CONSISTENT ORDER SELECTION FOR ARFIMA PROCESSES

BY HSUEH-HAN HUANG1, NGAI HANG CHAN2, KUN CHEN3, AND CHING-KANG ING1

1National Tsing Hua University
2Chinese University of Hong Kong
3Southwestern University of Finance and Economics

Estimating the orders of the autoregressive fractionally integrated moving average (ARFIMA) model has been a long-standing challenge in time series analysis. This paper tackles the challenge by establishing the consistency of the Bayesian information criterion (BIC) for ARFIMA models with independent errors. Since the memory parameter of the model can be any real number, this consistency result is valid for short-memory, long-memory, and non-stationary time series. This paper further extends the consistency of the BIC to ARFIMA models with conditional heteroscedastic errors, thereby extending its applications to encompass many real-life situations. Finite-sample implications of the theoretical results are illustrated via numerical examples.

1. Introduction. Model selection has always been one of the most important problems in statistical analysis. A correctly specified model not only fulfills the principle of parsimony, but also provides efficient prediction of future values thereby achieving the ultimate goal of model building. In the time series context, such a goal is manifested through the issue of order selection of a particular class of time series models. In particular, the class of autoregressive fractionally integrated moving average (ARFIMA) models has played a critical role over the past several decades because of its applicability in disciplines as diverse as economics, finance, hydrology, telecommunications, network engineering, and environmental sciences. For a comprehensive discussion about ARFIMA models, see the seminal monograph of Beran (1994). One of the key challenges in using ARFIMA models is consistently estimating its AR order \( p_0 \) and MA order \( q_0 \). As aforementioned, a correct model with the fewest number of parameters enhances both estimation efficiency and prediction accuracy. Due to the non-identifiability problem of over-parameterized candidates, the consistency issue of order selection has been only partially resolved, however. This issue becomes even more complicated when dealing with short-memory, long-memory, and non-stationary time series simultaneously, in which case the memory parameter \( d_0 \) in an ARFIMA model is allowed to take any (unknown) real value between \(-\infty\) and \(\infty\).

In the special case when \( d_0 = 0 \) is known, an ARFIMA model reduces to a stationary ARMA model; the problem of estimating \( p_0 \) and \( q_0 \) was pursued by Hannan (1980), Hannan and Rissanen (1982), and Hannan and Kavalieris (1984), among many others. These authors showed that \( p_0 \) and \( q_0 \) can be consistently estimated by means of the Bayesian information criterion (BIC) or its variants. If \( d_0 \) is known to be a positive integer, then an ARFIMA model becomes a well-known ARIMA model, which further simplifies to an ARI model when \( q_0 \) is zero. When the integration component of the model is treated as part of the AR component, it can be shown that the BIC-type criterion still boasts consistency in estimating \( p_0 + d_0 \) for ARI models; see Tsay (1984), Paulsen (1984), Wei (1992), and Ing, Sin and Yu (2012). Likewise, estimating \((p_0 + d_0, q_0)\) for ARIMA models, see Guo, Chen and Zhang (1989) and Huang and Guo (1990). Arguably, Beran, Bhansali and Ocker (1998) established the first result on order selection consistency without assuming \( d_0 \) to be an integer. When \( q_0 \) is zero, these authors further proved that \( p_0 \) can be consistently estimated using BIC, provided \( d_0 \) is
an unknown real number satisfying
\begin{equation}
\frac{\epsilon_0}{d_0} \geq -0.5, \quad d_0 \notin \{-0.5, 0.5, 1.5, 2.5, \ldots\}.
\end{equation}

Their results, however, preclude ARFIMA models with non-trivial MA components (that is, \(q_0 \neq 0\)). In addition, the condition (1.1) for \(d_0\) seems restrictive in practice. In fact, owing to identifiability problems when \(q_0 > 0\) (see, e.g., Hannan (1980)), consistent estimates of \((p_0, q_0)\) are yet to be established, even under the best scenario in which \(d_0\) belongs to the stationarity region \((-0.5, 0.5)\). One of the main objectives of this article is to fill this long-standing gap by establishing the consistency of BIC for ARFIMA models with few restrictions imposed on \(d_0\) and the error terms.

To fix ideas, suppose that \(\{y_t\}\) is generated according to the ARFIMA model,
\begin{equation}
1 - \sum_{j=1}^{p_0} \alpha_{0,j} \varepsilon^j = 0, \quad 1 - \sum_{j=1}^{q_0} \beta_{0,j} z^j = 0 \text{ for } |z| \leq 1,
\end{equation}
in which \(\sum_{a=b}^{c} = 0\) if \(a > b\),
\begin{equation}
1 - \sum_{j=1}^{p_0} \alpha_{0,j} z^j \text{ and } 1 - \sum_{j=1}^{q_0} \beta_{0,j} z^j \text{ have no common zeros},
\end{equation}
and
\begin{equation}
|\alpha_{0,p_0}| > 0, \quad |\beta_{0,q_0}| > 0,
\end{equation}
for non-zero \(p_0\) and \(q_0\). Following Hualde and Robinson (2011) and Chan, Huang and Ing (2013), the initial conditions are set to \(y_t = \varepsilon_t = 0\) for \(t \leq 0\). Let \(P\) and \(Q\) be prescribed upper bounds for \(p_0\) and \(q_0\). Having observed \(y_1, \ldots, y_n\), we are interested in choosing the unknown pair, \((p_0, q_0)\), from the set \(\{(p, q) | 0 \leq p \leq P, 0 \leq q \leq Q\}\). In the sequel, \((p, q)\) is referred to as the candidate model.

For a given candidate \((p, q)\), we estimate the vector formed by its AR, MA, and long-memory parameters using the conditional-sum-of-squares (CSS) estimate, \(\hat{\eta}_{n,pq}\), which is the minimizer of \(\sum_{t=1}^{n} \varepsilon_t^2(\eta_{pq})\) over \(\eta_{pq} = (\theta_{pq}^T, d^T) = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, d)^T\) in \(\Pi_{pq} \times D \subset \mathbb{R}^{p+q+1}\), where
\begin{equation}
\varepsilon_t(\eta_{pq}) = A_1 \theta_{pq}(B) A_2^{-1} \theta_{pq}(B)(1 - B)^d y_t,
\end{equation}
with
\begin{align*}
A_1 \theta_{pq}(z) &= 1 - \sum_{j=1}^{p} \alpha_j z^j, \\
A_2 \theta_{pq}(z) &= 1 - \sum_{j=1}^{q} \beta_j z^j,
\end{align*}
and \(\Pi_{pq} \times D\) is the parameter space to be specified in the next section. Note that \(\theta_{pq}\) vanishes if \(p = q = 0\), and \(\theta_{pq} = (\alpha_1, \ldots, \alpha_p)^T\) if \(p \geq 1\) and \(q = 0\), and \((\beta_1, \ldots, \beta_q)^T\) if \(q \geq 1\) and \(p = 0\). Consider a BIC-type criterion,
\begin{equation}
\phi(p, q) = n \log \hat{\sigma}_{pq}^2 + (p + q)p(n),
\end{equation}
where
\begin{equation}
\hat{\sigma}_{pq}^2 = n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\hat{\eta}_{n,pq}).
\end{equation}
is the CSS estimate of $\sigma^2$ when model $(p, q)$ is postulated, and $p(n)$ is a penalty term obeying
\begin{equation}
\lim_{n \to \infty} p(n) = \infty \quad \text{and} \quad \lim_{n \to \infty} \frac{p(n)}{n} = 0.
\end{equation}

Note that $n \log \hat{\sigma}^2_{p,q}$ in (1.7) is not exactly $-2 \log L(\hat{\eta}_{n,pq})$, where $L(\eta_{pq}) = L(\eta_{pq}|y_1, \ldots, y_n)$ is the Gaussian conditional likelihood. However, their difference, depending only on $n$, has no impact on order selection results. Let
\begin{equation}
(\hat{p}_n, \hat{q}_n) = \arg \min_{0 \leq p \leq P, 0 \leq q \leq Q} \phi(p, q).
\end{equation}

The main goal of this paper is to show that (1.10) is developed under the assumption that $\eta_{n,pq}$ does not possess a probability limit, and hence $\hat{\sigma}^2_{p,q}$ is difficult to analyze using standard methods that rely on the differences between the estimated parameters and their limits. We conquer this dilemma by establishing the intriguing result that
\begin{equation}
\lim_{n \to \infty} P\{(\hat{p}_n, \hat{q}_n) = (p_0, q_0)\} = 1.
\end{equation}

The major difficulty in proving (1.10) is to handle the asymptotic behavior of $\hat{\sigma}^2_{p,q}$ in the case of $p > p_0$ and $q > q_0$, which we call the two-sided overfitted candidate. In this case, the true parameters are not identifiable because there are infinitely many elements in the parameter space satisfying
\begin{equation}
A_{2, \theta_{pq}}(z)A^{-1}_{1, \theta_{pq}}(z) = A_{2, \theta_{pq_0}}(z)A^{-1}_{1, \theta_{pq_0}}(z).
\end{equation}

Therefore, $\hat{\eta}_{n,pq}$ does not possess a probability limit, and hence $\hat{\sigma}^2_{p,q}$ is difficult to analyze. The rest of the paper is organized as follows. In Section 2, (1.10) is developed under the assumption that $\{\varepsilon_t\}$ is a sequence of independent random variables. In Section 3, we establish (1.10) when $\{\varepsilon_t\}$ is conditionally heteroscedastic and satisfies the conditions described at the beginning of the section. Also, a refinement of our order selection method is proposed in this section to reduce the computational burden. The finite sample performance of the proposed methods is illustrated using simulations in Section 4. We conclude in Section 5. The proofs of all theorems are provided in Section 6. Further technical details are relegated to Appendix A and the supplementary material.
2. The Case of Independent Errors. Let the parameter space of the full model \((P, Q)\) be denoted by \(\Pi_{PQ} \times D\), where

\begin{equation}
D = [L, U], \text{ for some } -\infty < L < U < \infty,
\end{equation}

and \(\Pi_{PQ}\) is a compact set in \(\mathbb{R}^{P+Q}\) with element \(\theta_{PQ} = (\alpha_1, \ldots, \alpha_P, \beta_1, \ldots, \beta_Q)^T\) satisfying the stationarity condition,

\begin{equation}
A_1, \theta_{PQ}(z) \neq 0, A_2, \theta_{PQ}(z) \neq 0 \text{ for all } |z| \leq 1.
\end{equation}

For candidate model \((A)\), the stationarity condition, in model \((A)\) with the convention that the AR (MA) components vanish when \(p = 0\) (q = 0). It is clear that \(\Pi_{00} = \emptyset\) and \(\Pi_{pq} \times D\) is a compact set in \(\mathbb{R}^{p+q+1}\). The CSS estimate of the coefficient vector in model \((p, q)\) is given by

\begin{equation}
\hat{\theta}_{n, pq}^T \hat{d}_{n, pq} = \arg\min_{\eta_{pq} \in \Pi_{pq} \times D} \sum_{t=1}^{n} \varepsilon_t^2(\eta_{pq}).
\end{equation}

Recall the BIC-type criterion introduced in (1.7). The goal of this section is to establish (1.10) when \(\varepsilon_t\) are independent random variables satisfying \(E(\varepsilon_t) = 0\) and \(E(\varepsilon_t^2) = \sigma^2 \varepsilon > 0\) for all \(t\).

