p. Let p_0 be the true order of model (1.1), and $\boldsymbol{\theta}_0^p = \mathrm{argmin}_{\boldsymbol{\theta}\in\Theta^p} E\ell_t^G(\boldsymbol{\theta})$ be the unique minimizer. It is sufficient to show that, for any $p \neq p_0$,

$$\lim_{n\to\infty} P(\mathrm{BIC}^G(p) - \mathrm{BIC}^G(p_0) > 0) = 1. \qquad \text{(S3.55)}$$

We first consider the case that the model is overfitted, i.e. $p > p_0$. Similar to the proof of Theorem 4 for the overfitted case, we can show that $E\ell_t^G(\boldsymbol{\theta}_0^p) = E\ell_t^G(\boldsymbol{\theta}_0^{p_0}) + c$ for some $c > 0$. Then by ergodic theorem, it follows that

$$L_n^G(\boldsymbol{\theta}_0^p) = L_n^G(\boldsymbol{\theta}_0^{p_0}) + c + o_p(1).$$

Denote $\widetilde{\boldsymbol{u}}_n^{p_0} = \widetilde{\boldsymbol{\theta}}_n^{p_0} - \boldsymbol{\theta}_0^{p_0}$. Similar to the proof of Theorem 3, we can verify that $\widetilde{\boldsymbol{\theta}}_n^{p_0} \to_p \boldsymbol{\theta}_0^{p_0}$, $\sqrt{n}\partial L_n^G(\boldsymbol{\theta}_0^{p_0})/\partial\boldsymbol{\theta} = O_p(1)$ and $\sqrt{n}\widetilde{\boldsymbol{u}}_n^{p_0} = O_p(1)$ as $n \to \infty$. By Taylor's expansion and Slutsky's theorem, it can be shown that

$$(n-p_{\max})[L_n^G(\widetilde{\boldsymbol{\theta}}_n^{p_0}) - L_n^G(\boldsymbol{\theta}_0^{p_0})]$$

$$=(\sqrt{n}\widetilde{\boldsymbol{u}}_n^{p_0})'\frac{n-p_{\max}}{\sqrt{n}}\frac{\partial L_n^G(\boldsymbol{\theta}_0^{p_0})}{\partial\boldsymbol{\theta}}-(\sqrt{n}\widetilde{\boldsymbol{u}}_n^{p_0})'\Sigma_1^{p_0}\sqrt{n}\widetilde{\boldsymbol{u}}_n^{p_0}+o_p(1)=O_p(1). \tag{S3.56}$$

Similar to the proof of Theorem 3 and (S3.56), it can be verified that $\sqrt{n}\widetilde{\boldsymbol{u}}_n^p=O_p(1)$ and $(n-p_{\max})[L_n^G(\widetilde{\boldsymbol{\theta}}_n^p)-L_n^G(\boldsymbol{\theta}_0^p)]=O_p(1)$. Therefore, we have

$$\begin{aligned}&L_n^G(\widetilde{\boldsymbol{\theta}}_n^p)-L_n^G(\widetilde{\boldsymbol{\theta}}_n^{p_0})\\=&[L_n^G(\widetilde{\boldsymbol{\theta}}_n^p)-L_n^G(\boldsymbol{\theta}_0^p)]-[L_n^G(\widetilde{\boldsymbol{\theta}}_n^{p_0})-L_n^G(\boldsymbol{\theta}_0^{p_0})]+[L_n^G(\boldsymbol{\theta}_0^p)-L_n^G(\boldsymbol{\theta}_0^{p_0})]=O_p\left(\frac{1}{n}\right)+c+o_p(1).\end{aligned}$$

Hence, we have

$$\begin{aligned}&\mathrm{BIC}^G(p)-\mathrm{BIC}^G(p_0)\\=&2(n-p_{\max})[L_n^G(\widetilde{\boldsymbol{\theta}}_n^p)-L_n^G(\widetilde{\boldsymbol{\theta}}_n^{p_0})]+[(2p+1)\ln(n-p_{\max})-(2p_0+1)\ln(n-p_{\max})]\\=&O_p(1)+2(n-p_{\max})(c+o_p(1))+2(p-p_0)\ln(n-p_{\max})\to\infty\ \text{ as }\ n\to\infty.\end{aligned}\tag{S3.57}$$

