**Abstract:** This supplementary material includes additional results for the simulation and empirical analysis, as well as technical details for Theorems ??–?? and Remarks ??–??.

To show Theorems ??–??, Theorem ?? and Remark ??, Lemmas ??–?? 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 $\|A\| = \sqrt{\text{tr}(AA')} = \sqrt{\sum_{i,j}|a_{ij}|^2}$.

**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 lag order $M$ and 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 ??, ?? and ??, 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 \( \hat{\theta}_n \) and \( \tilde{\theta}_n \) when the innovations follow the \( t_2 \), \( t_5 \) or \( st_{3,-1.5} \) distribution. The following findings in Section ?? remain unchanged: (1) for the E-QMLE \( \hat{\theta}_n \), its ESDs and ASDs of scale-type estimators \( \hat{\omega}_n \) and \( \hat{\beta}_n \) increase as the distribution of \( \eta_t \) gets more heavy-tailed, while that of location-type estimator \( \hat{\alpha}_n \) are decreasing; (2) in terms of the G-QMLE \( \tilde{\theta}_n \), its ESDs and ASDs increase as the distribution of \( \eta_t \) gets more heavy-tailed. If \( \eta_t \sim t_2 \), then \( E(\eta^2_t) = \infty \) 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: (3) the E-QMLE performs satisfactorily with small biases and ESDs close to ASDs even for \( t_2 \) distribution with \( E(\eta^2_t) = \infty \), which indicates that the E-QMLE is more robust for thick tails than G-QMLE; (4) 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 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 ?? of the manuscript are unchanged. Moreover, we find that nonzero skewness could increase the selection accuracy of BIC while decrease that of
BIC\(^G\). It’s probably because nonzero skewness almost has no effects on ESDs and ASDs of the E-QMLE \(\hat{\theta}_n\), but it increases ESDs and ASDs of the G-QMLE \(\tilde{\theta}_n\).

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 ?? remain unchanged: (1) all sizes are close to the nominal rate except for the \(t_2\) case, and all powers increase as the sample size \(n\) or the departure level increases. This is as expected since the E-QMLE and G-QMLE are not applicable if \(\eta_t\) follows \(t_2\) distribution such that \(E(\eta_t^2) = \infty\), which make the size inaccurate; (2) for the same level of departures, \(Q(6)\) and \(Q^G(6)\) are more powerful in detecting the misspecification in the conditional mean \((c_1 \neq 0, c_2 = 0)\) than that in the conditional standard deviation \((c_1 = 0, c_2 > 0)\); (3) as the innovation distribution gets 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 some 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 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, while \(\{\eta_t\}\) 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 \(p = 1\) to obtain the G-QMLE and E-QMLE. Figure \(\square\)
plots the empirical size of $Q(M)$ and $Q^G(M)$ at 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 $\eta_t$. Therefore, we suggest $M \leq 20$ to avoid undersize situations 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:

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

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

$$(M_1, M_2, \ldots, M_K) = (\lfloor \ln(n) \rfloor, 2\lfloor \ln(n) \rfloor, \ldots, K \lfloor \ln(n) \rfloor),$$

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 ??, 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 5% significance level for $M = 12$ or 18. The following findings in Section ?? in the manuscript remain unchanged: (1) all powers increase 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 detecting the misspecification in the conditional mean ($c_1 \neq 0, c_2 = 0$) than that in the conditional standard deviation ($c_1 = 0, c_2 > 0$); (3) as the innovation distribution gets more heavy-tailed, $Q(M)$ and $Q^G(M)$ perform worse in detecting misspecifications in the conditional standard deviation. In addition, we have the following findings: (4) most of the sizes of $Q(M)$ and $Q^G(M)$ with $M = 12$ or 18 are less than the nominal level; (5) 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 $\tilde{\theta}_n$ and G-QMLE $\tilde{\theta}_n$ when the innovations follow the $t_2$, $t_5$ or $st_{3.15}$ distribution.

<table>
<thead>
<tr>
<th></th>
<th>$t_2$</th>
<th></th>
<th></th>
<th>$t_5$</th>
<th></th>
<th></th>
<th>$st_{3.15}$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>Bias</td>
<td>ESD</td>
<td>ASD</td>
<td>Bias</td>
<td>ESD</td>
<td>ASD</td>
<td>Bias</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.034</td>
<td>0.045</td>
<td>0.047</td>
<td>-0.040</td>
<td>0.059</td>
<td>0.061</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-0.022</td>
<td>0.031</td>
<td>0.033</td>
<td>-0.011</td>
<td>0.042</td>
<td>0.043</td>
<td>-0.004</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>500</td>
<td>0.112</td>
<td>0.138</td>
<td>0.144</td>
<td>0.066</td>
<td>0.078</td>
<td>0.083</td>
<td>-0.072</td>
</tr>
<tr>
<td>$\omega$</td>
<td>1000</td>
<td>0.057</td>
<td>0.098</td>
<td>0.108</td>
<td>0.031</td>
<td>0.055</td>
<td>0.059</td>
<td>-0.075</td>
</tr>
<tr>
<td>$\beta$</td>
<td>500</td>
<td>-0.096</td>
<td>0.115</td>
<td>0.100</td>
<td>-0.060</td>
<td>0.053</td>
<td>0.053</td>
<td>-0.043</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-0.051</td>
<td>0.079</td>
<td>0.076</td>
<td>-0.027</td>
<td>0.038</td>
<td>0.038</td>
<td>-0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$t_2$</th>
<th></th>
<th></th>
<th>$t_5$</th>
<th></th>
<th></th>
<th>$st_{3.15}$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n$</td>
<td>Bias</td>
<td>ESD</td>
<td>ASD</td>
<td>Bias</td>
<td>ESD</td>
<td>ASD</td>
<td>Bias</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-0.056</td>
<td>0.073</td>
<td>0.067</td>
<td>-0.037</td>
<td>0.054</td>
<td>0.054</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-0.023</td>
<td>0.053</td>
<td>0.049</td>
<td>-0.012</td>
<td>0.039</td>
<td>0.038</td>
<td>-0.005</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>500</td>
<td>-0.365</td>
<td>0.249</td>
<td>0.222</td>
<td>0.036</td>
<td>0.091</td>
<td>0.095</td>
<td>-0.016</td>
</tr>
<tr>
<td>$\omega$</td>
<td>1000</td>
<td>-0.312</td>
<td>0.213</td>
<td>0.202</td>
<td>0.020</td>
<td>0.068</td>
<td>0.071</td>
<td>-0.008</td>
</tr>
<tr>
<td>$\beta$</td>
<td>500</td>
<td>0.428</td>
<td>0.581</td>
<td>0.298</td>
<td>-0.072</td>
<td>0.090</td>
<td>0.085</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.479</td>
<td>0.569</td>
<td>0.284</td>
<td>-0.036</td>
<td>0.067</td>
<td>0.063</td>
<td>0.002</td>
</tr>
</tbody>
</table>
Table S.2: Percentages of underfitted, correctly selected and overfitted models by BIC and BIC\textsuperscript{G} when the innovations follow the $t_2$, $t_5$ or $st_{3, -1.5}$ distribution.

<table>
<thead>
<tr>
<th>$n$</th>
<th>$t_2$</th>
<th></th>
<th>$t_5$</th>
<th></th>
<th>$st_{3, -1.5}$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Under</td>
<td>Exact</td>
<td>Over</td>
<td>Under</td>
<td>Exact</td>
<td>Over</td>
</tr>
<tr>
<td>BIC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>7.2</td>
<td>89.7</td>
<td>3.1</td>
<td>30.9</td>
<td>68.9</td>
<td>0.2</td>
</tr>
<tr>
<td>500</td>
<td>0.5</td>
<td>95.8</td>
<td>3.7</td>
<td>7.2</td>
<td>92.7</td>
<td>0.1</td>
</tr>
<tr>
<td>1000</td>
<td>0.0</td>
<td>97.0</td>
<td>3.0</td>
<td>0.2</td>
<td>99.8</td>
<td>0.0</td>
</tr>
<tr>
<td>BIC\textsuperscript{G}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>29.2</td>
<td>48.6</td>
<td>22.2</td>
<td>36.6</td>
<td>61.2</td>
<td>2.2</td>
</tr>
<tr>
<td>500</td>
<td>10.7</td>
<td>59.5</td>
<td>29.8</td>
<td>11.2</td>
<td>87.5</td>
<td>1.3</td>
</tr>
<tr>
<td>1000</td>
<td>1.3</td>
<td>65.3</td>
<td>33.4</td>
<td>0.4</td>
<td>98.0</td>
<td>1.6</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>$c_1$</th>
<th>$c_2$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$t_2$</td>
<td>$t_5$</td>
<td>$st_{3, -1.5}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>500</td>
<td>1000</td>
<td>500</td>
<td>1000</td>
</tr>
<tr>
<td>Q</td>
<td>0.0</td>
<td>0.0</td>
<td>0.077</td>
<td>0.090</td>
<td>0.057</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0</td>
<td>0.253</td>
<td>0.478</td>
<td>0.203</td>
<td>0.475</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0</td>
<td>0.910</td>
<td>0.985</td>
<td>0.996</td>
<td>1.000</td>
</tr>
<tr>
<td>0.0</td>
<td>0.1</td>
<td>0.090</td>
<td>0.152</td>
<td>0.096</td>
<td>0.227</td>
</tr>
<tr>
<td>0.0</td>
<td>0.3</td>
<td>0.284</td>
<td>0.593</td>
<td>0.734</td>
<td>0.978</td>
</tr>
<tr>
<td>Q\textsuperscript{G}</td>
<td>0.0</td>
<td>0.0</td>
<td>0.151</td>
<td>0.168</td>
<td>0.048</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0</td>
<td>0.367</td>
<td>0.607</td>
<td>0.182</td>
<td>0.460</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0</td>
<td>0.937</td>
<td>0.994</td>
<td>0.998</td>
<td>1.000</td>
</tr>
<tr>
<td>0.0</td>
<td>0.1</td>
<td>0.143</td>
<td>0.177</td>
<td>0.064</td>
<td>0.126</td>
</tr>
<tr>
<td>0.0</td>
<td>0.3</td>
<td>0.231</td>
<td>0.367</td>
<td>0.420</td>
<td>0.871</td>
</tr>
</tbody>
</table>
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 $\eta_t$ 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 $\eta_t$ 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 $\eta_t$ 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 $\eta_t$ 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.

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Laplace</th>
<th>$t_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c_1$</td>
<td>$c_2$</td>
<td>500</td>
</tr>
<tr>
<td>$M=12$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Q$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0</td>
<td>0.140</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0</td>
<td>0.964</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.1</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.3</td>
<td>0.712</td>
</tr>
<tr>
<td>$Q^G$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0</td>
<td>0.143</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0</td>
<td>0.981</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.1</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.3</td>
<td>0.506</td>
</tr>
<tr>
<td>$M=18$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Q$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0</td>
<td>0.914</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.1</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.3</td>
<td>0.510</td>
</tr>
<tr>
<td>$Q^G$</td>
<td>0.0</td>
<td>0.0</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0</td>
<td>0.924</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.1</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.3</td>
<td>0.321</td>
</tr>
</tbody>
</table>
S2 Diagnosis for model (5.2)

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

To check ACFs of residuals \( \{\tilde{e}_\eta_t\} \) 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 \( \tilde{\rho}_k \)'s and \( \tilde{\gamma}_k \)'s up to lag 18, and they all fall within their corresponding 95% pointwise confidence intervals (CIs). Clearly almost all residual ACFs are insignificant both individually and jointly at the 5% significance level, and hence we can conclude that the fitted model at (??) is adequate.
Figure 3: Q-Q plots of the residuals \( \{ \tilde{e}_t \} \) against the Student’s \( t_2 \) (left panel), \( t_3 \) (middle panel) or \( t_4 \) (right panel) distribution.