Equation (1.10) is ensured by

\begin{equation}
\lim_{n \to \infty} P(\hat{p}_n < p_0 \text{ or } \hat{q}_n < q_0) = 0
\end{equation}

and

\begin{equation}
\lim_{n \to \infty} P(\hat{p}_n > p_0, \hat{q}_n \geq q_0 \text{ or } \hat{p}_n \geq p_0, \hat{q}_n > q_0) = 0.
\end{equation}

Throughout the paper, we assume that

\begin{equation}
\eta_0 = (\theta_0^T, d_0)^T \in \text{int} \ \Pi_{p_0q_0} \times D,
\end{equation}

where \(\theta_0^T = (\alpha_{0,1}, \ldots, \alpha_{0,p_0}, \beta_{0,1}, \ldots, \beta_{0,q_0})\).

The proof of (2.3) is relatively easy. To see this, note that in the case of \(p < p_0\) or \(q < q_0\), (1.5) implies that there exists a small positive constant \(\delta\) for which

\begin{equation}
\hat{\eta}_{n, pq} \notin B_\delta(\eta_0),
\end{equation}

where \(B_\delta(\eta_0)\) is the open ball of radius \(\delta\) centered at \(\eta_0\). With a slight abuse of notation, \(\hat{\eta}_{n, pq}\) and \(\eta_0\) in (2.6) are viewed as \((\max\{p, p_0\} + \max\{q, q_0\} + 1)\)-dimensional vectors with undefined entries set to 0. By (2.6) and an argument similar to that used in the proof of Theorem 2.1 of Hualde and Robinson (2011), we show in (6.11)–(6.13) that for \(p < p_0\) or \(q < q_0\).

\begin{equation}
\lim_{n \to \infty} P(\hat{\sigma}_{pq}^2 - \sigma_{pq}^2 \leq c) = 0,
\end{equation}

where \(c\) is some positive constant. Thus, (2.3) follows.

On the other hand, the proof of (2.4) is much more complicated owing to the aforementioned identifiability problem. Let

\begin{equation}
\Pi_{pq}^+ = \{\theta_{pq} \in \mathbb{R}^{p+q} | (\alpha_1, \ldots, \alpha_P, 0, \ldots, 0, \beta_1, \ldots, \beta_Q, 0, \ldots, 0)^T \text{ satisfies (2.2)}\}
\end{equation}

and

\begin{equation}
S_{0, pq}^+ = \{(\theta_{pq}^T, d)^T \in \Pi_{pq}^+ \times \{d_0\} | \theta_{pq} \text{ obeys (1.11)}\},
\end{equation}
noting that $\Pi_{pq} \subset \Pi_{pq}'$ and $S_{0,pq}^+ \subseteq \Pi_{pq}'$ contains all points in $\Pi_{pq} \times D$ that generate the true model (1.2). The key step in the proof of (2.4) is to establish that
\begin{equation}
 d(\hat{\theta}_{n,pq}, S_{0,pq}^+) = O_p(n^{-1/2}), \quad p_0 \leq p \leq P, \quad q_0 \leq q \leq Q,
\end{equation}
which is a slightly stronger version of (1.12). Indeed, for the case of $p = p_0$ and $q = q_0$, $p_0 = P$, and $q = q_0$, $S_{0,pq}^+$ only contains one point and (2.9) has already been established by Hu and Robinson (2011) and Chan, Huang and Ing (2013). Their proofs, however, are not applicable to the two-sided over fitted model, in which $S_{0,pq}^+$ contains uncountably many points and $\hat{\theta}_{n,pq}$ does not possess a probability limit. We tackle this difficulty by implementing the bijective parameter transformation of Hannan (1980) and Hannan and Kavalieris (1984), which was originally designed for the special case of $d_0 = 0$. The transformed parameter, $\eta_{pq}^* = F_0(\eta_{pq})$, contains $\max\{p_0 + q, q_0 + p\}$ 1 + 1 identifiability components in the sense that the first $\max\{p_0 + q, q_0 + p\}$ components, $\theta_{pq}^*$, and the last component, $d$, of $\eta_{pq}^*$ are the same for all $\eta_{pq} \in S_{0,pq}^+$. Note that $F_0(\cdot)$ is a one-to-one linear transformation depending on $\eta_0$, which is defined in (6.18) of Section 6.2. Denote $\eta_{1,pq}^* = (\theta_{1,pq}^T, d)^T$ by $\eta_{0,1,pq}^*$ when $\eta_{pq} \in S_{0,pq}^+$. Also, let the subvector of $\hat{\eta}_{pq}^* = F_0(\hat{\theta}_{n,pq})$ corresponding to $\eta_{0,1,pq}^*$ be denoted by $\hat{\eta}_{1,pq}^*$. The uniqueness of $\eta_{0,1,pq}^*$ allows one to obtain
\begin{equation}
 \|\hat{\eta}_{1,pq}^* - \eta_{0,1,pq}^*\| = O_p(n^{-1/2}), \quad p_0 \leq p \leq P, \quad q_0 \leq q \leq Q,
\end{equation}
by analyzing $\sum_{t=1}^{n} \nabla_1 \varepsilon_t^2(F_0^{-1}(\hat{\eta}_{pq}^*)) - \sum_{t=1}^{n} \nabla_1 \varepsilon_t^2(F_0^{-1}(\eta_{pq}^*))$ based on the mean value theorem, and establishing uniform moment/probability bounds (see (6.30)–(6.33)) for
\begin{equation}
 n^{-1/2} \sum_{t=1}^{n} \nabla_1 \varepsilon_t(F_0^{-1}(\eta_{pq}^*)) \varepsilon_t, \quad n^{-1/2} \sum_{t=1}^{n} \nabla_1^2 \varepsilon_t(F_0^{-1}(\eta_{pq}^*)) \varepsilon_t,
\end{equation}
where $\text{tr}(\cdot)$ stands for the trace operator and for a twice differentiable function $f(\cdot)$ on $\mathbb{R}^{p+q+1}$, $\nabla_1 f(\eta_{pq}^*) = \frac{\partial}{\partial \eta_{pq}^T} f(\eta_{pq}^*)$, and $\nabla_2^2 f(\eta_{pq}^*) = \frac{\partial^2}{\partial \eta_{pq}^T \partial \eta_{pq}^T} f(\eta_{pq}^*)$. Equation (2.10) serves as an important vehicle for deriving (2.9). By making use of (2.9), we obtain
\begin{equation}
 \sigma_{pq}^2 - \sigma_{pq,00}^2 = n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\hat{\eta}_{n,pq}) - n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\hat{\eta}_{n,pq,00}) \quad = O_p(n^{-1}), \quad p_0 \leq p \leq P, \quad q_0 \leq q \leq Q,
\end{equation}
leading immediately to (2.4). For the details on the proofs of (2.3), (2.4), (2.9), and (2.10), see Sections 6.1 and 6.2. We are now in a position to state the main result of this section.

**Theorem 2.1.** Assume (1.2)–(1.5), (2.1), (2.2), (2.5), $\varepsilon_t$’s are independent, and
\begin{equation}
 \sup_{-\infty < t < \infty} E|\varepsilon_t|^4 < \infty.
\end{equation}
Then, (1.10) holds.

**Remark 2.1.** Baillie, Kapetanios and Papailias (2014) have proposed a modified information criterion for choosing ARFIMA models in situations where $-\infty < d_0 < \infty$. However, to establish the criterion’s selection consistency, they have imposed an assumption that for each candidate $(p, q)$, with known $d_0$, the CSS estimators (or the maximum likelihood estimators) of the AR and MA parameters converge to non-random limits at a rate of $O_p(n^{-1/2})$. 
As explained previously, this assumption is obviously violated by any two-sided overfitted candidate owing to the non-identifiability issue. Moreover, they have assumed a high-level assumption similar to (2.12), whose justification, as shown in this study, is far from being trivial.