We next consider the case that the model is underfitted, i.e. $p<p_0$. Let $\boldsymbol{\theta}_{0,G}^p=\arg\min_{\boldsymbol{\theta}\in\Theta^p}E[\ell_t^G(\boldsymbol{\theta}^p)]$ and $\boldsymbol{\theta}_{0,E}^p=\arg\min_{\boldsymbol{\theta}\in\Theta^p}E[\ell_t^E(\boldsymbol{\theta}^p)]$. Similar to the proof of Theorem 3 and (S3.56), we can verify that $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n^p-\boldsymbol{\theta}_0^p)=O_p(1)$ and

$$(n-p_{\max})[L_n^G(\widetilde{\boldsymbol{\theta}}_n^p)-L_n^G(\boldsymbol{\theta}_0^p)]=O_p(1).$$

Note that $L_n^G(\cdot)$ is the negative likelihood. Since the model with order p corresponds to a smaller model than the true model, we have $E[\ell_t^G(\boldsymbol{\theta}_0^p)]\geq E[\ell_t^G(\boldsymbol{\theta}_0^{p_0})]+\varepsilon$ for some positive constant ε. By ergodic theorem, we have $L_n^G(\boldsymbol{\theta}_{0,G}^p)=E[\ell_t^G(\boldsymbol{\theta}_{0,G}^p)]+o_p(1)$. Thus it holds that

$$L_n^G(\boldsymbol{\theta}_0^p)-L_n^G(\boldsymbol{\theta}_0^{p_0})=E[\ell_t^G(\boldsymbol{\theta}_0^p)]-E[\ell_t^G(\boldsymbol{\theta}_0^{p_0})]+o_p(1)=\varepsilon+o_p(1).$$

Therefore, we have

$$\begin{aligned}L_n^G(\widetilde{\boldsymbol{\theta}}_n^p)-L_n^G(\widetilde{\boldsymbol{\theta}}_n^{p_0})=&[L_n^G(\widetilde{\boldsymbol{\theta}}_n^p)-L_n^G(\boldsymbol{\theta}_0^p)]-[L_n^G(\widetilde{\boldsymbol{\theta}}_n^{p_0})-L_n^G(\boldsymbol{\theta}_0^{p_0})]\\&+[L_n^G(\boldsymbol{\theta}_0^p)-L_n^G(\boldsymbol{\theta}_0^{p_0})]=O_p\left(\frac{1}{n}\right)+\varepsilon+o_p(1).\end{aligned}$$

This together with $(2p+1)\ln(n-p_{\max})-(2p_0+1)\ln(n-p_{\max})=O(\ln n)$, implies that

$$\begin{aligned}
&\mathrm{BIC}^G(p)-\mathrm{BIC}^G(p_0)\\
=&2(n-p_{\max})[L_n^G(\widetilde{\boldsymbol{\theta}}_n^p)-L_n^G(\widetilde{\boldsymbol{\theta}}_n^{p_0})]+[(2p+1)\ln(n-p_{\max})-(2p_0+1)\ln(n-p_{\max})]\\
=&2(n-p_{\max})\varepsilon+o_p(n-p_{\max})+O_p(1)+O(\ln n)\to\infty \ \text{ as } \ n\to\infty. \qquad \text{(S3.58)}
\end{aligned}$$

Hence, combining (S3.57) and (S3.58) implies that (S3.55) holds. The proof is accomplished.

□

S3.7 Proofs of Theorem 5 and Remark 3

Recall that the error function is defined as $\eta_t(\boldsymbol{\theta})=\varepsilon_t(\boldsymbol{\alpha})/h(\boldsymbol{\delta})$, $\kappa_1=E(\eta_t)$, $\tau_1=E[\mathrm{sgn}(\eta_t)]$, $\tau_2=E(|\eta_t|)$, $\sigma_1^2=\mathrm{var}(\eta_t)$ and $\sigma_2^2=\mathrm{var}(|\eta_t|)$. Note that $\eta_t=\eta_t(\boldsymbol{\theta}_0)$, $\widetilde{\eta}_t=\eta_t(\widetilde{\boldsymbol{\theta}}_n)$ and $\widehat{\eta}_t=\eta_t(\widehat{\boldsymbol{\theta}}_n)$. Let $\breve{\boldsymbol{\theta}}_n\in\Theta$ be a $\sqrt{n}-$estimator of $\boldsymbol{\theta}_0$ and have the asymptotic property $\sqrt{n}(\breve{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)=-D\sum_{t=1}^n\boldsymbol{G}_t+o_p(1)$, and the residual ACF and absolute residual ACF of model (1.1) fitted by $\breve{\boldsymbol{\theta}}_n$ can be defined as