Figure 4: Residual ACF plots for \( \tilde{\rho}_l \) (left panel) and \( \tilde{\gamma}_l \) (right panel), where the red dashed lines are the corresponding 95% pointwise confidence intervals.
S3 Proof of Theorem 1

To show Theorem $??$, we introduce the following lemma.

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

(i) $E \sup_{\theta \in \Theta} |\ell^E_i(\theta)| < \infty$;

(ii) $E [\ell^E_i(\theta)]$ has a unique minimum at $\theta_0$;

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

**Proof.** Recall that $\ell^E_i(\theta) = \ln h_t(\delta) + h_t^{-1}(\delta)|\varepsilon_i(\alpha)|$, where $\varepsilon_i(\alpha) = y_t - \sum_{i=1}^p \alpha_i y_{t-i}$ and $h_t(\delta) = \omega + \sum_{i=1}^p \beta_i |y_{t-i}|$. We first show (i). Denote $c_i = \sup_{\theta \in \Theta} |\alpha_{i0} - \alpha_i|$, then by Assumption $??$ we have $c_i < \infty$ for $i = 1, \ldots, p$. Note that $\varepsilon_i(\alpha) = \sum_{i=1}^p (\alpha_{i0} - \alpha_i) y_{t-i} + \varepsilon_i(\alpha_0)$ and $\varepsilon_i(\alpha_0) = \eta_i h_t(\delta_0) = \eta_i (\omega_0 + \sum_{i=1}^p \beta_{i0} |y_{t-i}|)$. Then by Assumption $??$ and $E(|\eta_i|) = 1$ by Assumption $??(i)$, it follows that

$$
E \sup_{\theta \in \Theta} \left[ \frac{\varepsilon_i(\theta)}{h_t(\delta)} \right] \\
\leq E \sup_{\theta \in \Theta} \left[ \frac{\sum_{i=1}^p |\alpha_{i0} - \alpha_i||y_{t-i}|}{\omega + \sum_{i=1}^p \beta_i |y_{t-i}|} \right] + E(|\eta_i|) E \sup_{\theta \in \Theta} \left[ \frac{\omega_0 + \sum_{i=1}^p \beta_{i0} |y_{t-i}|}{\omega + \sum_{i=1}^p \beta_i |y_{t-i}|} \right] \\
\leq \sum_{i=1}^p E \left[ \frac{c_i |y_{t-i}|}{\beta |y_{t-i}|} \right] + E \left[ \frac{\omega + \sum_{i=1}^p \beta_i |y_{t-i}|}{\omega + \sum_{i=1}^p \beta_i |y_{t-i}|} \right] + \sum_{i=1}^p E \left[ \frac{\beta_{i0} |y_{t-i}|}{\beta |y_{t-i}|} \right] \\
\leq \frac{1}{\beta} \sum_{i=1}^p c_i + \frac{\omega}{\omega} + \frac{1}{\beta} \sum_{i=1}^p \beta_{i0} < \infty. \quad (\text{S3.2})
$$

By (S6.44) and (S3.2), we have

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

Thus, (i) holds.
We next show (ii). Note that \( \varepsilon_t(\alpha) = \varepsilon_t(\alpha_0) - (\alpha - \alpha_0)' Y_1 t h_t(\delta_0) \) and \( \varepsilon_t(\alpha_0) = \eta h_t(\delta_0) \), where \( Y_1 t = h_t^{-1}(\delta_0)(y_{t-1}, \ldots, y_{t-p})' \). Denote \( F_t \) as the \( \sigma \)-field generated by \( \{y_s, s \leq t\} \).

Recall that \( \eta \) has zero median and \( E(|\eta|) = 1 \), and thus \( \min_a E|\eta_a - a| = E|\eta - \text{median}(\eta)| = E|\eta| = 1 \). Then by the law of iterated expectations, we have

\[
E \left[ \ell^E_t(\theta) \right] = E \left[ \ln h_t(\delta) + \frac{|\varepsilon_t(\alpha_0) - (\alpha - \alpha_0)' Y_1 t h_t(\delta_0)|}{h_t(\delta)} \right] = E \left[ \ln h_t(\delta) + \frac{h_t(\delta_0)}{h_t(\delta)} E \left\{ |\eta_t - (\alpha - \alpha_0)' Y_1 t | \big| F_{t-1} \right\} \right] \geq E \left[ \ln h_t(\delta) + \frac{h_t(\delta_0)}{h_t(\delta)} E \left( |\eta| \big| F_{t-1} \right) \right] = E \left[ \ln h_t(\delta) + \frac{h_t(\delta_0)}{h_t(\delta)} \right],
\]

where the minimum of the inequality is attained if and only if \( \alpha = \alpha_0 \) almost surely as \( \eta \) has zero median. Moreover, the function \( f(x) = \ln x + a/x, a \geq 0 \), reaches its minimum at \( x = a \). Therefore, \( E[\ell^E_t(\theta)] \) reaches its minimum if and only if \( h_t(\delta) = h_t(\delta_0) \) almost surely, and hence \( \theta = \theta_0 \). As a result, \( E[\ell^E_t(\theta)] \) is uniformly minimized at \( \theta_0 \), that is (ii) holds.

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

\[
\ell^E_t(\theta) - \ell^E_t(\theta^*) = \ln h_t(\delta) - \ln h_t(\delta^*) + \left[ \frac{\varepsilon_t(\alpha)}{h_t(\delta)} - \frac{\varepsilon_t(\alpha^*)}{h_t(\delta)} \right] + \left[ \frac{\varepsilon_t(\alpha^*)}{h_t(\delta)} - \frac{\varepsilon_t(\alpha^*)}{h_t(\delta^*)} \right].
\]

By Taylor’s expansion, we can see that

\[
\ln h_t(\delta) - \ln h_t(\delta^*) = \frac{(\delta - \delta^*)'(1, |y_{t-1}|, \ldots, |y_{t-p}|)}{h_t(\delta)},
\]

where \( \tilde{\delta} \) lies between \( \delta \) and \( \delta^* \). Then, by Assumption ??, we have

\[
E \left[ \sup_{\theta \in B_\eta(\theta^*)} |\ln h_t(\delta) - \ln h_t(\delta^*)| \right] \leq \eta E \left[ \frac{1 + \sum_{i=1}^p |y_{t-i}|}{\omega + \beta \sum_{i=1}^p |y_{t-i}|} \right] \leq \eta \left( \frac{1}{\omega} + \frac{p}{\beta} \right) \to 0 \text{ as } \eta \to 0.
\]
Similarly, by Taylor’s expansion and (S3.2), it can be verified that

\[
E \left[ \sup_{\theta \in B_{\eta}(\theta^*)} |\epsilon_t(\alpha^*)| \right] \left( \frac{1}{h_t(\delta)} - \frac{1}{h_t(\delta^*)} \right) = E \left[ \sup_{\theta \in B_{\eta}(\theta^*)} |\epsilon_t(\alpha^*)| \right] \left( \frac{1}{h_t(\delta)} - \frac{1}{h_t(\delta^*)} \right) \leq \eta E \left[ \sup_{\theta \in B_{\eta}(\theta^*)} |\epsilon_t(\alpha^*)| \right] \left( \frac{1}{h_t(\delta)} - \frac{1}{h_t(\delta^*)} \right) \leq \eta \left( \frac{1}{\beta} \sum_{i=1}^{p} c_i + \frac{\omega}{\beta} + \frac{1}{\beta} \sum_{i=1}^{p} \beta_{i0} \right) \left( \frac{1}{\omega} + \frac{p}{\beta} \right) \to 0 \text{ as } \eta \to 0,
\]

and

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

Therefore, we have

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

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

Proof of Theorem ?. To establish the strong consistency of \( \hat{\theta}_n \), we follow the method in [Huber, 1967]. Let \( V \) be any open neighborhood of \( \theta_0 \in \Theta \). By Lemma 1 (iii), for any \( \theta^* \in V^c = \Theta/V \) and \( \epsilon > 0 \), there exists an \( \eta_0 > 0 \) such that

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

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

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

Since \( V^c \) is compact, we can choose \( \{B_{\eta_0}(\theta_i) : \theta_i \in V^c, i = 1, 2, ..., k\} \) to be a finite covering of \( V^c \). For each \( \theta_i \in V^c \), from Lemma 1 (ii), there exists an \( \epsilon_0 > 0 \) such that

\[
E \left[ \inf_{\theta \in B_{\eta_0}(\theta_i)} \ell_t^E(\theta) \right] \geq E \left[ \ell_t^E(\theta_0) \right] + 3\epsilon_0. \tag{S3.5}
\]
Thus by (S3.3)–(S3.5), taking $\varepsilon = \varepsilon_0$, if $n$ is large enough, we have

$$
\inf_{\theta \in V^c} L_n^E(\theta) = \min_{1 \leq i \leq k} \inf_{\theta \in B_{\eta_0}(\theta_i)} L_n^E(\theta) \\
\geq \min_{1 \leq i \leq k} \frac{1}{n-p} \sum_{t=p+1}^n \inf_{\theta \in B_{\eta_0}(\theta_i)} \ell_t^E(\theta) \\
\geq \min_{1 \leq i \leq k} E \left[ \inf_{\theta \in B_{\eta_0}(\theta_i)} \ell_t^E(\theta) \right] - \varepsilon \\
\geq E \left[ \ell_t^E(\theta_0) \right] + 2\varepsilon_0. 
$$

(S3.6)

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

$$
\inf_{\theta \in V} L_n^E(\theta) \leq L_n^E(\theta_0) = \frac{1}{n-p} \sum_{t=p+1}^n \ell_t^E(\theta_0) \leq E \left[ \ell_t^E(\theta_0) \right] + \varepsilon_0 \quad \text{a.s..} \quad (S3.7)
$$

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

$$
\inf_{\theta \in V^c} L_n^E(\theta) \geq E \left[ \ell_t^E(\theta_0) \right] + 2\varepsilon_0 > E \left[ \ell_t^E(\theta_0) \right] + \varepsilon_0 \geq \inf_{\theta \in V} L_n^E(\theta) \quad \text{a.s..}
$$

Since $\hat{\theta}_n = \arg \min_{\Theta} L_n^E(\theta)$, if $n$ is large enough, then $\hat{\theta}_n \in V$ almost surely for $\forall V$. By the arbitrariness of $V$, it follows that $\hat{\theta}_n \to \theta_0$ almost surely. The proof is accomplished.

S4 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 $\hat{\theta}_n$.

**Lemma 2.** If Assumptions 2–4 hold, then for any sequence of random variables $u_n$ such that $u_n = o_p(1)$, it follows that

$$
\Pi_{1n}(u_n) = o_p(\sqrt{n} \|u_n\| + n \|u_n\|^2),
$$
where $\Pi_{1n}(u) = 2\sum_{t=p+1}^{n} \int_{0}^{u} K_{tt} \{X_{t}(s) - E[X_{t}(s)|\mathcal{F}_{t-1}]\} \, ds$ with $K_{tt} = (Y_{t}^{\prime}, 0_{p+1})^{\prime}$ and $X_{t}(s) = I(\eta_{t} \leq s) - I(\eta_{t} \leq 0)$.

**Proof.** Note that $\partial \varepsilon_{t}(\alpha_{0})/\partial \theta = (-Y_{1}^{\prime}h_{t}, 0_{p+1})^{\prime} = -h_{t}K_{tt}$. Then it follows that

$$\int_{0}^{u} K_{tt} \{X_{t}(s)\} \, ds = u^{2} K_{tt} \int_{0}^{u} X_{t}(u^{\prime} K_{tt}) \, ds = -\frac{u^{\prime}}{h_{t}} \frac{\partial \varepsilon_{t}(\alpha_{0})}{\partial \theta} M_{t}(u),$$

where $M_{t}(u) = \int_{0}^{u} X_{t}(u^{\prime} K_{tt}) \, ds$. It can be verified that

$$|\Pi_{1n}(u)| \leq 2\|u\|^{2p+1} \sum_{j=1}^{n} \left| \frac{1}{h_{t}} \frac{\partial \varepsilon_{t}(\alpha_{0})}{\partial \theta_{j}} \right| \sum_{t=p+1}^{n} \{M_{t}(u) - E[M_{t}(u)|\mathcal{F}_{t-1}]\},$$

where $\theta_{j}$ is the $j$th element of $\theta$ for $1 \leq j \leq 2p + 1$. Let $m_{t} = h_{t}^{-1}\partial \varepsilon_{t}(\alpha_{0})/\partial \theta_{j}$ and $f_{t}(u) = m_{t}M_{t}(u)$. Define

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

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

$$\sup_{\|u\| \leq \eta} \frac{|D_{n}(u)|}{1 + \sqrt{n}\|u\|} = o_{p}(1). \quad (S4.8)$$

Note that $m_{t} = \max\{m_{t}, 0\} - \max\{-m_{t}, 0\}$. We first prove the case when $m_{t} \geq 0$.