3. The Case of Conditional Heteroscedastic Errors. In this section, we assume that \( \{\varepsilon_t\} \) is a martingale difference sequence with respect to \( \{\mathcal{F}_t\} \), an increasing sequence of \( \sigma \)-fields. We further assume that \( \{\varepsilon_t^2\} \) admits an infinite-order moving average representation,

\[
\varepsilon_t^2 - E(\varepsilon_t^2) = \sum_{s=0}^{\infty} a_s^T w_{t-s},
\]

in which \( a_s \) are \( l \)-dimensional real vectors for some \( l \geq 1 \),

\[
||a_s|| = O((s+1)^{-t}), \text{ for some } t > 1,
\]

and \( \{w_t, \mathcal{F}_t\} \) is an \( L^1 \)-bounded martingale difference sequence. Assumptions (3.1) and (3.2) include stationary GARCH and GJR-GARCH processes as special cases. To see this, consider a stationary GJR-GARCH\((p'_0, q'_0)\) model,

\[
(3.3) \quad \varepsilon_t = \sigma_t z_t, \quad \sigma_t^2 = \varphi_{0,0} + \sum_{i=1}^{p'_0} \varphi_{0,i} \varepsilon_{t-i}^2 + \sum_{j=1}^{q'_0} \psi_{0,j} \sigma_{t-j}^2 + \sum_{k=1}^{p'_0} \zeta_{0,k} \varepsilon_{t-k}^2 I_{\{\varepsilon_{t-k} < 0\}},
\]

where \( p'_0 \) and \( q'_0 \) are some positive integers, \( \varphi_{0,0} > 0 \), \( z_t \) are i.i.d. and symmetric random variables with zero mean and common variance 1, and \( \varphi_{0,i}, \psi_{0,j}, \) and \( \zeta_{0,k} \) are non-negative constants obeying

\[
(3.4) \quad \sum_{i=1}^{p'_0} \varphi_{0,i} + \sum_{j=1}^{q'_0} \psi_{0,j} + \sum_{k=1}^{p'_0} \frac{\zeta_{0,k}}{2} < 1.
\]

According to (3.3), we express \( \varepsilon_t^2 \) as

\[
(3.5) \quad \varepsilon_t^2 = \varphi_{0,0} + \sum_{i=1}^{\max\{p'_0, q'_0\}} (\varphi_{0,i} + \psi_{0,i} + \frac{\zeta_{0,i}}{2}) \varepsilon_{t-i}^2 + w_{1,t} \sum_{j=1}^{q'_0} \psi_{0,j} w_{1,t-j} + \sum_{k=1}^{p'_0} \zeta_{0,k} w_{2,t-k},
\]

where \( w_{1,t} = \varepsilon_t^2 - \sigma_t^2, w_{2,t} = \varepsilon_t^2 I_{\{\varepsilon_{t} < 0\}} - \frac{1}{2} \varepsilon_t^2, \) and \( \varphi_{0,i}, \zeta_{0,i}, \) and \( \psi_{0,j} \) are set to 0 when \( i > p'_0 \) and \( j > q'_0 \). By (3.4), (3.5), and the fact that \( \{w_{1,t}, w_{2,t}\}^T \) is an \( L^1 \)-bounded martingale difference sequence with respect to \( \{\sigma(z_{s}, s \leq t)\} \), where \( \sigma(z_{s}, s \leq t) \) is the \( \sigma \)-field generated by \( \{z_t, z_{t-1}, \ldots\} \), it can be shown that (3.1) and (3.2) hold with \( l = 2 \), \( w_t = (w_{1,t}, w_{2,t})^T \), and \( a_s = (b_s, c_s)^T \), where \( b_j \) and \( c_j \), respectively, satisfy

\[
\sum_{j=0}^{\infty} (b_j z^j)^2 = \frac{1}{1 - \sum_{j=1}^{\max\{p'_0, q'_0\}} (\varphi_{0,i} + \psi_{0,i} + \frac{\zeta_{0,i}}{2}) z^j}
\]

and

\[
\sum_{j=0}^{\infty} (c_j z^j)^2 = \frac{\sum_{j=1}^{p'_0} \zeta_{0,j} z^j}{1 - \sum_{j=1}^{\max\{p'_0, q'_0\}} (\varphi_{0,i} + \psi_{0,i} + \frac{\zeta_{0,i}}{2}) z^j}.
\]

Moreover, since \( |b_j| \) and \( |c_j| \) decay exponentially as \( j \) increases, (3.2) is valid for arbitrarily large values of \( t \). When \( \zeta_{0,k} = 0 \) for all \( k \), (3.3) reduces to a stationary GARCH process. By the same argument, (3.1) and (3.2) hold with \( l = 1 \), \( w_t = w_{1,t} \), and \( a_s = b_s \).
Due to their nonparametric nature, (3.1) and (3.2) are much more flexible than assuming that \( \{\varepsilon_t\} \) is a stationary GJR-GARCH or GARCH model of finite order. The next theorem shows that the consistency of BIC established in the previous section carries over to conditional heteroscedastic errors obeying (3.1), (3.2), and a mild moment condition.

**Theorem 3.1.** Assume (1.2)–(1.5), (2.1), (2.2), (2.5), (3.1), (3.2), and

\[
\sup_{-\infty < t < \infty} E\|w_t\|^2 < \infty.
\]

Then, (1.10) follows.

A few comments are in order regarding Theorem 3.1. To start with, note that (3.1) and (3.2) are fulfilled when \( \{\varepsilon_t\} \) satisfies the assumptions of Theorem 2.1 or \( \{\varepsilon_t, F_t\} \) is a martingale difference sequence obeying

\[
E(\varepsilon_t^2 | F_{t-1}) = E(\varepsilon_1^2) = \sigma_\varepsilon^2 \text{ a.s.}
\]

Moreover, since (3.6) and (2.13) are equivalent under these assumptions on \( \{\varepsilon_t\} \), Theorem 3.1 includes Theorem 2.1 as a special case. Next, for stationary GARCH and GJR-GARCH models, it is easy to see that (3.6) holds when

\[
E|\sigma_t|^4 < \infty,
\]
a condition commonly made in the literature on GARCH-type models; see Ling and McAleer (2002a) and Ling and McAleer (2002b). In addition, under (3.1), (3.2), and (3.6), the uniform moment/probability bounds, (6.30)–(6.33), established for (2.11) in the case of independent errors are no longer applicable. To alleviate this difficulty, Lemma 6.2 extends the uniform moment bounds in Lemma B.1 of Chan and Ing (2011) to linear processes driven by conditionally heteroscedastic errors, thereby generalizing (6.30)–(6.33) to error terms satisfying (3.1), (3.2), and (3.6). These generalizations enable us to establish (2.9) under the assumptions of Theorem 3.1. Once (2.9) is obtained, Theorem 3.1 can be proved in a similar fashion as in the proof of Theorem 2.1. Last, Bardet, Kamila and Kengne (2020) have recently proposed a BIC-type criterion and proved its selection consistency for stationary ARMA-GARCH models of finite order. However, similar to the result of Baillie, Kapetanios and Papailias (2014), their result also relies on an identifiability condition, which is inevitably violated by a two-sided over fitted candidate.

In fact, when an identifiability condition is postulated, there is no fundamental difference between parameter estimation consistency and variable selection consistency. More specifically, when this type of condition is assumed, one can easily establish model selection consistency by applying hard thresholds on the consistent estimates of the parameters in the full model (the largest candidate model). This approach, however, is not proper when the parameters in the full model are not identifiable, and hence no consistent estimates are available. This difficulty becomes more severe in situations where \( d_0 \) is allowed to be any real number and \( \{\varepsilon_t\} \) can be conditionally heteroscedastic. The main advantage delivered by Theorem 3.1 is that BIC still works well for such a challenging situation.

Although Theorem 3.1 shows that (1.9) is consistent, it involves computing \( \phi(p, q) \) for all candidate models, which can be time consuming because \( \hat{\sigma}_{pq}^2 \) in \( \phi(p, q) \) is obtained by a nonlinear optimization. Inspired by Hannan and Rissanen (1982), we introduce a refinement of (1.9), referred to as a refinement of BIC (RBIC), that can substantially reduce the number of searches for the best candidate, in particular when \( P \) and \( Q \) are large. The details of the proposed method are as follows.
Algorithm 1: RBIC

1: Define

\[ \hat{r}_n^{(1)} = \arg \min_{0 \leq r \leq R} \phi(r, r), \]

where \( R = \max\{P, Q\} \).

2: Estimate \( p_0 \) and \( q_0 \) using

\[ \hat{p}_n^{(1)} = \arg \min_{0 \leq p \leq \hat{r}_n^{(1)}} \phi(p, \hat{r}_n^{(1)}) \]

and

\[ \hat{q}_n^{(1)} = \arg \min_{0 \leq q \leq \hat{r}_n^{(1)}} \phi(\hat{r}_n^{(1)}, q). \]

The consistency of RBIC is confirmed in the next corollary.

**Corollary 3.1.** Under the same assumptions as in Theorem 3.1,

\[
\lim_{n \to \infty} P\{ (\hat{p}_n^{(1)}, \hat{q}_n^{(1)}) = (p_0, q_0) \} = 1.
\]

Corollary 3.1 can be proved in the same manner as that in the proof of Theorem 3.1; details are omitted.

4. Simulation Studies. In this section, we illustrate the finite-sample performance of RBIC using simulations. The performance of BIC ((1.9)) is not reported here because both methods are asymptotically equivalent (see Theorem 3.1 and Corollary 3.1), and the latter is more time-consuming.

We first generate data from the following ARFIMA models,

(I) \((1 + 0.7B)(1 - B)^{d_0} y_t = \varepsilon_t,\)

(II) \((1 - 0.8B)(1 - B)^{d_0} y_t = (1 + 0.5B)\varepsilon_t,\)

(III) \((1 - 1.8B + 0.9B^2)(1 - B)^{d_0} y_t = (1 - 1.42B + 0.73B^2)\varepsilon_t,\)

in which \( \varepsilon_t \) are i.i.d. standard normal random variables and \( d_0 \in \{-0.5, 0, 0.25, 0.5, 0.75, 1, 1.5\} \).

The number of observations, \( n \), is set to 250 or 500, and the number of replications is given by \( M = 1000 \). To implement RBIC, we let \( P = Q = 4 \) and \( p(n) = \log n \), recalling that \( p(n) \) is the penalty term of RBIC. Denote by \( \hat{p}_l^{(1)}, \hat{q}_l^{(1)} \), and \( \hat{d}_l^{(1)} \) the estimators of \( p_0, q_0 \), and \( d_0 \) obtained in the \( l \)-th simulation, where \( 1 \leq l \leq M \). Note that \( \hat{d}_l^{(1)} \) is derived from model \( (\hat{p}_l^{(1)}, \hat{q}_l^{(1)}) \). We compute the following performance measures,

\[
\text{Frequency of overfitting (Over)} = \sum_{l=1}^{M} [1(\hat{p}_l^{(1)} \geq p_0, \hat{q}_l^{(1)} \geq q_0) - 1(\hat{p}_l^{(1)} = p_0, \hat{q}_l^{(1)} = q_0)],
\]

\[
\text{Frequency of exact selection (Ext)} = \sum_{l=1}^{M} 1(\hat{p}_l^{(1)} = p_0, \hat{q}_l^{(1)} = q_0),
\]

\[
\text{Frequency of underfitting (Under)} = \sum_{l=1}^{M} 1(\hat{p}_l^{(1)} < p_0 \text{ or } \hat{q}_l^{(1)} < q_0),
\]

\[
\text{Mean absolute error (MAE) of } \{\hat{d}_l^{(1)}\} \text{ (d-MAE)} = \frac{1}{M} \sum_{l=1}^{M} |\hat{d}_l^{(1)} - d_0|,
\]

\[
\text{where } R = \max\{P, Q\}.
\]
and summarize the results in Tables 1–3 for Models (I)–(III), respectively. These tables show that the performance of RBIC is quite satisfactory because for all models and all $d_0$ values, the Ext values of RBIC are not less than 852 when $n = 250$ and 922 when $n = 500$. The d-MAE values in these tables also reveal that $d_0$ is accurately estimated by our method.