$$\breve{\rho}_k=\frac{\sum_{t=p+k+1}^n(\breve{\eta}_t-\breve{\eta}_1)(\breve{\eta}_{t-k}-\breve{\eta}_1)}{\sum_{t=p+1}^n(\breve{\eta}_t-\breve{\eta}_1)^2} \text{ and } \breve{\gamma}_k=\frac{\sum_{t=p+k+1}^n(|\breve{\eta}_t|-\breve{\eta}_2)(|\breve{\eta}_{t-k}|-\breve{\eta}_2)}{\sum_{t=p+1}^n(|\breve{\eta}_t|-\breve{\eta}_2)^2},$$

where $\breve{\eta}_1=(n-p)^{-1}\sum_{t=p+1}^n\breve{\eta}_t$ and $\breve{\eta}_2=(n-p)^{-1}\sum_{t=p+1}^n|\breve{\eta}_t|$.

We introduce the following lemma to show Theorem 5 and Remark 3 in a unified framework.

Lemma 8. *Denote $\breve{\boldsymbol{\rho}}=(\breve{\rho}_1,\ldots,\breve{\rho}_M)'$ and $\breve{\boldsymbol{\gamma}}=(\breve{\gamma}_1,\ldots,\breve{\gamma}_M)'$. For any $\breve{\boldsymbol{\theta}}_n\in\Theta$, if the model (1.1) is correctly specified and $\sqrt{n}(\breve{\boldsymbol{\theta}}_n-\boldsymbol{\theta}_0)=-D\sum_{t=1}^n\boldsymbol{G}_t+o_p(1)$ holds, then*

$$\sqrt{n}(\breve{\boldsymbol{\rho}}',\breve{\boldsymbol{\gamma}}')'\to_{\mathcal{L}}N\left(0,VGV'\right) \ \ \textit{as} \ \ n\to\infty,$$

where $G=E(\boldsymbol{v}_t\boldsymbol{v}_t')$ with

$$\boldsymbol{v}_t=\big[(\eta_t-\kappa_1)(\eta_{t-1}-\kappa_1)/\sigma_1^2,\ldots,(\eta_t-\kappa_1)(\eta_{t-M}-\kappa_1)/\sigma_1^2,$$

$$(|\eta_t|-\tau_2)(|\eta_{t-1}|-\tau_2)/\sigma_2^2,\ldots,(|\eta_t|-\tau_2)(|\eta_{t-M}|-\tau_2)/\sigma_2^2,-\boldsymbol{G}_t'D'\big],$$

and $V=\begin{pmatrix} I_M & 0 & U_\rho/\sigma_1^2 \\ 0 & I_M & U_\gamma/\sigma_2^2 \end{pmatrix}$, $U_\rho=(\boldsymbol{U}_{\rho1}',\ldots,\boldsymbol{U}_{\rho M}')'$, $U_\gamma=(\boldsymbol{U}_{\gamma1}',\ldots,\boldsymbol{U}_{\gamma M}')'$, *and* $\boldsymbol{U}_{\rho k}=-\left(E\left[(\eta_{t-k}-\kappa_1)\boldsymbol{Y}_{1t}'\right],\kappa_1 E\left[(\eta_{t-k}-\kappa_1)\boldsymbol{Y}_{2t}'\right]\right)$, $\boldsymbol{U}_{\gamma k}=-\left(\tau_1 E[(|\eta_{t-k}|-\tau_2)\boldsymbol{Y}_{1t}'],\tau_2 E[(|\eta_{t-k}|-\tau_2)\boldsymbol{Y}_{2t}']\right)$ *for* $1\le k\le M$.

Proof of Lemma 8. When model (1.1) is correctly specified, by the ergodic theorem and the dominated convergence theorem, it can be shown that, as $n\to\infty$,

$$\breve{\eta}_1=\frac{1}{n-p}\sum_{t=p+1}^{n}\breve{\eta}_t\to_p\kappa_1 \quad\text{and}\quad \frac{1}{n}\sum_{t=p+1}^{n}(\breve{\eta}_t-\breve{\eta}_1)^2\to_p\sigma_1^2,$$
$$\breve{\eta}_2=\frac{1}{n-p}\sum_{t=p+1}^{n}|\breve{\eta}_t|\to_p\tau_2 \quad\text{and}\quad \frac{1}{n}\sum_{t=p+1}^{n}(|\breve{\eta}_t|-\breve{\eta}_2)^2\to_p\sigma_2^2,$$