We adopt the method in Lemma 4 of Pollard (1985) to verify (S4.8). Let $\mathcal{F} = \{f_{t}(u) : \|u\| \leq \eta\}$ be a collection of functions indexed by $u$. We first verify that $\mathcal{F}$ satisfies the bracketing condition on page 304 of Pollard (1985). Denote $B_{r}(\zeta)$ as an open neighborhood of $\zeta$ 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 $\{B_{\varepsilon\delta/C_{1}(u_{i})}\}_{i=1}^{K_{\varepsilon}}$ to cover $B_{\delta}(0)$, where $K_{\varepsilon}$ is an integer less than $C_{0}\varepsilon^{-(2p+1)}$ and $C_{0}$ is not depending on $\varepsilon$ and $\delta$. Denote $V_{i}(\delta) = B_{\varepsilon\delta/C_{0}(u_{i})} \cap B_{\delta}(0)$, 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 $B_{\delta}(0)$. For each $u_{i} \in U_{i}(\delta)$ with $1 \leq i \leq K_{\varepsilon}$, we define
the bracketing functions as follows

\[ f_t^\pm(u) = m_t \int_0^1 X_t \left( u' K_{1t} s \pm \frac{\varepsilon \delta}{C_1 h_t} \left\| \frac{\partial \varepsilon_t(\alpha_0)}{\partial \theta} \right\| \right) ds. \]

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

\[ f_t^-(ui) \leq f_t(u) \leq f_t^+(ui). \]

Moreover, by Taylor’s expansion and the fact that \( \partial \varepsilon_t(\alpha_0)/\partial \theta = (-Y_{1t}' h_t, 0_{p+1}')' \), it holds that

\[
E \left[ f_t^+(ui) - f_t^-(ui) \big| \mathcal{F}_{t-1} \right] \leq \frac{2 \varepsilon \delta}{C_1 h_t^2} \left\| \frac{\partial \varepsilon_t(\alpha_0)}{\partial \theta} \right\|^2 \sup_x \bar{f}(x) = \frac{\varepsilon \delta \Delta_t}{C_1}, \tag{S4.9}
\]

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

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

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

Put \( \delta_k = 2^{-k} \eta \). Define \( B(k) = B_{\delta_k}(0) \), 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 \( u \in U_i(\delta_k) \), by (S4.9) with \( \delta = \delta_k \), we have the upper tail

\[
D_n(u) \leq \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \left\{ f_t^+(ui) - E \left[ f_t^-(ui) \big| \mathcal{F}_{t-1} \right] \right\}
\]

\[
= D_n^+(ui) + \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} E \left[ f_t^+(ui) - f_t^-(ui) \big| \mathcal{F}_{t-1} \right]
\]

\[
\leq D_n^+(ui) + \sqrt{n} \varepsilon \delta_k \left( \frac{1}{nC_1} \sum_{t=p+1}^{n} \Delta_t \right),
\]

and the lower tail

\[
D_n(u) \geq \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \left\{ f_t^-(ui) - E \left[ f_t^+(ui) \big| \mathcal{F}_{t-1} \right] \right\}
\]
\[ D_n^-(u_i) = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \left\{ f_t^+(u_i) - E[f_t^+(u_i)|\mathcal{F}_{t-1}] \right\}, \]

where

\[ D_n^+(u_i) = \frac{1}{\sqrt{n}} \sum_{t=p+1}^{n} \left\{ f_t^-(u_i) - E[f_t^-(u_i)|\mathcal{F}_{t-1}] \right\} \]

and

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

Denote the event

\[ E_n = \left\{ \omega : \frac{1}{n C_1} \sum_{t=p+1}^{n} \Delta_t(\omega) < 2 \right\}. \]

On \( E_n \) with \( u \in U_i(\delta_k) \), it follows that

\[ D_n^-(u_i) - 2\sqrt{n}\varepsilon\delta_k \leq D_n(u) \leq D_n^+(u_i) + 2\sqrt{n}\varepsilon\delta_k. \] (S4.10)

For \( u \in A(k) \), i.e. \( \delta_{k+1} \leq \|u\| \leq \delta_k \), we have \( 1 + \sqrt{n}\|u\| > \sqrt{n}\delta_{k+1} = \sqrt{n}\delta_k/2 \). Thus, by the upper tail in (S4.10) and Chebyshev’s inequality, it holds that

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

\[ \leq K_\varepsilon \max_{1 \leq i \leq K_\varepsilon} P(D_n^+(u_i) > \sqrt{n}\varepsilon\delta_k) \]

\[ \leq K_\varepsilon \max_{1 \leq i \leq K_\varepsilon} \frac{E[(D_n^+(u_i))^2]}{n\varepsilon^2\delta_k^2}, \] (S4.11)

and by the lower tail in (S4.10) and Chebyshev’s inequality, we have

\[ P \left( \inf_{u \in A(k)} \frac{D_n(u)}{1 + \sqrt{n}\|u\|} < -6\varepsilon, E_n \right) \leq P \left( \min_{1 \leq i \leq K_\varepsilon} \inf_{u \in U_i(\delta_k) \cap A(k)} D_n(u) < -3\sqrt{n}\varepsilon\delta_k, E_n \right) \]

\[ \leq K_\varepsilon \max_{1 \leq i \leq K_\varepsilon} P(D_n^-(u_i) < -\sqrt{n}\varepsilon\delta_k) \]

\[ \leq K_\varepsilon \max_{1 \leq i \leq K_\varepsilon} \frac{E[(D_n^-(u_i))^2]}{n\varepsilon^2\delta_k^2}. \] (S4.12)
Recall that $\mathbf{K}_{1t} = (\mathbf{Y}_{1t}, 0_{p+1})'$ with $\mathbf{Y}_{1t} = h_t^{-1}(\delta_0)(y_{t-1}, \ldots, y_{t-p})'$ and $h_t(\delta) = \omega + \sum_{i=1}^{p} \beta_i |y_{t-i}|$.

By Assumption ??, it holds that

$$
\|\mathbf{K}_{1t}\| = \left[ \sum_{i=1}^{p} \frac{y_{t-i}}{(\omega_0 + \sum_{i=1}^{p} \beta_0 |y_{t-i}|)^2} \right]^{\frac{1}{2}} \leq \left[ \sum_{i=1}^{p} \frac{\beta_0^2 |y_{t-i}|}{(\beta |y_{t-i}|)^2} \right]^{\frac{1}{2}} \leq \left( \frac{p}{\beta^2} \right)^{\frac{1}{2}} \leq \frac{p}{\beta}. \tag{S4.13}
$$

Then it follows that

$$
|u' \mathbf{K}_{1t}| \leq \|u\| \|\mathbf{K}_{1t}\| \leq \frac{p \delta_k}{\beta} \quad \text{and} \quad |m_t| \leq \frac{1}{h_t} \left\| \frac{\partial \varepsilon_i(\alpha_0)}{\partial \theta} \right\| = \|\mathbf{K}_{1t}\| \leq \frac{p}{\beta}.
$$

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

$$
E[f_t^+(u)] = E \{ E[f_t^+(u)|\mathcal{F}_{t-1}] \} \\
\leq E \left\{ m_t^2 \int_0^1 E \left[ \left| X_t \left( u' \mathbf{K}_{1t} s + \varepsilon \delta_k \frac{\partial \varepsilon_i(\alpha_0)}{\partial \theta} \right) \right| \right] \mathcal{F}_{t-1} \right\} ds \\
\leq \frac{p^2}{\beta^2} E \left( \sup_{|x| \leq C \delta_k} |F(x) - F(0)| \right) \leq C \delta_k \sup_{x} f(x).
$$

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

$$
E \left[ (D_n^+(u_i))^2 \right] = \frac{1}{n} \sum_{t=p+1}^{n} E \left\{ f_t^+(u_i) - E[f_t^+(u_i)|\mathcal{F}_{t-1}] \right\}^2 \\
\leq \frac{1}{n} \sum_{t=p+1}^{n} E \left[ (f_t^+(u_i))^2 \right] \leq C \delta_k \sup_{x} f(x) := \pi_n(\delta_k). \tag{S4.14}
$$

Similarly, we have

$$
E \left[ (D_n^-(u_i))^2 \right] \leq \pi_n(\delta_k). \tag{S4.15}
$$

Combining (S4.11) and (S4.14), we have

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

Combining (S4.12) and (S4.15), we have

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

\[ P \left( \sup_{u \in A(k)} \frac{|D_n(u)|}{1 + \sqrt{n} \|u\|} > 6\varepsilon, E_n \right) \leq \frac{2K \pi_n(\delta_k)}{n\varepsilon^2\delta_k^2} . \] (S4.16)

Since \( \pi_n(\delta_k) \to 0 \) as \( k \to \infty \), we can choose \( k_\varepsilon \) so that \( 2K \pi_n(\delta_k)/(\varepsilon^2\eta^2) < \varepsilon \) 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 \( \{ u : \|u\| \leq \eta \} \) into two sets \( B := B(k_n + 1) \) and \( B^c := B(0) - B(k_n + 1) = \bigcup_{k=0}^{k_n} A(k) \). Then by (S4.16), we have

\[ P \left( \sup_{u \in B_c} \frac{|D_n(u)|}{1 + \sqrt{n} \|u\|} > 6\varepsilon \right) \leq \sum_{k=0}^{k_n} P \left( \sup_{u \in A(k)} \frac{|D_n(u)|}{1 + \sqrt{n} \|u\|} > 6\varepsilon, E_n \right) + P(E_n^c) \]

\[ \leq \frac{1}{n} \sum_{k=0}^{k_\varepsilon - 1} 2K \pi_n(\delta_k)2^{2k} + \frac{\varepsilon}{n} \sum_{k=k_\varepsilon}^{k_n} 2^{2k} + P(E_n^c) \]

\[ \leq O \left( \frac{1}{n} \right) + 4\varepsilon + P(E_n^c) . \] (S4.17)

Since \( 1 + \sqrt{n} \|u\| > 1 \) and \( \sqrt{n}\delta_{k_n+1} < 1 \), similar to the proof of (S4.11) and (S4.14), we have

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

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

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

\[ \leq \frac{2K_\varepsilon \pi_n(\delta_{k_n+1})}{\varepsilon^2} + P(E_n^c) . \] (S4.18)

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, (S4.8) follows by (S4.17) and (S4.18). We can verify (S4.8) similarly for the case of \( m_t < 0 \). This completes the proof of this lemma. \( \square \)

**Lemma 3.** Suppose that Assumptions ??, ?? and ?? hold, then for any sequence of random variables \( u_n \) such that \( u_n = o_p(1) \), it follows that

\[ (n - p)[L_n(\theta_0 + u_n) - L_n(\theta_0)] = \sqrt{n}u_n^T \Sigma + (\sqrt{n}u_n)' \Sigma(\sqrt{n}u_n) + o_p(\sqrt{n} \|u_n\| + n\|u_n\|^2) , \]
where \( T_n = T_{1n} + T_{2n} \) with \( T_{1n} = n^{-1/2} \sum_{t=p+1}^{n} K_{1t} [I(\eta_t < 0) - I(\eta_t > 0)] \) and \( T_{2n} = n^{-1/2} \sum_{t=p+1}^{n} K_{2t} (1 - |\eta_t|) \), and \( \Sigma = \text{diag} \{ f(0)E(Y_{1t}'Y_{1t}), E(Y_{2t}'Y_{2t})/2 \} \). Moreover,

\[
T_n \rightarrow_{p} N(0, \Omega) \quad \text{as} \quad n \rightarrow \infty,
\]

where

\[
\Omega = \begin{pmatrix}
E(Y_{1t}'Y_{1t}') & \kappa_1 E(Y_{1t}'Y_{2t}') \\
\kappa_1 E(Y_{2t}'Y_{1t}') & \kappa_2 E(Y_{2t}'Y_{2t}')
\end{pmatrix}
\]

with \( \kappa_1 = E(\eta_t) \) and \( \kappa_2 = E(\eta_t^2) - 1 \).