Next, we focus on model (II), but with $\{\varepsilon_t\}$ generated by the following conditionally heteroscedastic models,

- **ARCH (1):** $\varepsilon_t = \sigma_t z_t$, $\sigma_t^2 = 0.4 + 0.5 \varepsilon_{t-1}^2$,
- **GARCH (1, 1):** $\varepsilon_t = \sigma_t z_t$, $\sigma_t^2 = 0.4 + 0.2 \varepsilon_{t-1}^2 + 0.7 \sigma_{t-1}^2$,
- **GARCH (2, 2):** $\varepsilon_t = \sigma_t z_t$, $\sigma_t^2 = 0.4 + 0.3 \varepsilon_{t-1}^2 + 0.2 \varepsilon_{t-2}^2 + 0.2 \sigma_{t-1}^2 + 0.1 \sigma_{t-2}^2$,
- **GJR-GARCH (1, 1):** $\varepsilon_t = \sigma_t z_t$, $\sigma_t^2 = 0.4 + 0.2 \varepsilon_{t-1}^2 + 0.6 \sigma_{t-1}^2 + 0.1 \varepsilon_{t-1}^2 I(\varepsilon_{t-1} < 0)$,

where $z_t$ are i.i.d. standard normal random variables. All other settings are the same as those in the case of independent errors. The performance of RBIC in model (II) with these four different errors is summarized in Table 4–7. As shown in Tables 2 and 4–7, the performance of RBIC is slightly deteriorated by the conditional heteroscedasticity. However, the method’s Ext values are still maintained at a range of 795–875 when $n = 250$, and 899–962 when $n = 500$.

As seen from these simulated scenarios, RBIC identifies the true orders near or over 80% of the time when $n = 250$, and increases to near or over 90% when $n = 500$; we conclude that the finite-sample behavior of RBIC concurs with the asymptotic results developed in Section 3.

5. Concluding Remarks. In this work, we propose using BIC-type criteria to choose ARFIMA models of finite order. The major contribution is to show that the proposed criteria achieve order selection consistency in very challenging situations where the memory parameter is allowed to be any real number, the error terms can be conditionally heteroscedastic,
and the candidate models are not necessarily identifiable. This result substantially enhances the applicability of the BIC, which is further confirmed by numerical simulations.

On the other hand, the performance of Akaike’s information criterion (AIC) in choosing fractionally integrated AR models of infinite order is yet to be explored. In the case of $d_0 = 0$, the asymptotic efficiency of AIC for independent-realization and same-realization predictions has been proved by Shibata (1980) and Ing and Wei (2005), respectively. The latter result has

**Table 3**
Simulation results of RBIC in model (III) with independent errors.

<table>
<thead>
<tr>
<th>n</th>
<th>d=-0.5</th>
<th>0</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>250 Over</td>
<td>29</td>
<td>27</td>
<td>38</td>
<td>28</td>
<td>33</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td>Ext</td>
<td>969</td>
<td>971</td>
<td>959</td>
<td>971</td>
<td>967</td>
<td>964</td>
<td></td>
</tr>
<tr>
<td>Under</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>d-MAE</td>
<td>0.068</td>
<td>0.107</td>
<td>0.116</td>
<td>0.076</td>
<td>0.097</td>
<td>0.113</td>
<td>0.079</td>
</tr>
<tr>
<td>500 Over</td>
<td>3</td>
<td>10</td>
<td>8</td>
<td>6</td>
<td>7</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Ext</td>
<td>997</td>
<td>990</td>
<td>992</td>
<td>994</td>
<td>993</td>
<td>985</td>
<td></td>
</tr>
<tr>
<td>Under</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>d-MAE</td>
<td>0.049</td>
<td>0.062</td>
<td>0.065</td>
<td>0.051</td>
<td>0.060</td>
<td>0.068</td>
<td>0.056</td>
</tr>
</tbody>
</table>

**Table 4**
Simulation results of RBIC in model (II) with ARCH(1) errors.

<table>
<thead>
<tr>
<th>n</th>
<th>d=-0.5</th>
<th>0</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>250 Over</td>
<td>13</td>
<td>16</td>
<td>15</td>
<td>13</td>
<td>11</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Ext</td>
<td>808</td>
<td>869</td>
<td>858</td>
<td>854</td>
<td>875</td>
<td>875</td>
<td></td>
</tr>
<tr>
<td>Under</td>
<td>179</td>
<td>115</td>
<td>127</td>
<td>133</td>
<td>114</td>
<td>113</td>
<td>124</td>
</tr>
<tr>
<td>d-MAE</td>
<td>0.215</td>
<td>0.179</td>
<td>0.184</td>
<td>0.183</td>
<td>0.175</td>
<td>0.179</td>
<td>0.184</td>
</tr>
<tr>
<td>500 Over</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Ext</td>
<td>939</td>
<td>958</td>
<td>962</td>
<td>948</td>
<td>952</td>
<td>954</td>
<td></td>
</tr>
<tr>
<td>Under</td>
<td>57</td>
<td>35</td>
<td>32</td>
<td>44</td>
<td>41</td>
<td>38</td>
<td>49</td>
</tr>
<tr>
<td>d-MAE</td>
<td>0.162</td>
<td>0.121</td>
<td>0.120</td>
<td>0.129</td>
<td>0.124</td>
<td>0.130</td>
<td>0.154</td>
</tr>
</tbody>
</table>

**Table 5**
Simulation results of RBIC in model (II) with GARCH(1,1) errors.

<table>
<thead>
<tr>
<th>n</th>
<th>d=-0.5</th>
<th>0</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>250 Over</td>
<td>17</td>
<td>17</td>
<td>15</td>
<td>14</td>
<td>18</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>Ext</td>
<td>813</td>
<td>856</td>
<td>853</td>
<td>848</td>
<td>863</td>
<td>865</td>
<td></td>
</tr>
<tr>
<td>Under</td>
<td>170</td>
<td>127</td>
<td>132</td>
<td>138</td>
<td>119</td>
<td>116</td>
<td>127</td>
</tr>
<tr>
<td>d-MAE</td>
<td>0.218</td>
<td>0.180</td>
<td>0.185</td>
<td>0.187</td>
<td>0.177</td>
<td>0.180</td>
<td>0.189</td>
</tr>
<tr>
<td>500 Over</td>
<td>11</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>10</td>
</tr>
<tr>
<td>Ext</td>
<td>933</td>
<td>954</td>
<td>955</td>
<td>944</td>
<td>954</td>
<td>950</td>
<td></td>
</tr>
<tr>
<td>Under</td>
<td>56</td>
<td>34</td>
<td>33</td>
<td>44</td>
<td>34</td>
<td>38</td>
<td>44</td>
</tr>
<tr>
<td>d-MAE</td>
<td>0.164</td>
<td>0.123</td>
<td>0.123</td>
<td>0.133</td>
<td>0.123</td>
<td>0.132</td>
<td>0.153</td>
</tr>
</tbody>
</table>

**Table 6**
Simulation results of RBIC in model (II) with GARCH(2,2) errors.

<table>
<thead>
<tr>
<th>n</th>
<th>d=-0.5</th>
<th>0</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>250 Over</td>
<td>29</td>
<td>26</td>
<td>25</td>
<td>29</td>
<td>28</td>
<td>28</td>
<td>24</td>
</tr>
<tr>
<td>Ext</td>
<td>795</td>
<td>818</td>
<td>806</td>
<td>800</td>
<td>820</td>
<td>828</td>
<td></td>
</tr>
<tr>
<td>Under</td>
<td>206</td>
<td>156</td>
<td>169</td>
<td>171</td>
<td>152</td>
<td>144</td>
<td>155</td>
</tr>
<tr>
<td>d-MAE</td>
<td>0.229</td>
<td>0.189</td>
<td>0.193</td>
<td>0.200</td>
<td>0.188</td>
<td>0.186</td>
<td>0.196</td>
</tr>
<tr>
<td>500 Over</td>
<td>27</td>
<td>33</td>
<td>29</td>
<td>30</td>
<td>38</td>
<td>29</td>
<td>30</td>
</tr>
<tr>
<td>Ext</td>
<td>899</td>
<td>908</td>
<td>910</td>
<td>905</td>
<td>903</td>
<td>910</td>
<td></td>
</tr>
<tr>
<td>Under</td>
<td>74</td>
<td>59</td>
<td>61</td>
<td>65</td>
<td>59</td>
<td>61</td>
<td>66</td>
</tr>
<tr>
<td>d-MAE</td>
<td>0.178</td>
<td>0.138</td>
<td>0.142</td>
<td>0.146</td>
<td>0.138</td>
<td>0.146</td>
<td>0.170</td>
</tr>
</tbody>
</table>
Simulation results of RBIC in model (II) with GJR-GARCH(1,1) errors.

<table>
<thead>
<tr>
<th>n</th>
<th>d=−0.5</th>
<th>0</th>
<th>0.25</th>
<th>0.5</th>
<th>0.75</th>
<th>1</th>
<th>1.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>Over</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>12</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Ext</td>
<td>809</td>
<td>853</td>
<td>837</td>
<td>840</td>
<td>853</td>
<td>857</td>
</tr>
<tr>
<td></td>
<td>Under</td>
<td>177</td>
<td>132</td>
<td>147</td>
<td>148</td>
<td>132</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td>d-MAE</td>
<td>0.217</td>
<td>0.175</td>
<td>0.180</td>
<td>0.183</td>
<td>0.176</td>
<td>0.178</td>
</tr>
<tr>
<td>500</td>
<td>Over</td>
<td>11</td>
<td>15</td>
<td>11</td>
<td>13</td>
<td>14</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>Ext</td>
<td>924</td>
<td>943</td>
<td>945</td>
<td>937</td>
<td>946</td>
<td>947</td>
</tr>
<tr>
<td></td>
<td>Under</td>
<td>65</td>
<td>42</td>
<td>44</td>
<td>50</td>
<td>40</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>d-MAE</td>
<td>0.172</td>
<td>0.129</td>
<td>0.130</td>
<td>0.139</td>
<td>0.130</td>
<td>0.134</td>
</tr>
</tbody>
</table>

been generalized by Ing, Sin and Yu (2012) to the case where $d_0$ is a non-negative integer. However, in the case of $-\infty < d_0 < \infty$, AIC’s asymptotic efficiency has not been established, which will be dealt with in further research.