Therefore, it follows that

$$\sqrt{n}\left(\breve{\boldsymbol{\rho}}',\breve{\boldsymbol{\gamma}}'\right)'=\sqrt{n}\left(\breve{\boldsymbol{\rho}}',\breve{\boldsymbol{\gamma}}'\right)'+o_p(1), \tag{S3.59}$$

where $\breve{\boldsymbol{\rho}}=(\breve{\rho}_1,\ldots,\breve{\rho}_M)'$ and $\breve{\boldsymbol{\gamma}}=(\breve{\gamma}_1,\ldots,\breve{\gamma}_M)'$ with $\breve{\rho}_k=(n\sigma_1^2)^{-1}\sum_{t=p+k+1}^{n}(\breve{\eta}_t-\kappa_1)(\breve{\eta}_{t-k}-\kappa_1)$ and $\breve{\gamma}_k=(n\sigma_2^2)^{-1}\sum_{t=p+k+1}^{n}(|\breve{\eta}_t|-\tau_2)(|\breve{\eta}_{t-k}|-\tau_2)$. It can be shown that

$$\begin{aligned}\sigma_1^2\sqrt{n}\breve{\rho}_k=&\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}(\eta_t-\kappa_1)(\eta_{t-k}-\kappa_1)+\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}A_{1nt}\\&+\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}A_{2nt}+\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}A_{3nt},\\ \sigma_2^2\sqrt{n}\breve{\gamma}_k=&\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}(|\eta_t|-\tau_2)(|\eta_{t-k}|-\tau_2)+\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}B_{1nt}\\&+\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}B_{2nt}+\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}B_{3nt},\end{aligned}$$

where

$$A_{1nt}=(\breve{\eta}_t-\eta_t)(\eta_{t-k}-\kappa_1),\qquad B_{1nt}=(|\breve{\eta}_t|-|\eta_t|)(|\eta_{t-k}|-\tau_2);$$

$$A_{2nt} = (\eta_t - \kappa_1)\left(\breve{\eta}_{t-k} - \eta_{t-k}\right),\ B_{2nt} = (|\eta_t| - 1)\left(|\breve{\eta}_{t-k}| - |\eta_{t-k}|\right);$$

$$A_{3nt} = (\breve{\eta}_t - \eta_t)\left(\breve{\eta}_{t-k} - \eta_{t-k}\right),\quad B_{3nt} = (|\breve{\eta}_t| - |\eta_t|)\left(|\breve{\eta}_{t-k}| - |\eta_{t-k}|\right).$$

Since $\sqrt{n}(\breve{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_p(1)$, moreover, note that $\partial\eta_t(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta} = (-\boldsymbol{Y}'_{1t}, -\eta_t\boldsymbol{Y}'_{2t})'$ by (S3.39). Then by Taylor's expansion and the ergodic theorem, we have

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n} A_{1nt} =& \frac{1}{n}\sum_{t=p+k+1}^{n}(\eta_{t-k} - \kappa_1)\frac{\partial\eta_t(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}\sqrt{n}(\breve{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1)\\
=& -\frac{1}{n}\sum_{t=p+k+1}^{n}(\eta_{t-k} - \kappa_1)\boldsymbol{Y}'_{1t}\sqrt{n}(\breve{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0)\\
&-\frac{1}{n}\sum_{t=p+k+1}^{n}(\eta_{t-k} - \kappa_1)\eta_t\boldsymbol{Y}'_{2t}\sqrt{n}(\breve{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_0) + o_p(1)\\
=& \boldsymbol{U}_{\rho k}\sqrt{n}(\breve{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1),
\end{aligned} \tag{S3.60}$$

where $\boldsymbol{U}_{\rho k} = -\left(E\left[(\eta_{t-k} - \kappa_1)\boldsymbol{Y}'_{1t}\right], \kappa_1 E\left[(\eta_{t-k} - \kappa_1)\boldsymbol{Y}'_{2t}\right]\right)$. Similarly, we can show that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n} A_{2nt} &= \frac{1}{n}\sum_{t=p+k+1}^{n}(\eta_t - \kappa_1)\sqrt{n}\left[\eta_{t-k}(\breve{\boldsymbol{\theta}}_n) - \eta_{t-k}(\boldsymbol{\theta}_0)\right] = o_p(1),\ \text{and}\\
\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n} A_{3nt} &= \frac{1}{n}\sum_{t=p+k+1}^{n}(\breve{\eta}_t - \eta_t)\sqrt{n}\left[\eta_{t-k}(\breve{\boldsymbol{\theta}}_n) - \eta_{t-k}(\boldsymbol{\theta}_0)\right] = o_p(1).
\end{aligned} \tag{S3.61}$$