**Proof.** We first re-parameterize the objective function as

\[
H_n(u) = (n - p) \left[ L_n^E(\theta_0 + u) - L_n^E(\theta_0) \right],
\]

where \( u \in \Lambda = \{ u = (u_1', u_2') : u + \theta_0 \in \Theta \} \). Let \( \hat{u}_n = \hat{\theta}_n - \theta_0 \). Then, \( \hat{u}_n \) is the minimizer of \( H_n(u) \) in \( \Lambda \). Furthermore, we have

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

where

\[
A_t(u) = \frac{1}{h_t(\delta_0)} \left[ |\varepsilon_t(\alpha_0 + u_1)| - |\varepsilon_t(\alpha_0)| \right],
\]

\[
B_t(u) = \ln h_t(\delta_0 + u_2) - \ln h_t(\delta_0) + \frac{|\varepsilon_t(\alpha_0)|}{h_t(\delta_0 + u_2)} - \frac{|\varepsilon_t(\alpha_0)|}{h_t(\delta_0)},
\]

\[
C_t(u) = \left[ \frac{1}{h_t(\delta_0 + u_2)} - \frac{1}{h_t(\delta_0)} \right] \left[ |\varepsilon_t(\alpha_0 + u_1)| - |\varepsilon_t(\alpha_0)| \right].
\]

Recall that \( K_{1t} = (Y_{1t}', 0'_{p+1})' \). By Taylor’s expansion, we can show that

\[
\frac{\varepsilon_t(\alpha_0 + u_1)}{h_t(\delta_0)} - \frac{\varepsilon_t(\alpha_0)}{h_t(\delta_0)} = -\frac{u_1'(y_{t-1}, \ldots, y_{t-p})'}{h_t(\delta_0)} = -u'K_{1t}.
\] (S4.19)

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

\[
|x - y| - |x| = -y[I(x > 0) - I(x < 0)] + 2 \int_{0}^{y} [I(x \leq s) - I(x \leq 0)] ds,
\] (S4.20)
together with (S4.19), we have
\[
|\varepsilon_t(\alpha_0 + u_1)| - |\varepsilon_t(\alpha_0)| = - u' K_{1t} h_t[I(\eta_t > 0) - I(\eta_t < 0)] \\
+ 2 h_t \int_0^t u' K_{1t} X_t(s) ds.
\] (S4.21)

Therefore, it follows that
\[
A_t(u) = u' K_{1t}[I(\eta_t < 0) - I(\eta_t > 0)] + 2 \int_0^t u' K_{1t} X_t(s) ds,
\]
where \(X_t(s) = I(\eta_t \leq s) - I(\eta_t \leq 0)\). Then we have
\[
\sum_{t=p+1}^n A_t(u) = \sqrt{n} u'^T \mathbf{T}_{1n} + \Pi_{1n}(u) + \Pi_{2n}(u),
\] (S4.22)

where
\[
T_{1n} = \frac{1}{\sqrt{n}} \sum_{t=p+1}^n K_{1t}[I(\eta_t < 0) - I(\eta_t > 0)],
\]
\[
\Pi_{1n}(u) = 2 \sum_{t=p+1}^n \int_0^t u' K_{1t} \{X_t(s) - E[X_t(s)|\mathcal{F}_{t-1}]\} ds,
\]
\[
\Pi_{2n}(u) = 2 \sum_{t=p+1}^n \int_0^t u' K_{1t} E[X_t(s)|\mathcal{F}_{t-1}] ds.
\]

For \(\Pi_{1n}(u_n)\) with \(u_n = o_p(1)\), by Lemma 2, it holds that
\[
\Pi_{1n}(u_n) = o_p(\sqrt{n}\|u_n\| + n\|u_n\|^2).
\] (S4.23)

For \(\Pi_{2n}(u)\), by iterated-expectation and Taylor’s expansion, we have
\[
\Pi_{2n}(u) = 2 \sum_{t=p+1}^n \int_0^t u' K_{1t} E[I(\eta_t \leq s) - I(\eta_t \leq 0)|\mathcal{F}_{t-1}] ds
\]
\[
= 2 \sum_{t=p+1}^n \int_0^t u' K_{1t} [F(s) - F(0)] ds
\]
\[
= 2 \sum_{t=p+1}^n \int_0^t u' K_{1t} s f(0) ds + 2 \sum_{t=p+1}^n \int_0^t u' K_{1t} s [f(s^*) - f(0)] ds
\]
\[
= (\sqrt{n} u')' [K_{1n} + K_{2n}(u)] (\sqrt{n} u),
\] (S4.24)
where \( s^* \) lies between 0 and \( s \),

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

By the ergodic theorem, it holds that

\[
K_{1n} = f(0)E[K_{1t}K_{1t}'] + o_p(1) = \Sigma + o_p(1), \quad (S4.25)
\]

where \( \Sigma = \text{diag} \{ f(0)E(Y_{1t}Y_{1t}'), 0\times (p+1) \} \).

Since \( \|K_{1t}\| \leq p/\beta \) by (S4.13) and \( \|u\| \leq \eta \), we have \( |u'K_{1t}| \leq p\eta/\beta \).

Furthermore, it follows that

\[
\lim_{\eta \to 0} \sup_{0 \leq s \leq p\eta/\beta} |f(s) - f(0)| = 0 \quad \text{as} \quad \sup_x f(x) < \infty \quad \text{by Assumption ??(ii)}.
\]

Then for any \( \eta > 0 \), it holds that

\[
\sup_{\|u\| < \eta} |K_{2n}(u)| \leq \sup_{\|u\| < \eta} \frac{2}{\eta \|u\|^2} \sum_{t=p+1}^n \left| u'K_{1t} \right| s[f(s^*) - f(0)]ds
\]

\[
\leq \frac{1}{\eta \|u\|^2} \sum_{t=p+1}^n \|u\|^2 \|K_{1t}\|^2 \left[ \sup_{0 \leq s \leq p\eta/\beta} |f(s) - f(0)| \right]
\]

\[
\leq \frac{p^2}{\eta \beta^2} \sum_{t=p+1}^n \left[ \sup_{0 \leq s \leq p\eta/\beta} |f(s) - f(0)| \right]
\]

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

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

\[
P \left( \sup_{\|u\| \leq \eta_0} |K_{2n}(u)| > \delta \right) < \frac{\varepsilon}{2}, \quad (S4.26)
\]

for all \( n \geq 1 \). On the other hand, since \( u_n = o_p(1) \), it follows that

\[
P(\|u_n\| > \eta_0) < \frac{\varepsilon}{2}, \quad (S4.27)
\]

as \( n \) is large enough. By (S4.26) and (S4.27), for \( \forall \varepsilon, \delta > 0 \), we have

\[
P \left( |K_{2n}(u_n)| > \delta \right) = P(\|K_{2n}(u_n)\| > \delta, \|u_n\| \leq \eta_0) + P(\|K_{2n}(u_n)\| > \delta, \|u_n\| > \eta_0)
\]

\[
= P(\|u_n\| > \eta_0) + P(\|K_{2n}(u_n)\| > \delta, \|u_n\| > \eta_0)
\]

\[
< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\]
\[ \leq P \left( \sup_{\|u\| \leq \eta_0} |K_2n(u)| > \delta \right) + P(\|u_n\| > \eta_0) < \varepsilon \]

as \( n \) is large enough. Hence, it holds that \( K_{2n}(u_n) = o_p(1) \). Furthermore, combining (S4.24) and (S4.25), we can show that

\[ \Pi_{2n}(u_n) = (\sqrt{n}u_n)' \Sigma_{21}(\sqrt{n}u_n) + o_p(n\|u_n\|^2). \tag{S4.28} \]

As a result, combining (S4.22), (S4.23) and (S4.28), we have

\[ \sum_{t=\eta+1}^{n} A_t(u_n) = \sqrt{n}u'_n T_{1n} + (\sqrt{n}u_n)' \Sigma_{21}(\sqrt{n}u_n) + o_p(\sqrt{n}\|u_n\| + n\|u_n\|^2). \tag{S4.29} \]

We next consider \( B_t(u) \). Note that \( \partial h_t(\delta_0)/\partial \delta = h_t Y_{2t} \) and \( \varepsilon_t(\alpha_0) = h_t \eta_t \). By Taylor’s expansion, we can show that

\[ B_t(u) = \ln h_t(\delta_0 + u_2) + \frac{|\varepsilon_t(\alpha_0)|}{h_t(\delta_0 + u_2)} - \left( \ln h_t(\delta_0) + \frac{|\varepsilon_t(\alpha_0)|}{h_t(\delta_0)} \right) \]

\[ = u'_2 Y_{2t} - |\eta_t| u'_2 Y_{2t} - \frac{u'_2}{2} \frac{\partial h_t(\delta)}{\partial \delta} - \frac{1}{2} \frac{1}{h^2_t(\delta)} \frac{\partial h_t(\delta)}{\partial \theta} \frac{\partial h_t(\delta)}{\partial \theta'} u_2 \]

\[ = u' K_{2t}(1 - |\eta_t|) + u' \left( \frac{|\varepsilon_t(\alpha_0)|}{h_t(\delta)} - \frac{1}{2} \frac{1}{h^2_t(\delta)} \frac{\partial h_t(\delta)}{\partial \theta} \frac{\partial h_t(\delta)}{\partial \theta'} u_2 \right), \]

where \( \tilde{\delta} \) lies between \( \delta_0 \) and \( \delta_0 + u_2 \), and \( K_{2t} = (0', Y_{2t}'). \) Then, we have

\[ \sum_{t=\eta+1}^{n} B_t(u) = \sqrt{n}u' T_{2n} + \Pi_{3n}(u), \tag{S4.30} \]

where

\[ T_{2n} = \frac{1}{\sqrt{n}} \sum_{t=\eta+1}^{n} K_{2t}(1 - |\eta_t|) \] and \( \Pi_{3n}(u) = u' \sum_{t=\eta+1}^{n} \left( \frac{|\varepsilon_t(\alpha_0)|}{h_t(\delta)} - \frac{1}{2} \frac{1}{h^2_t(\delta)} \frac{\partial h_t(\delta)}{\partial \theta} \frac{\partial h_t(\delta)}{\partial \theta'} u_2 \right). \]

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

\[ J_t(\tilde{\delta}) = \left( \frac{|\varepsilon_t(\alpha_0)|}{h_t(\delta)} - \frac{1}{2} \frac{1}{h^2_t(\delta)} \frac{\partial h_t(\delta)}{\partial \theta} \frac{\partial h_t(\delta)}{\partial \theta'} \right). \]

By the ergodic theorem, it is easy to see that

\[ \frac{1}{n} \sum_{t=\eta+1}^{n} J_t(\tilde{\delta}) = E[J_t(\tilde{\delta})] + o_p(1). \]
Since $\bar{\delta} \to_p \delta_0$, by the dominated convergence theorem, we have

$$\lim_{n \to \infty} E[J_t(\bar{\delta})] = E[J_t(\theta_0)] = \Sigma_{22},$$

where $\Sigma_{22} = \text{diag} \{0_{p \times p}, E(Y_{2t}Y_{2t}')/2\}$. Thus, for $u_n = o_p(1)$, we can get that

$$\Pi_{3n}(u_n) = (\sqrt{n}u_n)'\Sigma(\sqrt{n}u_n) + o_p(n\|u_n\|^2). \quad (S4.31)$$