Acknowledgements. We thank Ruey S. Tsay for insightful suggestions and comments.

SUPPLEMENTARY MATERIAL

Supplement to "Consistent Order Selection for ARFIMA Processes" (DOI:). The supplementary material contains the proofs of (6.5), (6.6), Lemma 6.2, and (A.12).

6. Proofs. Throughout the rest of the paper, $C$ denotes a generic positive constant independent of $n$.

6.1. Proof of Theorem 2.1.

Proof. We first prove (2.4). Let $p_0 \leq p \leq P$ and $q_0 \leq q \leq Q$ be given. For any $\delta > 0$, define

$$S_{\delta,pq} = \{ \eta_{pq} \in \Pi_{pq} \times D | d(\eta_{pq}, S_{0,pq}^+) < \delta \}.$$

Let $\tilde{\eta}_{n,pq} = \arg \min_{\eta_{pq} \in S_{0,pq}^+} \| \tilde{\eta}_{n,pq} - \eta_{pq} \|$. Then,

$$\lim_{n \to \infty} P(\tilde{\eta}_{n,pq} \text{ exists}) = 1,$$

which is ensured by Lemma 6.1 and the compactness of $\Pi_{pq}$. In the rest of the section, we abbreviate $\tilde{\eta}_{n,pq}, \tilde{\eta}_{n,pq}, \eta_{pq}, S_{0,pq}^+, \tilde{\eta}_{n}, \eta, S_{0}^+$, and $S_{\delta}$ for notational simplicity.

On $\Omega_n = \{ \tilde{\eta}_n \text{ exists} \}$,

$$\hat{\sigma}_{p_0,q_0}^2 - \hat{\sigma}_{pq}^2 = n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\hat{\eta}_{n,p_0,q_0}) - n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\eta_0)$$

$$\quad - (n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\tilde{\eta}_n) - n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\tilde{\eta}_n)).$$
Let $\nabla \epsilon_t(\eta) = \partial \epsilon_t(\eta)/\partial \eta = (\nabla \epsilon_t(\eta)_1, \ldots, \nabla \epsilon_t(\eta)_{\bar{r}})^T$, with $\bar{r} = p + q + 1$. By the mean value theorem,

$$\left| n^{-1} \sum_{t=1}^{n} \epsilon_t^2(\tilde{\eta}_n) - n^{-1} \sum_{t=1}^{n} \epsilon_t^2(\hat{\eta}_n) \right|$$

(6.3) \leq \left| \int_0^1 n^{-1} \sum_{t=1}^{n} \epsilon_t \nabla \epsilon_t(\hat{\eta}_n + r(\hat{\eta}_n - \tilde{\eta}_n))dr \right|^T (\hat{\eta}_n - \tilde{\eta}_n) 

+ \left| \int_0^1 n^{-1} \sum_{t=1}^{n} r(\hat{\eta}_n - \tilde{\eta}_n)^T \nabla \epsilon_t(\eta^*_{t,r}) \nabla T \epsilon_t(\hat{\eta}_n + r(\hat{\eta}_n - \tilde{\eta}_n))dr \right| (\hat{\eta}_n - \tilde{\eta}_n),$$

where $\eta^*_{t,r}$ satisfies $\|\eta^*_{t,r} - \tilde{\eta}_n\| \leq r \|\hat{\eta}_n - \tilde{\eta}_n\|$. Given $N > 0$, define $\Lambda_n(M) = \{d(\hat{\eta}_n, S_0^n) < \nu_n\}$, where $\nu_n = \min\{M^{1/4}n^{-1/2}, \delta_1\}$, for some $0 < \delta_1 < 1/2$. It follows from (6.3), Jensen’s inequality, and the Cauchy–Schwarz inequality that

$$P\left( \left| \sum_{t=1}^{n} \epsilon_t^2(\tilde{\eta}_n) - \sum_{t=1}^{n} \epsilon_t^2(\hat{\eta}_n) \right| > M, \Omega_n \cap \Lambda_n(M) \right)$$

$$\leq P(r^{1/2} \max_{1 \leq i \leq r} \sup_{\eta \in S_{\nu_n}} \left| \sum_{t=1}^{n} \epsilon_t \nabla \epsilon_t(\eta)_i \right| \|\hat{\eta}_n - \tilde{\eta}_n\| > 2^{-1}M, \Omega_n \cap \Lambda_n(M) \right)$$

(6.4) + $P\left( r \max_{1 \leq i \leq r} \sup_{\eta \in S_{\nu_n}} \left( \sum_{t=1}^{n} (\nabla \epsilon_t(\eta)_i)^2 \right) \|\hat{\eta}_n - \tilde{\eta}_n\|^2 > 2^{-1}M, \Omega_n \cap \Lambda_n(M) \right)$$

$$\leq P\left( \max_{1 \leq i \leq r} \sup_{\eta \in S_{\nu_n}} \left| \sum_{t=1}^{n} \epsilon_t \nabla \epsilon_t(\eta)_i \right| > 2^{-1}r^{-1/2}M^{3/4}, \Omega_n \right)$$

+ $P\left( \max_{1 \leq i \leq r} \sup_{\eta \in S_{\nu_n}} \left\{ \sum_{t=1}^{n} (\nabla \epsilon_t(\eta)_i)^2 \right\} > 2^{-1}r^{-1}M^{1/2}, \Omega_n \right)$. 

By virtue of (2.13), Lemma B.1 of Chan and Ing (2011), and Markov’s inequality, it is shown in Section S1 of the supplementary material that

$$E\left( \max_{1 \leq i \leq r} \sup_{\eta \in S_{\nu_n}} \left| \sum_{t=1}^{n} \epsilon_t \nabla \epsilon_t(\eta)_i \right|^2 \right) = O(1)$$

(6.5) and

$$P\left( \max_{1 \leq i \leq r} \sup_{\eta \in S_{\nu_n}} \left\{ \sum_{t=1}^{n} (\nabla \epsilon_t(\eta)_i)^2 \right\} > M \right) = o(1),$$

(6.6) for some $M > 0$. Combining (6.4)–(6.6) yields that for any $\epsilon > 0$, there exist $M_1, N_1 > 0$ such that for all $n > N_1$,

$$P\left( \left| \sum_{t=1}^{n} \epsilon_t^2(\tilde{\eta}_n) - \sum_{t=1}^{n} \epsilon_t^2(\hat{\eta}_n) \right| > M_1, \Omega_n \cap \Lambda_n(M_1) \right) < \epsilon/2.$$ 

Moreover, Lemma 6.1 ensures that for any $\epsilon > 0$, there exist $M_2, N_2 > 0$ such that for all $n > N_2$,

$$P\left( \Lambda_n^c(M_2) \right) < \epsilon/2.$$ 

(6.8)
Thus, by (6.7) and (6.8), for any $\epsilon > 0$, there exists $M = \max\{M_1, M_2\}$ such that for all $n >\max\{N_1, N_2\},$

\begin{equation}
(6.9) \quad P\left(\left| \sum_{t=1}^{n} \varepsilon_t^2(\hat{\eta}_n) - \sum_{t=1}^{n} \varepsilon_t^2(\tilde{\eta}_n) \right| > M, \Omega_n \right) < \epsilon.
\end{equation}

Similarly, for any $\epsilon > 0$, there exist $M_3, N_3 > 0$ such that for all $n > N_3,$

\begin{equation}
(6.10) \quad P\left(\left| \sum_{t=1}^{n} \varepsilon_t^2(\hat{\eta}_{n,pq,0}) - \sum_{t=1}^{n} \varepsilon_t^2(\eta_0) \right| > M_3 \right) < \epsilon.
\end{equation}

Combining (6.1), (6.2), (6.9), and (6.10) yields (2.4).

It remains to prove (2.3). For a given $(p, q)$, with $p < p_0$ or $q < q_0$, we can treat (1.2) as an ARFIMA($p^\ast, d_0, q^\ast$) model,

\begin{equation}
(6.11) \quad (1 - \alpha_{0,1} B - \cdots - \alpha_{0,p_0} B^p)(1 - B)^{d_0} y_t = (1 - \beta_{0,1} B - \cdots - \beta_{0,q_0} B^q)\varepsilon_t,
\end{equation}

where $p^\ast = \max\{p, p_0\}$, $q^\ast = \max\{q, q_0\}$, and $\alpha_{0,i} = \beta_{0,j} = 0$ for $i > p_0$ and $j > q_0$. Define $\eta_{0,p^\ast q^\ast} = (\alpha_{0,1}, \ldots, \alpha_{0,n}, 0, \ldots, 0, \beta_{0,1}, \ldots, \beta_{0,n}, 0, \ldots, 0, d_0) \in \mathbb{R}^{p^\ast+q^\ast+1}$, which is $\eta_0$ extended by adding zeros to the overfitted entries. It follows from (1.4), (6.11), and the proofs of (2.8) and (2.18) of Hualde and Robinson (2011) that for any $\delta > 0$, there exists a small number $c_\delta > 0$ such that

\begin{equation}
(6.12) \quad P\left(\inf_{\eta_{p^\ast q^\ast} \in \Pi_{p^\ast q^\ast}} n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\eta_{p^\ast q^\ast}) - n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\eta_{0,p^\ast q^\ast}) \leq c_\delta \right) = o(1).
\end{equation}

Using (1.5) and (6.12) with a small enough $\delta$, we obtain for some small constant $c > 0$ that

\begin{equation}
(6.13) \quad P\left(n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\hat{\eta}_n) - n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\eta_0) \leq c\right)
\leq P\left(\inf_{\eta_{p^\ast q^\ast} \in \Pi_{p^\ast q^\ast}} n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\eta_{p^\ast q^\ast}) - n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\eta_{0,p^\ast q^\ast}) \leq c\right) = o(1).
\end{equation}

As a result, (2.3) follows from (6.10), (6.13), and $\hat{\sigma}_{pq}^2 - \hat{\sigma}_{pq} = n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\hat{\eta}_{n,pq,0}) - n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\eta_0) - (n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\hat{\eta}_n) - n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\eta_0)).$ Equation (1.10) now is a straightforward consequence of (2.3) and (2.4).