As a result, we have

$$\sqrt{n}\breve{\rho}_k = \frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}\frac{(\eta_t - \kappa_1)(\eta_{t-k} - \kappa_1)}{\sigma_1^2} + \frac{\boldsymbol{U}_{\rho k}}{\sigma_1^2}\sqrt{n}(\breve{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1). \tag{S3.62}$$

We next consider $\breve{\gamma}_k$. By the identity (S3.20), we have

$$|\breve{\eta}_t| - |\eta_t| = (\breve{\eta}_t - \eta_t)\,\mathrm{sgn}(\eta_t) + 2\int_0^{-(\breve{\eta}_t - \eta_t)}[I(\eta_t \le s) - I(\eta_t \le 0)]ds.$$

where $\mathrm{sgn}(\eta_t) = I(\eta_t > 0) - I(\eta_t < 0)$. Then similar to the proof of (S3.60), we can show that

$$\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n} B_{1nt} = \boldsymbol{U}_{\gamma k}\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1),$$

where $\boldsymbol{U}_{\gamma k} = -\left(\tau_1 E[(|\eta_{t-k}| - \tau_2)\boldsymbol{Y}'_{1t}], \tau_2 E[(|\eta_{t-k}| - \tau_2)\boldsymbol{Y}'_{2t}]\right)$. Similar to the proof of (S3.61), we can verify that $n^{-1/2}\sum_{t=p+k+1}^{n} B_{2nt} = o_p(1)$ and $n^{-1/2}\sum_{t=p+k+1}^{n} B_{3nt} = o_p(1)$. Therefore, we have

$$\sqrt{n}\check{\gamma}_k = \frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}\frac{(|\eta_t| - \tau_2)(|\eta_{t-k}| - \tau_2)}{\sigma_2^2} + \frac{\boldsymbol{U}_{\gamma k}}{\sigma_2^2}\sqrt{n}(\breve{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_p(1). \tag{S3.63}$$

Combining (S3.59), (S3.62) and (S3.63), by Theorem 2, we can obtain that

$$\sqrt{n}(\breve{\boldsymbol{\rho}}', \breve{\boldsymbol{\gamma}}')' = V\frac{1}{\sqrt{n}}\sum_{t=p+k+1}^{n}\boldsymbol{v}_t + o_p(1). \tag{S3.64}$$

Then by the martingale central limit theorem and the Cramér–Wold device, we have

$$\sqrt{n}(\breve{\boldsymbol{\rho}}', \breve{\boldsymbol{\gamma}}')' \to_{\mathcal{L}} N\left(0, VGV'\right) \text{ as } n \to \infty,$$

where $G = E(\boldsymbol{v}_t\boldsymbol{v}_t')$. Hence, the proof of this theorem is accomplished. □

Proof of Theorem 5. Since $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\Sigma^{-1}/2\sum_{t=1}^{n}\boldsymbol{G}_t$ with $\boldsymbol{G}_t = (\boldsymbol{Y}'_{1t}[I(\eta_t \le 0) - I(\eta_t > 0)], \boldsymbol{Y}'_{2t}(1 - |\eta_t|))'$ by Theorem 2 and $\tau_1 = 0$, $\tau_2 = 1$ by Assumption 3(i), we have established all the conditions for Lemma 8, and hence Theorem 5 is established. □

Proof of Remark 3. Since $\sqrt{n}(\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = -\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{t=p+1}^{n}\frac{\partial \ell_t^G(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + o_p(1)$ by (S3.50) and $\kappa_1 = 0$, $\sigma_1^2 = 1$ by Assumption 4(i), we have established all the conditions for Lemma 8, and hence Remark 3 is verified. □

Bibliography

Amemiya, T. (1985). *Advanced econometrics*. Harvard University Press.

Box, G. E., G. M. Jenkins, G. C. Reinsel, and G. M. Ljung (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Fristedt, B. E. and L. F. Gray (2013). *A modern approach to probability theory.* Springer Science & Business Media.

Huber, P. (1967). The behavior of maximum likelihood estimates under nonstandard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability 1*, 221–233.

Jiang, F., D. Li, and K. Zhu (2020). Non-standard inference for augmented double autoregressive models with null volatility coefficients. *Journal of Econometrics 215*, 165–183.

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

Pollard, D. (1985). New ways to prove central limit theorems. *Econometric Theory 1*, 295–313.

Tsay, R. S. (2005). *Analysis of financial time series.* John wiley & sons.