Therefore, $(S4.30)$ and $(S4.31)$ imply that

$$\sum_{t=p+1}^n B_t(u_n) = \sqrt{n}u_n'T_{2n} + (\sqrt{n}u_n)'\Sigma_{22}(\sqrt{n}u_n) + o_p(n\|u_n\|^2). \quad (S4.32)$$

Finally, we consider $C_t(u)$. By Taylor’s expansion, we have

$$\frac{1}{h_t(\delta_0 + u_2)} - \frac{1}{h_t(\delta_0)} = -\frac{u'}{h_t(\delta_0)} \frac{\partial h_t(\bar{\delta})}{\partial \theta},$$

where $\bar{\delta}$ lies between $\delta_0$ and $\delta_0 + u_2$. This together with $(S4.21)$, implies that

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

where

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

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

Since $\eta_t$ has zero median by Assumption ??(i), then by the ergodic theorem, we have

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

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

$$\sum_{t=p+1}^n C_t(u_n) = o_p(n\|u_n\|^2) \quad (S4.33)$$
Combining \((S4.29), (S4.32)\) and \((S4.33)\), we have
\[
(n-p)[L_n(\theta_0 + u_n) - L_n(\theta_0)] = \sqrt{n}u_n' \mathbf{T}_n + (\sqrt{n}u_n)' \Sigma(\sqrt{n}u_n) + o_p(\sqrt{n}\|u_n\| + n\|u_n\|^2),
\]
where \(\mathbf{T}_n = \mathbf{T}_{1n} + \mathbf{T}_{2n}\) and \(\Sigma = \Sigma_{21} + \Sigma_{22} = \text{diag}\{f(0)E(Y_{1t}Y'_{1t}), E(Y_{2t}Y'_{2t})/2\} \).

Moreover, let \(\mathbf{G}_t = (Y'_{1t}[I(\theta_t < 0) - I(\theta_t > 0)], Y'_{2t}(1 - |\theta_t|))'\), we have that \(\mathbf{T}_n = \mathbf{T}_{1n} + \mathbf{T}_{2n} = \sum_{t=p+1}^n \mathbf{G}_t\). If Assumptions \(\text{??, ?? and ??}\) hold, by the Central Limit Theorem, we have
\[
\mathbf{T}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega) \quad \text{as} \quad n \to \infty, \quad (S4.34)
\]
where \(\Omega\) is defined as in Lemma \(\text{??}\). We accomplish the proof of this lemma.

**Proof of Theorem \(??\).** We have \(\hat{u}_n = \hat{\theta}_n - \theta_0 = o_p(1)\) by Theorem \(??\). Furthermore, by Lemma \(??\), we have
\[
H_n(\hat{u}_n) = \sqrt{n}u_n' \mathbf{T}_n + (\sqrt{n}u_n)' \Sigma(\sqrt{n}u_n) + o_p(\sqrt{n}\|u_n\| + n\|u_n\|^2)
\]
\[
\geq -\sqrt{n}\|u_n\| [\|\mathbf{T}_n\| + o_p(1)] + n\|u_n\|^2[\lambda_{\text{min}} + o_p(1)],
\]
where \(\lambda_{\text{min}} > 0\) is the minimum eigenvalue of \(\Sigma\). Note that \(H_n(\hat{u}_n) = (n-p)[L_n^E(\hat{\theta}_n) - L_n^E(\theta_0)] \leq 0\). Then it follows that
\[
n\|\hat{u}_n\| \leq [\lambda_{\text{min}} + o_p(1)]^{-1} [\|\mathbf{T}_n\| + o_p(1)] = O_p(1). \quad (S4.36)
\]
This together with Theorem \(??\), verifies the \(\sqrt{n}\)-consistency. Hence, Statement (i) holds.

Next, let \(\sqrt{n}u_n^* = -\Sigma^{-1}\mathbf{T}_n/2\), then, by \((S4.34)\) from Lemma \(??\), we have
\[
\sqrt{n}u_n^* \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{1}{4}\Sigma^{-1}\Omega\Sigma^{-1}\right) \quad \text{as} \quad n \to \infty.
\]
As a result, it is sufficient to show that \(\sqrt{n}\hat{u}_n - \sqrt{n}u_n^* = o_p(1)\). By \((S4.35)\) and \((S4.36)\), we have
\[
H_n(\hat{u}_n) = (\sqrt{n}\hat{u}_n)' \mathbf{T}_n + (\sqrt{n}\hat{u}_n)' \Sigma(\sqrt{n}\hat{u}_n) + o_p(1)
\]
\begin{equation}
-2 \left( \sqrt{n} \mathbf{u}_n \right)' \Sigma \left( \sqrt{n} \mathbf{u}_n \right) + \left( \sqrt{n} \mathbf{u}_n \right)' \Sigma \left( \sqrt{n} \mathbf{u}_n \right) + o_p(1).
\end{equation}

Note that (S4.35) still holds when \( \mathbf{u}_n \) is replaced by \( \mathbf{u}_n^* \). Thus,

\begin{equation}
H_n(\mathbf{u}_n^*) = \left( \sqrt{n} \mathbf{u}_n^* \right)' \mathbf{T}_n + \left( \sqrt{n} \mathbf{u}_n^* \right)' \Sigma \left( \sqrt{n} \mathbf{u}_n^* \right) + o_p(1)
\end{equation}

By the previous two equations, it follows that

\begin{equation}
H_n(\mathbf{u}_n) - H_n(\mathbf{u}_n^*) = \left( \sqrt{n} \mathbf{u}_n - \sqrt{n} \mathbf{u}_n^* \right)' \Sigma \left( \sqrt{n} \mathbf{u}_n - \sqrt{n} \mathbf{u}_n^* \right) + o_p(1)
\end{equation}

\begin{equation}
\geq \lambda_{\min} \left\| \sqrt{n} \mathbf{u}_n - \sqrt{n} \mathbf{u}_n^* \right\|^2 + o_p(1).
\end{equation}

(S4.37)

Since \( H_n(\mathbf{u}_n) - H_n(\mathbf{u}_n^*) = (n-p) \left[ L_n^E(\theta_0 + \mathbf{u}_n) - L_n^E(\theta_0 + \mathbf{u}_n^*) \right] \leq 0 \) almost surely, by (S4.37) we have \( \| \sqrt{n} \mathbf{u}_n - \sqrt{n} \mathbf{u}_n^* \| = o_p(1) \). Therefore, we have

\begin{equation}
\sqrt{n} \mathbf{u}_n \to_L N \left( 0, \frac{1}{4} \Sigma^{-1} \Omega \Sigma^{-1} \right) \text{ as } n \to \infty.
\end{equation}

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

\section*{S5 Proof of Remark 1}

\textbf{Proof.} Recall that the error function is \( \eta_t(\theta) = \varepsilon_t(\alpha)/h_t(\delta) \), where \( \varepsilon_t(\alpha) = y_t - \sum_{i=1}^p \alpha_i y_{t-i} \) and \( h_t(\delta) = \omega + \sum_{i=1}^p \beta_i |y_{t-i}| \), and the residuals are defined as \( \hat{\eta}_t = \eta_t(\hat{\theta}_n) = \varepsilon_t(\hat{\alpha}_n)/h_t(\hat{\delta}_n) \). Note that \( \eta_t = \eta_t(\theta_0) \) and \( \hat{f}(0) = (nb_n)^{-1} \sum_{t=p+1}^n K(\hat{\eta}_t/b_n) \). Since \( |K(x) - K(y)| \leq L|x - y| \) for some \( L > 0 \), by Taylor’s expansion, it holds that

\begin{equation}
\left| \hat{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{\hat{\eta}_t}{b_n} \right) - K \left( \frac{\eta_t}{b_n} \right) \right| \leq \frac{L}{nb_n^2} \sum_{t=p+1}^n |\hat{\eta}_t - \eta_t|.
\end{equation}
where $\theta^*$ lies between $\theta_0$ and $\hat{\theta}_n$. It can be verified that

$$\frac{\partial \eta_t(\theta)}{\partial \theta} = \frac{1}{h_t(\delta)} \frac{\partial \varepsilon_t(\alpha)}{\partial \theta} - \varepsilon_t(\alpha) \frac{\partial h_t(\delta)}{\partial \theta} = - \frac{h_t}{h_t(\delta)} K_{1t} - \frac{h_t}{h_t(\delta)} \frac{\varepsilon_t(\alpha)}{\partial \theta} = K_{2t},$$

where $K_{1t} = (Y_{1t}', 0_{p+1}')'$ and $K_{2t} = (0_p', Y_{2t}')'$. Since $\|K_{1t}\| \leq p/\omega$ by (S4.13), we can show that $\|K_{2t}\| \leq \omega^{-1} + p/\omega$, and $\sup_{\theta \in \Theta} h_t/\|h_t(\delta)\| \leq \omega_0/\omega + \sum_{i=1}^p \beta_i/\beta$ by Assumption (S3.2).

These together with (S3.2), imply that

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

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

$$\frac{1}{n} \sum_{t=p+1}^n \left\| \frac{\partial \eta_t(\theta^*)}{\partial \theta} \right\| = E \left\| \frac{\partial \eta_t(\theta^*)}{\partial \theta} \right\| + O_p(1) = E \left\| \frac{\partial \eta_t(\theta_0)}{\partial \theta} \right\| + O_p(1) = O_p(1).$$

Since $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$ by Theorem (S3.2) and $nb_n^4 \to \infty$ as $n \to \infty$, by (S5.38) and (S5.40), we have

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

Moreover, since $\sup_{x} f(x) < \infty$ by Assumption (S3.2)(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

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

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

$$\left| E \left[ \frac{1}{b_n} K \left( \frac{\eta_t}{b_n} \right) \right] - f(0) \right| = \int_{-\infty}^{\infty} K(x) [f(b_n x) - f(0)] dx.$$
\begin{equation}
\leq b_n \sup_x |f'(x)| \int_{-\infty}^{\infty} |x| K(x) dx \to 0 \text{ as } n \to \infty. \tag{S5.43}
\end{equation}

By (S5.41)–(S5.43), we have \( \hat{f}(0) = f(0) + o_p(1) \). Finally, by a similar argument as for (S5.40), we can show that \((n-p)^{-1} \sum_{t=p+1}^n Y_{1t} Y_{1t}' = E(Y_{1t} Y_{1t}') + o_p(1), (n-p)^{-1} \sum_{t=p+1}^n Y_{2t} Y_{2t}' = E(Y_{2t} Y_{2t}') + o_p(1), (n-p)^{-1} \sum_{t=p+1}^n Y_{1t} Y_{2t}' = E(Y_{1t} Y_{2t}') + o_p(1) \), \( \kappa_1 = (n-p)^{-1} \sum_{t=p+1}^n \kappa_t = E(\kappa_t) + o_p(1) \) and \( \kappa_2 = (n-p)^{-1} \sum_{t=p+1}^n \kappa^2_t - 1 = E(\kappa_t^2) - 1 + o_p(1) \). As a result, \( \hat{\Sigma}_n \to_p \Sigma \) and \( \hat{\Omega}_n \to_p \Omega \) as \( n \to \infty \). The proof is accomplished.