\section*{6.2. Proof of (2.9) in the case of independent errors.}

\textbf{Lemmal 6.1.} \textit{Under the assumptions of Theorem 2.1, (2.9) holds.}

\textbf{Proof of Lemma 6.1.} For given $p_0 \leq p \leq P$ and $q_0 \leq q \leq Q$, denote $\Pi_{pq}, \Pi_{pq}^+, \hat{\theta}_{n,pq}, \hat{d}_{n,pq}$, and $\theta_{pq}$ by $\Pi, \Pi^+, \hat{\theta}_n, \hat{d}_n$, and $\theta$. First, we show that for any $\delta > 0,$

\begin{equation}
(6.14) \quad P(\hat{\eta}_n \in \Pi \times D - S_\delta) = o(1),
\end{equation}

which is, in turn, ensured by

\begin{equation}
(6.15) \quad \lim_{n \to \infty} P\left(\inf_{\eta \in \Pi \times D - S_\delta} n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\eta) - n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\eta_0) \leq c_\delta \right) = 0,
\end{equation}

where $c_\delta = \max\{C_{\delta,1}, C_{\delta,2}, C_{\delta,3}\} $ is a constant.
where \( c_\delta \) is a small positive constant depending on \( \delta \). Note that for \( \eta \in \Pi \times D - S_\delta \),

\[
(1 - z)^{d-d_0} A_1 \theta (z) A_2^{-1}_2 (z) A_1^{-1}_1 \theta_0 (z) A_2 \theta_0 (z) \neq 1.
\]

Thus, (6.15) follows from the same arguments as those in the proofs of (2.8) and (2.18) in Hualde and Robinson (2011) except that the open ball centered at the true parameter \( \eta_0 \) is replaced by \( S_\delta \).

Next, consider the linear transformation introduced in Theorem 1 of Hannan (1980) (or Section 3 of Hannan and Kavalieris (1984)), which asserts that there exists a \((p+q) \times (p+q)\) full rank matrix \( A \), depending only on \( \alpha_{0,i} \) and \( \beta_{0,j} \), such that

\[
A (\theta - \theta_{0,pq}) = A \left\{ \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{p} \\ \beta_1 \\ \vdots \\ \beta_{q} \end{pmatrix} \right\} - \left\{ \begin{pmatrix} \alpha_{0,1} \\ \vdots \\ \alpha_{0,p} \\ \beta_{0,1} \\ \vdots \\ \beta_{0,q} \end{pmatrix} \right\} = \begin{pmatrix} \theta_1^* \\ \theta_2^* \end{pmatrix}
\]

where \( \theta^*_1 \) is the \( \max \{p+q, q + p_0\} \)-dimensional vector formed by the coefficients of the polynomial \( A_1 \theta (z) A_2 \theta_0 (z) - A_1 \theta_0 (z) A_2 \theta (z) \), and \( \theta^*_2 \) is a \( \min \{p-p_0, q - q_0\} \)-dimensional vector. Let \( \theta^* = (\theta_1^*, \theta_2^*)^T \), \( s_1^* = \max \{p+q, q + p_0\} \), \( s_2^* = \min \{p-p_0, q - q_0\} \), \( \eta^*_1 = (\theta_1^* T, d)^T \), and \( F_0 \) be the \((p+q+1) \times (p+q+1)\) full rank matrix satisfying

\[
\eta^* := F_0 (\eta) = (\theta^* T, d)^T.
\]

Denote by \( \hat{\theta}_{2n}, \hat{\eta}_{1n}^*, \hat{\eta}_{0,1}^* \), and \( \hat{\eta}_n^* \) the vectors corresponding to \( \theta_2^*, \eta_1^*, \) and \( \eta^* \) when \( \eta \) is replaced by \( \eta_n \). Equation (6.14) immediately leads to

\[
\hat{\eta}_{1n}^* - \eta_{0,1}^* = o_p (1),
\]

where \( \eta_{0,1}^* = (0^T, d_0)^T \).

Define \( \Pi^* = \{ \theta^* \in \mathbb{R}^{p+q} | \theta \in \Pi \} \) and \( \Pi^*_1 = \{ \theta_1^* \in \mathbb{R}^{s_1^*} : \theta^* \in \Pi^* \} \). It follows from (2.5) and (6.17) that

\[
\eta_{0,1}^* \in \text{int} \ \Pi^*_1 \times D.
\]

Relation (6.20) ensures that there is a small constant \( 0 < \delta_1^* < \min \{ \hat{\delta}_1, r_1^{-1} (M_2^*)^{-2} \} \) such that \( B_{\delta_1^*} (\hat{\eta}_{0,1}^*) \subset \text{int} \ \Pi^*_1 \times D \), where \( r_1 = s_1^* + 1 \) and \( \hat{\delta}_1 \) and \( M_2^* \) are positive constants to be specified later. Also, define

\[
S_n (\eta^*) = \sum_{t=1}^{n} \hat{\varepsilon}^2_t (\eta^*) \text{ and } \bar{\varepsilon}_t (\eta^*) = \varepsilon_t (F_0^{-1} (\eta^*)),
\]

noting that \( \varepsilon_t (F_0^{-1} (\eta^*)) = (1 - B)^{d-d_0} A_1 G_0^{-1} (\theta^* ) B A_2 G_0^{-1} (\theta^* ) B A_1 \theta_0 (B) A_2 \theta_0 (B) \varepsilon_t \), with \( G_0^{-1} (\theta^*) = A^{-1} \theta^* + \theta_{0,pq} \). For ease of exposition, we write \( S_n (\eta^*) = S_n (\eta_1^*, \theta_2^*) \) and \( \bar{\varepsilon}_t (\eta^*) = \bar{\varepsilon}_t (\eta_1^*, \theta_2^*) \). Then, by the mean value theorem for vector-valued functions, one obtains on set \( A_n = \{ \hat{\eta}_{1n}^* \in S_{\delta_1^*} \}, \)

\[
0 = \nabla_1 S_n (\hat{\eta}_{1n}^*, \hat{\theta}_{2n}^*) = \nabla_1 S_n (\eta_{0,1}^*, \hat{\theta}_{2n}^*)
\]

\[
+ \{ \int_0^1 \nabla_1^2 S_n (\eta_{0,1}^* + r (\hat{\eta}_{1n}^* - \eta_{0,1}^*), \hat{\theta}_{2n}^*) dr \} (\hat{\eta}_{1n}^* - \eta_{0,1}^*),
\]

(6.21)
where the integral of a matrix is to be understood component-wise. In view of (6.21),
\[ \nabla_1 S_n(\eta^*) = 2 \sum_{t=1}^{n} \tilde{e}_t(\eta^*) \nabla_1 \tilde{e}_t(\eta^*), \]
and \[ \nabla_1^2 S_n(\eta^*) = 2 \sum_{t=1}^{n} \nabla_1 \tilde{e}_t(\eta^*)(\nabla_1 \tilde{e}_t(\eta^*))^T + 2 \sum_{t=1}^{n} \tilde{e}_t(\eta^*) \nabla_1^2 \tilde{e}_t(\eta^*), \]
it holds that
\[ \sum_{t=1}^{n} \varepsilon_t \nabla_1 \tilde{e}_t(\eta^*) = -\{L(\hat{\eta}_1, \hat{\theta}_{2n}) + Q(\hat{\eta}_1, \hat{\theta}_{2n})\}(\hat{\eta}_1 - \eta_0^*) \text{ on } A_n, \]
where
\[ L(\hat{\eta}_1, \hat{\theta}_{2n}) = \int_0^1 \sum_{t=1}^{n} \nabla_1 \tilde{e}_t(\eta^* + r(\hat{\eta}_1 - \eta_0^*), \hat{\theta}_{2n})(\nabla_1 \tilde{e}_t(\eta^* + r(\hat{\eta}_1 - \eta_0^*), \hat{\theta}_{2n}))^T dr, \]
\[ Q(\hat{\eta}_1, \hat{\theta}_{2n}) = \int_0^1 \sum_{t=1}^{n} \tilde{e}_t(\eta^* + r(\hat{\eta}_1 - \eta_0^*), \hat{\theta}_{2n})^2 \nabla_1^2 \tilde{e}_t(\eta^* + r(\hat{\eta}_1 - \eta_0^*), \hat{\theta}_{2n}^*) dr. \]

For \( \delta > 0 \), define \( S_\delta^* = \{ \eta^* \in \Pi^* \times D| \eta^* \in B_\delta(\eta_0^*) \} \), and \( S^*_\delta,+ = S_\delta^* \cup \{ \lambda \eta^* + (1 - \lambda)(\theta^T, \theta^2, d_0) \eta^* \in S^*_\delta, 0 \leq \lambda \leq 1 \} \). Choose a small enough \( \delta_1 \in (0, 1/2) \). Then, by the compactness of \( S^*_\delta,+ \) (the closure of \( S^*_\delta,+, \)) there exists a set of finite points \( \{ \eta_1^*, \ldots, \eta_l^* \} \subset \tilde{S}_{\delta,+}^* \) and a small positive number \( 0 < \delta_2 < 1/2 - \delta_1 \), depending possibly on \( \Pi^* \) (and \( \Pi \)), such that
\[ \tilde{S}^*_{\delta_1,+} \subset \bigcup_{k=1}^{l} B_{\delta_2}(\eta_k^*). \]
Moreover, for all \( |z| \leq 1 \) and all \( \theta^* \) for which \( (\theta^*, d)^T \in \bigcup_{k=1}^{l} B_{\delta_2}(\eta_k^*) \),
\[ A_{1, G_{\theta^*}}(z) \neq 0 \text{ and } A_{2, G_{\theta^*}}(z) \neq 0. \]

Direct algebraic manipulation gives
\[ \lambda_{\min}(L(\hat{\eta}_1, \hat{\theta}_{2n})) \geq \inf_{\eta^* \in \bigcup_{k=1}^{l} B_{\delta_2}(\eta_k^*)} \lambda_{\min}(\sum_{t=1}^{n} \nabla_1 \tilde{e}_t(\eta^*)(\nabla_1 \tilde{e}_t(\eta^*))^T) \text{ on } A_n, \]
where \( \lambda_{\min}(M) \) denotes the minimum eigenvalue of matrix \( M \). It is shown in Appendix A that for some \( M^*_1 > 0 \),
\[ P\{ \sup_{\eta^* \in \bigcup_{k=1}^{l} B_{\delta_2}(\eta_k^*)} \lambda_{\min}^{\frac{1}{-1}}(\hat{\eta}_1(\eta^*)) > M^*_1 \} = o(1), \]
where \( \hat{\eta}_1(\eta^*) = n^{-1} \sum_{t=1}^{n} \nabla_1 \tilde{e}_t(\eta^*)(\nabla_1 \tilde{e}_t(\eta^*))^T \). In view of (6.25) and (6.26), we can assume without loss of generality that \( L^{-1}(\hat{\eta}_1(\theta^*), \theta^*_2) \) exists on \( A_n \). Therefore, by (6.22),
\[ \|n^{1/2}(\hat{\eta}_1 - \eta_0^*)\|_{L_A} \leq \|nL^{-1}(\hat{\eta}_1, \theta^*_2)\| \|n^{-1/2} \sum_{t=1}^{n} \varepsilon_t \nabla_1 \tilde{e}_t(\eta_0^*, \hat{\theta}_2)\|_{L_A} \]
\[ + \|nL^{-1}(\hat{\eta}_1, \theta^*_2)\| \int_0^1 n^{-1/2} \sum_{t=1}^{n} \varepsilon_t \nabla_1^2 \tilde{e}_t(\eta_0^*, \hat{\theta}_2)dr \]
\[ \times \|\hat{\eta}_1 - \eta_0^*\|_{L_A}, \]
\[ + \|nL^{-1}(\hat{\eta}_1, \theta^*_2)\| \int_0^1 n^{-1/2} \sum_{t=1}^{n} r(\hat{\eta}_1 - \eta_0^*)(\nabla_1 \tilde{e}_t(\eta_0^*, \theta^*_2)\|_{L_A} \]
\[ \times \|n^{1/2}(\hat{\eta}_1 - \eta_0^*)\|_{L_A}, \]
where \( \eta_{1,t,r}^\ast \) satisfies \( \| \eta_{1,t,r}^\ast - \eta_{0,1}^\ast \| \leq r \| \hat{\eta}_{1,n}^\ast - \eta_{0,1}^\ast \| \).

It follows from the Cauchy–Schwarz inequality and Jensen’s inequality that on the set \( A_n \),

\[
\begin{align*}
\| \int_0^1 n^{-1/2} \sum_{t=1}^n \varepsilon_t \nabla_1^2 \hat{\varepsilon}_t (\eta_{0,1}^\ast + r (\hat{\eta}_{1,n}^\ast - \eta_{0,1}^\ast), \theta_{2n}^\ast) \, dr \| \\
\leq r^* \max_{1 \leq i, j \leq r^*} (\max_{\eta^\ast \in S_{i,t}^\ast} n^{-1/2} \sum_{t=1}^n | \varepsilon_t (\nabla_1^2 \hat{\varepsilon}_t (\eta^\ast))_{i,j} | )
\end{align*}
\]

(6.28)

and

\[
\begin{align*}
\| \int_0^1 n^{-1} \sum_{t=1}^n r (\hat{\eta}_{1,n}^\ast - \eta_{0,1}^\ast)^T \nabla_1 \hat{\varepsilon}_t (\eta_{0,1}^\ast + r (\hat{\eta}_{1,n}^\ast - \eta_{0,1}^\ast), \theta_{2n}^\ast) \, dr \| \\
\leq r^* \| \hat{\eta}_{1,n}^\ast - \eta_{0,1}^\ast \| ( \max_{\eta^\ast \in S_{i,t}^\ast} n^{-1/2} \sum_{t=1}^n \| \nabla_1 \hat{\varepsilon}_t (\eta^\ast) \|^2 )^{1/2} \\
\times \{ \max_{1 \leq i, j \leq r^*} (\max_{\eta^\ast \in S_{i,t}^\ast} n^{-1} \sum_{t=1}^n (\nabla_1^2 \hat{\varepsilon}_t (\eta^\ast))_{i,j}^2 )^{1/2} \}.
\end{align*}
\]

(6.29)

By Lemma B.1 of Chan and Ing (2011), (2.13), (A.7)–(A.10), and an argument similar to that used to prove (A.6), there exists \( M^*_2 \geq M^*_1 > 0 \) such that

\[
P( \sup_{\eta^\ast \in \bigcup_{i=1}^m B_{i,z}(\eta_i^\ast)} n^{-1} \sum_{t=1}^n \| \nabla_1 \hat{\varepsilon}_t (\eta^\ast) \|^2 > M^*_2 ) = o(1),
\]

(6.30)

\[
P( \max_{1 \leq i, j \leq r^*} (\sup_{\eta^\ast \in \bigcup_{i=1}^m B_{i,z}(\eta_i^\ast)} n^{-1} \sum_{t=1}^n (\nabla_1^2 \hat{\varepsilon}_t (\eta^\ast))_{i,j}^2 ) > M^*_2 ) = o(1),
\]

(6.31)

\[
E( \sup_{\eta^\ast \in \bigcup_{i=1}^m B_{i,z}(\eta_i^\ast)} \| n^{-1/2} \sum_{t=1}^n \varepsilon_t \nabla_1 \hat{\varepsilon}_t (\eta^\ast) \| ) = O(1),
\]

(6.32)

and

\[
E( \max_{1 \leq i, j \leq r^*} (\sup_{\eta^\ast \in \bigcup_{i=1}^m B_{i,z}(\eta_i^\ast)} \| n^{-1/2} \sum_{t=1}^n \varepsilon_t (\nabla_1^2 \hat{\varepsilon}_t (\eta^\ast))_{i,j} \| ) ) = O(1),
\]

(6.33)

noting that (A.6)–(A.10) are given in Appendix A. Using (6.19) and (6.27)–(6.33), we obtain that for any \( \epsilon > 0 \), there exist \( M^*, N^* > 0 \) such that for all \( n > N^* \),

\[
P( \| n^{1/2} (\hat{\eta}_{1,n}^\ast - \eta_{0,1}^\ast) \| > M^* ) < \epsilon.
\]

(6.34)

Let \( \theta_n = G_0^\ast ([0^T, \theta_{2n}^\ast T]^T) \). Since when \( \hat{\eta}_n \) (defined in (6.1)) exists,

\[
d(\hat{\eta}_n, \theta_n^\ast) \leq (\| \hat{\theta}_n - \theta_n \|^2 + | \hat{d}_n - d_0 |^2 )^{1/2}
\]

(6.35)

\[
\leq (\| A^{-1} \|^2 (\theta_{1,n}^T, \theta_{2n}^T T - (0^T, \theta_{2n}^T T)^T)^2 + | \hat{d}_n - d_0 |^2 )^{1/2}
\]

\[
\leq \max \{ \| A^{-1} \|^2, 1 \} \| \hat{\eta}_n^\ast - \eta_{0,1}^\ast \|
\]

the desired conclusion (2.9) is ensured by (6.1), (6.34), and (6.35).

\[ \square \]
6.3. Proof of Theorem 3.1. The proof of Theorem 3.1 relies heavily on Lemmas 6.2 and 6.3, whose proofs are given in the supplementary material and Appendix A, respectively. Lemma 6.2 establishes uniform moment bounds for linear/quadratic forms of a linear process driven by conditional heteroscedastic errors, which are of independent interest. To state Lemma 6.2, for any \(1 \leq m \leq k\), define \(J(m, k) = \{ (j_1, \ldots, j_m) | j_1 < \cdots < j_m, j_i \in \{1, \ldots, k\}, 1 \leq i \leq m \}\). Moreover, for \(j = (j_1, \ldots, j_m) \in J(m, k)\) and a smooth function \(w = w(\xi_1, \ldots, \xi_t)\), let \(D_j w = \partial^m w/\partial \xi_{j_1}\cdots \partial \xi_{j_m}\).

**Lemma 6.2.** Assume (3.1) and (3.2). Let \(\theta_a = (\theta_{a,1}, \ldots, \theta_{a,k})^T\) be some point in \(\mathbb{R}^k\), \(k \geq 1\), and \(\delta\) be a positive number. For \(t \geq 2\), define \(K_t(\theta) = \sum_{i=1}^{t-1} c_i(\theta) \varepsilon_{t-i} + Q_t(\theta) = \sum_{i=1}^{t-1} d_i(\theta) \varepsilon_{t-i},\) where \(c_i(\theta)\) and \(d_i(\theta)\) are real-valued functions on \(B_\delta(\theta_a)\). Assume for any \(i \geq 1, j = (j_1, \ldots, j_m)^T \in J(m, k)\), and \(1 \leq m \leq k\), \(D_j c_i(\theta)\) are continuous on \(B_\delta(\theta_a)\) and

\[
\sup_{-\infty < i < \infty} E\|w_i\|^{m_1} < \infty,
\]

for some \(m_1 \geq 2\). Then, there exists \(C > 0\) such that for all \(n \geq 2\),

\[
E(\sup_{\theta \in B_\delta(\theta_a)} |n^{-1/2} \sum_{t=2}^{n} K_t(\theta) \varepsilon_t|^{m_1}) \leq C[\{n^{-1} \sum_{t=2}^{n} \sum_{i=1}^{t-1} \max_{j \in J(m, k), 1 \leq m \leq k} \sup_{\theta \in B_\delta(\theta_a)} (D_j c_i(\theta))^2\}^{m_1 / 2} + \{n^{-1} \sum_{t=2}^{n} \sum_{i=1}^{t-1} c_i^2(\theta)\}]^{m_1 / 2}.
\]

Furthermore, if for any \(i, j \geq 1, j = (j_1, \ldots, j_m)^T \in J(m, k),\) and \(1 \leq m \leq k\), \(D_j\{c_i(\theta)d_j(\theta)\}\) are continuous on \(B_\delta(\theta_a)\), then there exists \(C > 0\) such that for all \(n \geq 3\),

\[
E(\sup_{\theta \in B_\delta(\theta_a)} |n^{-1/2} \sum_{t=2}^{n} K_t(\theta) Q_t(\theta) - E(K_t(\theta) Q_t(\theta))|^{m_1}) \leq C[\{(n^{-1} \sum_{i=1}^{n-1} \sum_{t=i+1}^{n} |c_{t-i}(\theta) d_{t-i}(\theta_a)|)^2\}^{m_1 / 2} + \{n^{-1} \sum_{t=2}^{n} \sum_{i=1}^{t-1} \max_{j \in J(m, k), 1 \leq m \leq k} \sup_{\theta \in B_\delta(\theta_a)} |D_j\{c_{t-i}(\theta) d_{t-i}(\theta)\}|^2\}^{m_1 / 2} + \{n^{-1} \sum_{t=2}^{n} \sum_{i=1}^{t-1} \max_{j \in J(m, k), 1 \leq m \leq k} \sup_{\theta \in B_\delta(\theta_a)} |D_j\{c_{t-i}(\theta) d_{t-i}(\theta)\}|^2\}^{m_1 / 2} + \{n^{-1} \sum_{t=2}^{n} \sum_{i=1}^{t-1} \max_{j \in J(m, k), 1 \leq m \leq k} \sup_{\theta \in B_\delta(\theta_a)} |D_j\{c_{t-i}(\theta) d_{t-i}(\theta)\}|^2\}^{m_1 / 2}]^{m_1 / 2}.
\]

Lemma 6.3 plays the same role as Lemma 6.1 in the proof of Theorem 2.1.