\section*{S6 Proof of Theorem 3}

To show Theorem 3, we introduce the following Lemmas

\textbf{Lemma 4.} If Assumptions ?? hold, then it holds that

\begin{enumerate}[(i)]
\item \( E \sup_{\theta \in \Theta} |\ell^G_t(\theta)| < \infty \);
\item \( E \sup_{\theta \in \Theta} \left\| \frac{\partial \ell^G_t(\theta)}{\partial \theta} \right\| < \infty \);
\item \( E \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \ell^G_t(\theta)}{\partial \theta \partial \theta'} \right\| < \infty \).
\end{enumerate}

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

\begin{align*}
\frac{\partial \ell^G_t(\theta)}{\partial \alpha_i} &= \frac{\varepsilon_t(\alpha) y_{t-i}}{h_t^2(\delta)} \left( 1 - \frac{\varepsilon_t^2(\alpha)}{h_t^2(\delta)} \right), \\
\frac{\partial \ell^G_t(\theta)}{\partial \delta} &= \frac{1}{h_t(\delta)} \left( 1 - \frac{\varepsilon_t^2(\alpha)}{h_t^2(\delta)} \right), \\
\frac{\partial^2 \ell^G_t(\theta)}{\partial \alpha_i \partial \alpha_j} &= \frac{2 \varepsilon_t(\alpha) y_{t-i}}{h_t^2(\delta)} \left( 1 - \frac{3 \varepsilon_t^2(\alpha)}{h_t^2(\delta)} \right), \\
\frac{\partial^2 \ell^G_t(\theta)}{\partial \alpha_i \partial \delta} &= \frac{2 \varepsilon_t(\alpha) y_{t-i}}{h_t^2(\delta)} \left( 1 - \frac{3 \varepsilon_t^2(\alpha)}{h_t^2(\delta)} \right), \\
\frac{\partial^2 \ell^G_t(\theta)}{\partial \delta \partial \beta_j} &= \frac{2 \varepsilon_t(\alpha) y_{t-i} y_{t-j}}{h_t^2(\delta)} \left( 1 - \frac{3 \varepsilon_t^2(\alpha)}{h_t^2(\delta)} \right) \text{ and}
\end{align*}

\begin{align*}
\frac{\partial^2 \ell^G_t(\theta)}{\partial \beta_i \partial \beta_j} &= \frac{2 \varepsilon_t(\alpha) y_{t-i} y_{t-j}}{h_t^2(\delta)} \left( 1 - \frac{3 \varepsilon_t^2(\alpha)}{h_t^2(\delta)} \right) .
\end{align*}

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

\begin{equation}
E \ln \left( \overline{\omega}^* + \overline{\beta} \sum_{i=1}^p |y_{t-i}| \right) = \frac{1}{\kappa} \ln \left( \overline{\omega}^* + \overline{\beta} \sum_{i=1}^p |y_{t-i}| \right) .
\end{equation}
Thus, (i) is verified. Similarly, we can show that (ii) and (iii) hold.

By Assumption ??, implies that

\[
E \sup_{\theta \in \Theta} |\ln h_i(\delta)| \leq E \sup_{\theta \in \Theta} \left[ I \left( \omega + \beta \sum_{i=1}^{P} |y_{t-i}| \geq 1 \right) \ln \left( \omega + \beta \sum_{i=1}^{P} |y_{t-i}| \right) \right] + E \sup_{\theta \in \Theta} \left[ -I \left( \omega + \beta \sum_{i=1}^{P} |y_{t-i}| \leq 1 \right) \ln \left( \omega + \beta \sum_{i=1}^{P} |y_{t-i}| \right) \right] 
\leq E \ln \left( \omega^* + \beta \sum_{i=1}^{P} |y_{t-i}| \right) - I\{\omega < 1\} \ln \omega < \infty. 
\] (S6.44)

By Assumption ??(i), we have \(E(\eta_t) = 0\) and \(E(\eta_t^2) = 1\). Moreover, note that \(\eta_t\) is independent of \(\mathcal{F}_{t-1}\), and \(y_t - \sum_{i=1}^{P} \alpha_i y_{t-i} = \sum_{i=1}^{P} (\alpha_i \eta_0 - \alpha_i) y_{t-i} + \eta_t (\omega_0 + \sum_{i=1}^{P} \beta_i |y_{t-i}|)\), then by Assumption ?? and \(c_r\) inequality, it can be verified that

\[
E \sup_{\theta \in \Theta} \left[ \frac{(y_t - \sum_{i=1}^{P} \alpha_i y_{t-i})^2}{(\omega + \sum_{i=1}^{P} \beta_i |y_{t-i}|)^2} \right] \leq E \sup_{\theta \in \Theta} \left[ \left( \sum_{i=1}^{P} \frac{(\alpha_i \eta_0 - \alpha_i) y_{t-i}}{\omega + \beta \sum_{i=1}^{P} |y_{t-i}|} \right)^2 \right] + E \left[ \left( \frac{\omega + \beta \sum_{i=1}^{P} |y_{t-i}|}{\omega + \beta \sum_{i=1}^{P} |y_{t-i}|} \right)^2 \right] 
\leq E \sup_{\theta \in \Theta} \left[ \left( \sum_{i=1}^{P} \frac{\alpha_i \eta_0 - \alpha_i}{\beta |y_{t-i}|} \right)^2 \right] + 2E \left[ \left( \frac{\omega}{\omega + \beta \sum_{i=1}^{P} |y_{t-i}|} \right)^2 \right] + 2E \left[ \left( \sum_{i=1}^{P} \frac{\beta |y_{t-i}|}{\omega + \beta \sum_{i=1}^{P} |y_{t-i}|} \right)^2 \right] 
\leq \frac{P}{\beta^2} \sum_{i=1}^{P} \sup_{\theta \in \Theta} (\alpha_i \eta_0 - \alpha_i)^2 + \frac{2\omega^2}{\omega^2} + \frac{2(\omega^2 + \beta^2)}{\beta^2} < \infty. 
\] (S6.45)

By (S6.44), (S6.45) and the triangle inequality, we have

\[
E \sup_{\theta \in \Theta} |\ell_{G}(\theta)| \leq E \sup_{\theta \in \Theta} |\ln h_i(\delta)| + \frac{1}{2} E \sup_{\theta \in \Theta} \frac{\epsilon^2_i(\alpha)}{h_i^2(\delta)} < \infty. 
\]

Thus, (i) is verified. Similarly, we can show that (ii) and (iii) hold. \(\square\)
Lemma 5. If Assumptions ?? and ?? hold, then

(i) \( \sup_{\theta \in \Theta} |L_n^G(\theta) - E[\ell_t^G(\theta)]| = o_p(1); \)

(ii) \( \sup_{\theta \in \Theta} \left\| \frac{\partial L_n^G(\theta)}{\partial \theta} - E \left[ \frac{\partial \ell_t^G(\theta)}{\partial \theta} \right] \right\| = o_p(1); \)

(iii) \( \sup_{\theta \in \Theta} \left\| \frac{\partial^2 L_n^G(\theta)}{\partial \theta \partial \theta'} - E \left[ \frac{\partial^2 \ell_t^G(\theta)}{\partial \theta \partial \theta'} \right] \right\| = o_p(1). \)

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

Lemma 6. If Assumptions ??, ?? and ??(i) hold, then \( E[\ell_t^G(\theta)] \) has a unique minimum at \( \theta_0. \)

Proof. We first prove that

\[ c_1 = 0_p \text{ if } c_1' Y_{1t} = 0 \text{ a.s.} \quad \text{and} \quad c_2 = 0_{p+1} \text{ if } c_2' Y_{2t} = 0 \text{ a.s.}, \]

(S6.46)

where \( c_1 \) and \( c_2 \) are \( p \times 1 \) and \( (p + 1) \times 1 \) constant vectors, respectively. If \( c_1' Y_{1t} = 0 \text{ a.s.} \) and \( c_1 = (c_1, \ldots, c_p)' \neq 0, \) without loss of generality, we can assume \( c_1 = 1, \) thus \( y_t = -\sum_{i=2}^{p} c_i y_{t-i+1} \text{ a.s.} \) Recall that \( \eta_t = \varepsilon_t(\alpha_0)/h_t(\delta_0) \) with \( \varepsilon_t(\alpha_0), h_t(\delta_0) \in \mathcal{F}_{t-1}, \) and \( \eta_t \) is independent of \( \mathcal{F}_{t-1}, \) we have

\[ E(\eta_t^2) = E(\eta_t)E \left( \frac{-\sum_{i=2}^{p} c_i y_{t-i+1} - \varepsilon_t(\alpha_0)}{h_t(\delta_0)} \right) = 0, \]

(S6.47)

which is a contradiction with \( E(\eta_t^2) = 1 \) by Assumption ??(i), thus \( c_1 = 0_p. \) Similarly, we can show that \( c_2 = 0_{p+1}. \)

Note that \( \varepsilon_t(\alpha) = \varepsilon_t(\alpha_0) - (\alpha - \alpha_0)' Y_{1t} h_t(\delta_0) \) and \( \varepsilon_t(\alpha_0) = \eta_t h_t(\delta_0). \) As for (S6.45), then by the law of iterated expectations and Assumption ??(i), we can show that

\[ E[\ell_t^G(\theta)] = E \left\{ \ln h_t(\delta) + \frac{\left[ \varepsilon_t(\alpha_0) - (\alpha_0)' Y_{1t} h_t(\delta_0) \right]^2}{2 h_t^2(\delta)} \right\} \]

\[ = E \left[ \ln h_t(\delta) + \frac{1}{2} \left( \frac{h_t(\delta_0)}{h_t(\delta)} \right)^2 \right] - \frac{1}{2} E \left[ \frac{(\alpha - \alpha_0)' Y_{1t} h_t(\delta_0)}{h_t(\delta)} \right]^2. \]

(S6.48)
The second term in (S6.48) reaches its minimum at zero, and this happens if and only if 
\((\alpha - \alpha_0)Y_{1t} = 0\) a.s., which holds if and only if \(\alpha = \alpha_0\) by (S6.46). For the first term in (S6.48), denote \(f(x) = -\ln(x) - 0.5a^2/x^2\), where \(x = h_t(\delta)\) and \(a = h_t(\delta_0)\). We can prove that \(f(x)\) reaches its minimum at \(x = a\), i.e. \(h_t(\delta) = h_t(\delta_0)\), which holds if and only if \(\delta = \delta_0\) by (S6.46). Therefore, \(E[\ell_i^G(\theta)]\) is uniquely minimized at \(\theta_0\).

\[\text{Lemma 7. Suppose Assumptions ??, ?? and ?? hold, then} \]

(i) \(\Omega_1\) and \(\Sigma_1\) are finite and positive definite;

(ii) \(\sqrt{n}\partial L_n^G(\theta_0)/\partial \theta \rightarrow_L N(0, \Omega_1)\) as \(n \rightarrow \infty\).

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

\[
\Sigma_1 = \text{diag} \{E(Y_{1t}Y'_{1t}), 2E(Y_{2t}Y'_{2t})\} \text{ and } \Omega_1 = \begin{pmatrix}
E(Y_{1t}Y'_{1t}) & \kappa_3 E(Y_{1t}Y'_{2t}) \\
\kappa_3 E(Y_{2t}Y'_{1t}) & \kappa_4 E(Y_{2t}Y'_{2t})
\end{pmatrix}.
\]

By Assumptions ??-??, for some constant \(C\), we have

\[
||E(Y_{1t}Y'_{1t})|| < C, \quad ||E(Y_{1t}Y'_{2t})|| < C \quad \text{and} \quad ||E(Y_{2t}Y'_{2t})|| < C.
\]

Thus, \(\Sigma_1\) is finite. Since \(E(\eta_t^4) < \infty\) by Assumption ??(ii), we have \(\kappa_1, \kappa_2 < \infty\), then \(\Omega_1\) is also finite.

Let \(\mathbf{x} = (x_1', x_2')'\), where \(x_1 \in \mathbb{R}^p\) and \(x_2 \in \mathbb{R}^{p+1}\) are arbitrary non-zero constant vectors. It follows that

\[
x'\Omega_1 x = E \left\{ (x_1'Y_{1t})^2 + \kappa_4(x_2'Y_{2t})^2 + 2\kappa_3 x_2'Y_{2t}Y'_{1t}x_1 \right\} = E \left\{ (x_1'Y_{1t} + \kappa_3 x_2'Y_{2t})^2 + (\kappa_4 - \kappa_3^2)(x_2'Y_{2t})^2 \right\}. \tag{S6.49}
\]

By Cauchy-Schwarz inequality and Assumption ??(i), we have \(\kappa_3^2 = [\text{cov}(\eta_t, \eta_t^2)]^2 \leq \text{var}(\eta_t) \times \text{var}(\eta_t^2) = \kappa_4\), 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 \( \mathbf{x}'\Omega_1\mathbf{x} > 0 \), i.e. \( \Omega_1 \) is positive definite. Moreover, by (S6.46), it can be verified that

\[
\mathbf{x}'\Sigma_1\mathbf{x} = E \left\{ \left( \mathbf{x}_1'\mathbf{Y}_1\right)^2 + 2\left( \mathbf{x}_2'\mathbf{Y}_2\right)^2 \right\} > 0.
\]

As a result, \( \Sigma_1 \) is positive definite. Hence, (i) holds.

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

**Proof of Theorem 4**.

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 \( \mathbf{\tilde{\theta}}_n \to_p \mathbf{\theta}_0 \) as \( n \to \infty \).

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

\[
\sqrt{n}(\mathbf{\tilde{\theta}}_n - \mathbf{\theta}_0) = -\Sigma_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=p+1}^{n} \frac{\partial \ell_t^G(\mathbf{\theta}_0)}{\partial \mathbf{\theta}} + o_p(1).
\]

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.

**Proof of Theorem 9**.

In the following proof, notations \( \Theta^p, \mathbf{\theta}_0^p, E, \mathbf{\tilde{\theta}}_n^p, \mathbf{T}_n^p \), and \( \Sigma^p \) are employed to emphasize their dependence on the order \( p \). Let \( p_0 \) be the true order of model (9). It is sufficient to show that, for any \( p \neq p_0 \),

\[
\lim_{n \to \infty} P(\text{BIC}(p) - \text{BIC}(p_0) > 0) = 1.
\]

We first consider the case that the model is overfitted, i.e. \( p > p_0 \). Since the model with order \( p \) corresponds to a larger model, then it holds that

\[
\ell_t^E(\mathbf{\theta}_0^p) = \ell_t^E(\mathbf{\theta}_0^{p_0}), \quad L_n^E(\mathbf{\theta}_0^p) = L_n^E(\mathbf{\theta}_0^{p_0}).
\]
Denote $\hat{\mathbf{u}}_n^{p_0} = \hat{\theta}_n^{p_0} - \theta_0^{p_0}$.

Similar to the proof of Theorems ?? and ??, we can show that $\hat{\theta}_n^{p_0} \to \theta_0^{p_0}$ almost surely and $\sqrt{n} \hat{\mathbf{u}}_n^{p_0} = O_p(1)$ as $n \to \infty$. Then it follows that

$$\sqrt{n} \hat{\mathbf{u}}_n^{p_0} ((n - p_{\max})[L_n^E(\hat{\theta}_n^{p_0}) - L_n^E(\theta_0^{p_0})]$$

$$= (\sqrt{n} \hat{\mathbf{u}}_n^{p_0})^T T_n^{p_0} + (\sqrt{n} \hat{\mathbf{u}}_n^{p_0})^T \Sigma_n^{p_0} (\sqrt{n} \hat{\mathbf{u}}_n^{p_0}) + o_p(1) = O_p(1), \quad (S7.52)$$

where $T_n^{p_0}$ is defined as in Lemma 8. Similarly, it can be verified that $(n - p_{\max})[L_n^E(\hat{\theta}_n^{p}) - L_n^E(\theta_0^{p})] = O_p(1)$. Therefore,

$$L_n^E(\hat{\theta}_n^{p}) - L_n^E(\hat{\theta}_n^{p_0}) = \left[ L_n^E(\hat{\theta}_n^{p_0}) - L_n^E(\theta_0^{p_0}) \right] - \left[ L_n^E(\hat{\theta}_n^{p_0}) - L_n^E(\theta_0^{p_0}) \right] + [L_n^E(\theta_0^{p}) - L_n^E(\theta_0^{p_0})] = O_p \left( \frac{1}{n} \right).$$

Hence, it follows that

$$\text{BIC}^E(p) - \text{BIC}^E(p_0)$$

$$= 2(n - p_{\max})[L_n^E(\hat{\theta}_n^{p}_0) - L_n^E(\hat{\theta}_n^{p_0})] + [(2p + 1) \ln(n - p_{\max}) - (2p_0 + 1) \ln(n - p_{\max})]$$

$$= O_p(1) + 2(p - p_0) \ln(n - p_{\max}) \to \infty \text{ as } n \to \infty. \quad (S7.53)$$

We next consider the case that the model is underfitted, i.e. $p < p_0$. Let $\theta_{0,E} = \arg \min_{\theta \in \Theta^p} E[\ell_t^E(\theta^p)]$. Similar to the proof of Theorems ??--?? and (S7.52), we can verify that $\sqrt{n}(\hat{\theta}_n - \theta_{0,E}) = O_p(1)$ and

$$(n - p_{\max})[L_n^E(\hat{\theta}_n) - L_n^E(\theta_{0,E}^{p})] = O_p(1).$$

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 $E[\ell_t^E(\theta_{0,E}^{p})] \geq E[\ell_t^E(\theta_0^{p_0})] + \varepsilon$ for some positive constant $\varepsilon$. By ergodic theorem, we have $L_n^E(\theta_{0,E}^{p}) = E[\ell_t^E(\theta_{0,E}^{p})] + o_p(1)$. Thus it holds that

$$L_n^E(\theta_{0,E}^{p}) - L_n^E(\theta_0^{p_0}) = E[\ell_t^E(\theta_{0,E}^{p})] - E[\ell_t^E(\theta_0^{p_0})] + o_p(1) = \varepsilon + o_p(1).$$
Hence, we have
\[
L_n^E(\hat{\theta}_n^p) - L_n^E(\hat{\theta}_n^{p_0}) = [L_n^E(\hat{\theta}_n^p) - L_n^E(\theta_0^{p_0})] + [L_n^G(\hat{\theta}_n^p) - L_n^G(\theta_0^{p_0})]
\]
\[
= [L_n^E(\theta_0^{p,E}) - L_n^E(\theta_0^{p_0})] = O_p\left(\frac{1}{n}\right) + \varepsilon + o_p(1).
\]
This together with \((2p + 1)\ln(n - p_{\text{max}}) - (2p_0 + 1)\ln(n - p_{\text{max}}) = O(\ln n)\), implies that
\[
BIC^E(p) - BIC^E(p_0)
\]
\[
= 2(n - p_{\text{max}})[L_n^E(\hat{\theta}_n^p) - L_n^E(\hat{\theta}_n^{p_0})] + [(2p + 1)\ln(n - p_{\text{max}}) - (2p_0 + 1)\ln(n - p_{\text{max}})]
\]
\[
= 2(n - p_{\text{max}})\varepsilon + o_p(n - p_{\text{max}}) + O_p(1) + O(\ln n) \to \infty \text{ as } n \to \infty.
\] (S7.54)

Hence, combines (S7.53) and (S7.54) implies that (S7.51) holds. The proof is accomplished.

**S8 Proof of Remark 2**

**Proof.** In the following proof, notations \(\Theta^p, \theta_0^p, \tilde{\theta}_n^p, \Sigma_1^p\) are employed to emphasize their dependence on the order \(p\). Let \(p_0\) be the true order of model (??). It is sufficient to show that, for any \(p \neq p_0\),
\[
\lim_{n \to \infty} P(BIC^G(p) - BIC^G(p_0) > 0) = 1.
\] (S8.55)

We first consider the case that the model is overfitted, i.e. \(p > p_0\). Since the model with order \(p\) corresponds to a larger model, then it holds that
\[
\ell_t^G(\theta_0^p) = \ell_t^G(\theta_0^{p_0}) \text{ and } L_n^G(\theta_0^p) = L_n^G(\theta_0^{p_0}).
\]
Denote \(\tilde{\theta}_n^{p_0} = \hat{\theta}_n^{p_0} - \theta_0^{p_0}\). Similar to the proof of Theorem ??, we can verify that \(\tilde{\theta}_n^{p_0} \to_p \theta_0^{p_0}\), \(\sqrt{n}\partial L_n^G(\theta_0^{p_0})/\partial \theta = O_p(1)\) and \(\sqrt{n}\tilde{u}_n^{p_0} = O_p(1)\) as \(n \to \infty\). By Taylor’s expansion and Slutsky’s theorem, it can be shown that
\[
(n - p_{\text{max}})[L_n^G(\tilde{\theta}_n^p) - L_n^G(\theta_0^{p_0})]
\]
Hence, we have

\[
\begin{align*}
&= (\sqrt{n} \widehat{u}_n^{(p)})^T \nabla \ell_n^{(\theta^0)} - \frac{\partial L_n^{(\theta^0)}}{\partial \theta} - (\sqrt{n} \widehat{u}_n^{(p)})^T \Sigma_n^{(p)} \sqrt{n} \widehat{u}_n^{(p)} + o_p(1) = O_p(1).
\end{align*}
\] (S8.56)

Similarly, it can be verified that \((n - p_{\max})[L_n^{(\tilde{\theta}_n^{p})} - L_n^{(\theta_0^{p})}] = O_p(1)\). As a result,

\[
L_n^{(\tilde{\theta}_n^{p})} - L_n^{(\theta_0^{p})} = [L_n^{(\tilde{\theta}_n^{p})} - L_n^{(\theta_0^{p})}] - [L_n^{(\tilde{\theta}_n^{p})} - L_n^{(\theta_0^{p})}] + [L_n^{(\theta_0^{p})} - L_n^{(\theta_0^{p})}] = O_p\left(\frac{1}{n}\right).
\]

Hence, we have

\[
\text{BIC}^{(p)} - \text{BIC}^{(p_0)} = 2(n - p_{\max})[L_n^{(\tilde{\theta}_n^{p})} - L_n^{(\theta_0^{p})}] + [(2p + 1) \ln(n - p_{\max}) - (2p_0 + 1) \ln(n - p_{\max})]\n
= O_p(1) + 2(p - p_0) \ln(n - p_{\max}) \to \infty \text{ as } n \to \infty.
\] (S8.57)

We next consider the case that the model is underfitted, i.e. \(p < p_0\). Let \(\theta_{0,G}^{p} = \arg\min_{\theta \in \Theta} E[\ell_t^{(\theta)}] \) and \(\theta_{0,E}^{p} = \arg\min_{\theta \in \Theta} E[\ell_t^{(\theta)}]\). Similar to the proof of Theorem ?? and (S8.56), we can verify that \(\sqrt{n}(\tilde{\theta}_n^{p} - \theta_{0,G}^{p}) = O_p(1)\) and

\[
(n - p_{\max})[L_n^{(\tilde{\theta}_n^{p})} - L_n^{(\theta_{0,G}^{p})}] = O_p(1).
\]

Note that \(L_n^{(\cdot)}\) is the negative likelihood. Since the model with order \(p\) corresponds to a smaller model than the true model, we have \(E[\ell_t^{(\theta_{0,G}^{p})}] \geq E[\ell_t^{(\theta_0^{p})}] + \varepsilon\) for some positive constant \(\varepsilon\). By ergodic theorem, we have \(L_n^{(\theta_{0,G}^{p})} = E[\ell_t^{(\theta_{0,G}^{p})}] + o_p(1)\). Thus it holds that

\[
L_n^{(\theta_{0,G}^{p})} - L_n^{(\theta_0^{p})} = E[\ell_t^{(\theta_{0,G}^{p})}] - E[\ell_t^{(\theta_0^{p})}] + o_p(1) = \varepsilon + o_p(1).
\]

Therefore, we have

\[
L_n^{(\tilde{\theta}_n^{p})} - L_n^{(\theta_0^{p})} = [L_n^{(\tilde{\theta}_n^{p})} - L_n^{(\theta_0^{p})}] - [L_n^{(\tilde{\theta}_n^{p})} - L_n^{(\theta_0^{p})}] + [L_n^{(\theta_{0,G}^{p})} - L_n^{(\theta_0^{p})}] = O_p\left(\frac{1}{n}\right) + \varepsilon + o_p(1).
\]
This together with \((2p + 1) \ln(n - p_{\text{max}}) - (2p_0 + 1) \ln(n - p_{\text{max}}) = O(\ln n)\), implies that

\[
\text{BIC}^G(p) - \text{BIC}^G(p_0) = 2(n - p_{\text{max}}) [L_n^G(\tilde{\theta}_n^p) - L_n^G(\tilde{\theta}_n^{p_0})] + [(2p + 1) \ln(n - p_{\text{max}}) - (2p_0 + 1) \ln(n - p_{\text{max}})]
\]

\[
= 2(n - p_{\text{max}}) \varepsilon + o_p(n - p_{\text{max}}) + O_p(1) + O(\ln n) \rightarrow \infty \quad \text{as} \quad n \rightarrow \infty.
\]  

\(\text{(S8.58)}\)

\[\square\]

S9 Proof of Theorem 5 and Remark 3

Recall that the error function is defined as \(\eta_t(\theta) = \varepsilon_t(\alpha)/h(\delta)\), \(\kappa_1 = E(\eta_t)\), \(\tau_1 = E[\text{sgn}(\eta_t)]\), \(\tau_2 = E(|\eta_t|)\), \(\sigma_1^2 = \text{var}(\eta_t)\) and \(\sigma_2^2 = \text{var}(|\eta_t|)\). Note that \(\eta_t = \eta_t(\theta_0)\), \(\tilde{\eta}_t = \eta_t(\tilde{\theta}_n)\) and \(\hat{\eta}_t = \eta_t(\hat{\theta}_n)\). Let \(\tilde{\theta}_n \in \Theta\) be a \(\sqrt{n}\)-estimator of \(\theta_0\) and have the asymptotic property

\(\sqrt{n}(\tilde{\theta}_n - \theta_0) = -D \sum_{t=1}^n G_t + o_p(1)\), and the residual ACF and absolute residual ACF of model (\(\theta_0\)) fitted by \(\tilde{\theta}_n\) can be defined as

\[
\hat{\rho}_k = \frac{\sum_{t=p+k+1}^n (\hat{\eta}_t - \hat{\eta}_1)(\hat{\eta}_{t-k} - \hat{\eta}_1)}{\sum_{t=p+1}^n (\hat{\eta}_t - \hat{\eta}_1)^2}
\]

\[
\hat{\gamma}_k = \frac{\sum_{t=p+k+1}^n (|\hat{\eta}_t| - \hat{\eta}_2)(|\hat{\eta}_{t-k}| - \hat{\eta}_2)}{\sum_{t=p+1}^n (|\hat{\eta}_t| - \hat{\eta}_2)^2},
\]

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

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

**Lemma 8.** Denote \(\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_M)'\) and \(\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_M)'\). For any \(\tilde{\theta}_n \in \Theta\), if the model (\(\theta_0\)) is correctly specified and \(\sqrt{n}(\tilde{\theta}_n - \theta_0) = -D \sum_{t=1}^n G_t + o_p(1)\) holds, then

\(\sqrt{n}(\hat{\rho}', \hat{\gamma}') \rightarrow \mathcal{N}(0, GV')\)

as \(n \rightarrow \infty\),

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

\[
\mathbf{v}_t = [(\eta_t - \kappa_1)(\eta_{t-1} - \kappa_1)/\sigma_1^2, \ldots, (\eta_t - \kappa_1)(\eta_{t-M} - \kappa_1)/\sigma_1^2],
\]
Therefore, it follows that

\[
(|\eta_{t}| - \tau_2)(|\eta_{t-1}| - \tau_2)/\sigma_2^2, \ldots, (|\eta_{t}| - \tau_2)(|\eta_{t-M}| - \tau_2)/\sigma_2^2, -\mathcal{G}_t'D',
\]

and \(V = \begin{pmatrix} I_M & 0 \\ 0 & \sigma_1^2 \end{pmatrix}, U_{\rho} = (U_{\rho1}', \ldots, U_{\rho M}', U_{\gamma} = (U_{\gamma1}', \ldots, U_{\gamma M}'), \) and \(U_{\rho k} = (\eta_{t} - \rho_{k})^2 \rightarrow_p \sigma_1^2,\)

\[
-(E[(\eta_{t-k} - \kappa_1)Y_{1t}], \kappa_1 E[(\eta_{t-k} - \kappa_1)Y_{2t}]), U_{\gamma k} = -(\tau_1 E[(|\eta_{t-k}| - \tau_2)Y_{1t}], \tau_2 E[(|\eta_{t-k}| - \tau_2)Y_{2t}])
\]

for \(1 \leq k \leq M.\)

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

\[
\begin{align*}
\bar{\eta}_1 &= \frac{1}{n-p} \sum_{t=p+1}^{n} \bar{\eta}_t \rightarrow_p \kappa_1 \quad \text{and} \quad \frac{1}{n} \sum_{t=p+1}^{n} (\bar{\eta}_t - \bar{\eta}_1)^2 \rightarrow_p \sigma_1^2, \\
\bar{\eta}_2 &= \frac{1}{n-p} \sum_{t=p+1}^{n} |\bar{\eta}_t| \rightarrow_p \tau_2 \quad \text{and} \quad \frac{1}{n} \sum_{t=p+1}^{n} (|\bar{\eta}_t| - \bar{\eta}_2)^2 \rightarrow_p \sigma_2^2,
\end{align*}
\]

Therefore, it follows that

\[
\sqrt{n} (\hat{\rho}', \hat{\gamma}')' = \sqrt{n} (\hat{\rho}', \hat{\gamma}')' + o_p(1), \quad (S9.59)
\]

where \(\hat{\rho} = (\hat{\rho}_1, \ldots, \hat{\rho}_M)'\) and \(\hat{\gamma} = (\hat{\gamma}_1, \ldots, \hat{\gamma}_M)'\) with \(\hat{\rho}_k = (n\sigma_1^2)^{-1} \sum_{t=p+k+1}^{n} (\bar{\eta}_t - \kappa_1)(\bar{\eta}_{t-k} - \kappa_1)\)

and \(\hat{\gamma}_k = (n\sigma_2^2)^{-1} \sum_{t=p+k+1}^{n} (|\bar{\eta}_t| - \tau_2)(|\bar{\eta}_{t-k}| - \tau_2).\) It can be shown that

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

where

\[
A_{1nt} = (\bar{\eta}_t - \eta_t)(\eta_{t-k} - \kappa_1), \quad B_{1nt} = (|\bar{\eta}_t| - |\eta_t|)(|\eta_{t-k}| - \tau_2); 
\]
\[ A_{2nt} = (\eta_t - \kappa_1)(\bar{\eta}_{t-k} - \eta_{t-k}), \quad B_{2nt} = (|\eta_t| - 1)(|\bar{\eta}_{t-k}| - |\eta_{t-k}|); \]

\[ A_{3nt} = (\bar{\eta}_t - \eta_t)(\bar{\eta}_{t-k} - \eta_{t-k}), \quad B_{3nt} = (|\bar{\eta}_t| - |\eta_t|)(|\bar{\eta}_{t-k}| - |\eta_{t-k}|). \]

Since \( \sqrt{n}(\dot{\theta}_n - \theta_0) = O_p(1) \), moreover, note that \( \vartheta_n(\theta_0)/\vartheta = (-Y_{1t}, -\eta Y'_{2t})' \) by (S9.60). Then by Taylor’s expansion and the ergodic theorem, we have

\[
\frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} A_{1nt} = \frac{1}{n} \sum_{t=p+k+1}^{n} (\eta_{t-k} - \kappa_1) \frac{\partial \eta_t(\theta_0)}{\partial \vartheta} \sqrt{n}(\dot{\theta}_n - \theta_0) + o_p(1) \\
= -\frac{1}{n} \sum_{t=p+k+1}^{n} (\eta_{t-k} - \kappa_1) Y'_{1t} \sqrt{n}(\dot{\alpha}_n - \alpha_0) \\
- \frac{1}{n} \sum_{t=p+k+1}^{n} (\eta_{t-k} - \kappa_1) \eta_t Y'_{2t} \sqrt{n}(\dot{\delta}_n - \delta_0) + o_p(1) \\
= U_{\rho k} \sqrt{n}(\dot{\theta}_n - \theta_0) + o_p(1), \quad (S9.60)
\]

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

\[
\frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} A_{2nt} = \frac{1}{n} \sum_{t=p+k+1}^{n} (\eta_t - \kappa_1) \sqrt{n} \left[ \eta_{t-k}(\dot{\theta}_n) - \eta_{t-k}(\theta_0) \right] = o_p(1), \quad \text{and} \\
\frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} A_{3nt} = \frac{1}{n} \sum_{t=p+k+1}^{n} (\bar{\eta}_t - \eta_t) \sqrt{n} \left[ \eta_{t-k}(\dot{\theta}_n) - \eta_{t-k}(\theta_0) \right] = o_p(1). \quad (S9.61)
\]

As a result, we have

\[
\sqrt{n}\hat{\rho}_k = \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} \frac{(\eta_t - \kappa_1)(\eta_{t-k} - \kappa_1)}{\sigma_1^2} + \frac{U_{\rho k}}{\sigma_1^2} \sqrt{n}(\dot{\theta}_n - \theta_0) + o_p(1). \quad (S9.62)
\]

We next consider \( \hat{\gamma}_k \). By the identity (S4.20), we have

\[
|\bar{\eta}_t| - |\eta_t| = (\bar{\eta}_t - \eta_t) \text{sgn}(\eta_t) + 2 \int_{0}^{(\bar{\eta}_t - \eta_t)} [I(\eta_t \leq s) - I(\eta_t \leq 0)] ds.
\]

where \( \text{sgn}(\eta_t) = I(\eta_t > 0) - I(\eta_t < 0) \). Then similar to the proof of (S9.60), we can show that

\[
\frac{1}{\sqrt{n}} \sum_{t=p+k+1}^{n} B_{1nt} = U_{\gamma k} \sqrt{n}(\dot{\theta}_n - \theta_0) + o_p(1),
\]
where $U_{\gamma k} = -(\tau_1 E[(|\eta_{t-k}| - \tau_2)Y'_1], \tau_2 E[(|\eta_{t-k}| - \tau_2)Y'_2])$. Similar to the proof of (S9.61), 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

$$\sqrt{n} \gamma_k = \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^n \frac{(|\eta_t| - \tau_2)(|\eta_{t-k}| - \tau_2)}{\sigma_2^2} + \frac{U_{\gamma k}}{\sigma_2^2} \sqrt{n} \gamma_k + o_p(1). \quad (S9.63)$$

Combining (S9.59), (S9.62) and (S9.63), by Theorem ??, we can obtain that

$$\sqrt{n} \gamma_k = V \frac{1}{\sqrt{n}} \sum_{t=p+k+1}^n v_t + o_p(1). \quad (S9.64)$$

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

$$\sqrt{n} \gamma_k \rightarrow_L N(0, VG') \quad as \quad n \rightarrow \infty,$$

where $G = E(v_t v'_t)$. Hence, the proof of this theorem is accomplished.

**Proof of Theorem ??**. Since $\sqrt{n} \gamma_k = -\Sigma_{-1/2} \sum_{t=1}^n G_t$ with $G_t = (Y'_1[I(\eta_t \leq 0) - I(\eta_t > 0)], Y'_2(1 - |\eta_t|))'$ by Theorem ?? and $\tau_1 = 0, \tau_2 = 1$ by Assumption ??(i), we have established all the conditions for Lemma 8, and hence Theorem ?? is followed.

**Proof of Remark ??**. Since $\sqrt{n} \gamma_k = -\Sigma_{-1/2} \sum_{t=1}^n \frac{\partial G_t(\theta_0)}{\partial \theta} + o_p(1)$ by (S6.50) and $\kappa_1 = 0, \sigma_1^2 = 1$ by Assumption ??(i), we have established all the conditions for Lemma 8, and hence Remark ?? is followed.

**Bibliography**