**Lemma 6.3.** Under the same assumptions as in Theorem 3.1, (2.9) holds.
PROOF OF THEOREM 3.1. It suffices to prove (2.3) and (2.4) under the assumptions of Theorem 3.1. As indicated in the proof of Theorem 2.1, (2.4) follows from (6.5), (6.6), and (6.8), whereas (2.3) is ensured by (6.12). By making use of Lemma 6.2, we prove (6.5) and (6.6) in Section S1 of the supplementary material. Moreover, (6.8) and (6.12) are immediate consequences of Lemma 6.3 and (6.15), respectively. Note that the proof of (6.15) under the assumptions of Theorem 3.1 is given in Appendix A. Consequently, the desired conclusion follows.

APPENDIX A: PROOFS OF (6.26) AND LEMMA 6.3

PROOF OF (6.26). We first show that for all large $n$ and $1 \leq k \leq l$, there exists $c_k > 0$ such that

\[
(A.1) \quad \inf_{\eta^* \in B_{\delta_k}(\eta^*_k)} \lambda_{\min}(\Gamma_1(\eta^*)) > c_k,
\]

where $\Gamma_1(\eta^*) = E(\hat{\Gamma}_1(\eta^*))$. To prove (A.1), consider

\[
(A.2) \quad C_{\eta^*}(z) := (1 - z)^{d - d_0} A_{1, G_0^{-1} (\theta^*)} (z) A_{2, G_0^{-1} (\theta^*)} (z) A_{1, \theta_0} (z) A_{2, \theta_0} (z)
\]

\[
= (1 - z)^{d - d_0} (z) \{ A_{1, G_0^{-1} (\theta^*)} (z) A_{2, \theta_0} (z) - A_{2, G_0^{-1} (\theta^*)} (z) A_{1, \theta_0} (z) \} A_{2, \theta_0} (z)
\]

\[
+ A_{1, \theta_0} (z) A_{2, \theta_0} (z) A_{1, \theta_0} (z) A_{2, \theta_0} (z).
\]

According to the definition of $\theta^*_i$, we have for $i = 1, \ldots, s^*_1$,

\[
(A.3) \quad \frac{\partial}{\partial \eta^*_{1,i}} C_{\eta^*}(z) \bigg|_{\eta^*_{1,i} = \eta_0^*, \theta^*_{1,i} = \theta_0^*} = A_{1, \theta_0}^{-1} (z) A_{2, \theta_0}^{-1} (z) A_{1, \theta_0} (z) A_{2, \theta_0} (z)
\]

and

\[
(A.4) \quad \frac{\partial}{\partial \eta^*_{1,s^*_1}} C_{\eta^*}(z) \bigg|_{\eta^*_{1,s^*_1} = \eta_0^*, \theta^*_{1,s^*_1} = \theta_0^*} = \log(1 - z),
\]

where $(\eta_0^*, \ldots, \eta_{s^*_1}^*)^T = \eta_1^*$. Therefore, (A.1) follows from (A.2)–(A.4) and an argument similar to (0.3) and (0.4) in the supplementary material of Chan, Huang and Ing (2013).

Write

\[
(A.5) \quad \tilde{\epsilon}_t(\eta^*) = \sum_{s=0}^{t-1} b_s(\eta^*) \epsilon_{t-s},
\]

where $b_0(\eta^*) = 1$. Then $\nabla_1 \tilde{\epsilon}_t(\eta^*) = \sum_{s=0}^{t-1} b_{s,i}(\eta^*) \epsilon_{t-s}$, where $b_{s,i}(\eta^*) = \partial b_s(\eta^*) / \partial \eta^*_{1,i}$. Recall the definition of $D_i$ given in Section 6.3. It is clear that $b_{s,i}(\eta^*)$ has continuous partial derivatives $D_i b_{s,i}(\eta^*)$. Combining (A.1) with

\[
\inf_{\eta^* \in B_{\delta_k}(\eta^*_k)} \lambda_{\min}(\hat{\Gamma}_1(\eta^*)) \geq \inf_{\eta^* \in B_{\delta_k}(\eta^*_k)} \lambda_{\min}(\Gamma_1(\eta^*)) - \sup_{\eta^* \in B_{\delta_k}(\eta^*_k)} \| \Gamma_1(\eta^*) - \hat{\Gamma}_1(\eta^*) \|,
\]

we obtain the desired conclusion.
(2.13), Lemma B.1 of Chan and Ing (2011), and Markov’s inequality, one obtains for $M_{1,k}^* > 2/c_k$,

$$P\left\{ \sup_{\eta^* \in B_{1,k}^*} \lambda_{\min}(\Gamma_1(\eta^*)) > M_{1,k}^* \right\} \leq P\left( \sup_{\eta^* \in B_{1,k}^*} \|\Gamma_1(\eta^*) - \hat{\Gamma}_1(\eta^*)\| > c_k/2 \right)$$

\[
\leq (4r_1^4/c_k^2) \max_{1 \leq i,j \leq r_1^*} E\left[ \sup_{\eta^* \in B_{1,k}^*} |n^{-1} \sum_{t=1}^n (\nabla_1 \hat{\epsilon}_t(\eta^*))_i (\nabla_1 \hat{\epsilon}_t(\eta^*))_j \right] \\
- E\{ (\nabla_1 \hat{\epsilon}_t(\eta^*))_i (\nabla_1 \hat{\epsilon}_t(\eta^*))_j \} \right|^2 \]  \tag{A.6}

In view of (A.7) and (A.8), it follows that for all $s \geq 1$ and $1 \leq k \leq l$,

\[
\max_{1 \leq i,j \leq r_1^*} \sup_{\eta^* \in B_{1,k}^*} |b_{s,i}(\eta^*)| \leq C(\log(s + 1)) s^{-1+\delta_1+\delta_2}, \tag{A.7}
\]

and

\[
\max_{1 \leq i,j \leq r_1^*} \sup_{\eta^* \in B_{1,k}^*} |D_j b_{s,i}(\eta^*)| \leq C(\log(s + 1)) s^{-1+\delta_1+\delta_2}. \tag{A.8}
\]

In view of (A.7) and (A.8), it follows that for all $1 \leq i,j \leq r_1^*$, $(\sum_{w=1}^{\infty} S^{(k)}_{w,w}(i,j))^2$ and $(\sum_{w=1}^{\infty} V^{(k)}_{w,v-u,w}(i,j))^2$ are bounded by some constant $C$, and $(\sum_{w=1}^{\infty} S^{(k)}_{w,v+u-w}(i,j))^2$, $(\sum_{w=1}^{\infty} V^{(k)}_{w,v+u-v}(i,j))^2$, $(\sum_{w=1}^{\infty} V^{(k)}_{w,v+u-v}(i,j))^2$, and $(\sum_{w=1}^{\infty} V^{(k)}_{w,v+u-v}(i,j))^2$ are bounded by $C(v-u)^{-2+2(\delta_1+\delta_2)}$. These bounds together with (A.6) and $\delta_1 + \delta_2 < 1/2$ yield

$$P\left\{ \sup_{\eta^* \in B_{1,k}^*} \lambda_{\min}(\hat{\Gamma}_1(\eta^*)) > M_{1,k}^* \right\} \leq Cn^{-1+2(\delta_1+\delta_2)} = o(1).$$

Thus (6.26) holds with $M_{1,k}^* = \max_{1 \leq k \leq l} M_{1,k}^*$. \hfill \ensuremath{\square}

**Remark A.1.** Equation (A.5) implies that $\nabla_1^2 \hat{\epsilon}_t(\eta^*)_{i,j} = \sum_{s=1}^{t-1} c_{s,i,j}(\eta^*)_{t-s}$, where $c_{s,i,j}(\eta^*) = \partial^2 b_s(\eta^*)/\partial \eta_{t-s}^* \partial \eta_{t-s}^*$. By an argument similar to that used to prove (A.7) and (A.8), we have

$$\max_{1 \leq i,j \leq r_1^*} \sup_{\eta^* \in B_{1,k}^*} |c_{s,i,j}(\eta^*)| \leq C(\log(s + 1)) s^{-1+\delta_1+\delta_2}. \tag{A.9}$$
and

\[
\max_{1 \leq i, j \leq r} \max_{\eta \in \Pi} \sup_{i \leq r} |D_j c_{s,j}(\eta^*)| \leq C((\log(s + 1))^3 s^{1+\tilde{\delta}_1+\tilde{\delta}_2}).
\]

**Proof of Lemma 6.3.** Let \((p, q)\) satisfy \(p_0 \leq p \leq P\) and \(q_0 \leq q \leq Q\). If (6.15), (6.26), and (6.30)–(6.33) hold under the assumptions of Theorem 3.1, then the desired conclusion follows from the same argument as that in the proof of Lemma 6.1. Using Lemma 6.2 to replace Lemma B.1 of Chan and Ing (2011) in the proofs of (6.26) and (6.30)–(6.33), one can easily establish these equations under the assumptions of Theorem 3.1. The details are omitted. Therefore, it remains to prove (6.15).

Let \(z_t = z_t(d_0) = (1 - B)^{d_0} y_t\). Assume (A.11)–(A.18), which are listed as follows:

(A.11)

\[
E| \sum_{t=1}^{n} z_t^2 - E(z_t^2) | = O(n^{1/2}),
\]

(A.12)

\[
E| \sum_{r=j+1}^{k} \sum_{l=r-j+1}^{n-j} \{z_t z_{l-r+j} - E(z_t z_{l-r+j})\} \]
\[\leq C(k-j)^{1/2}(n-j)^{1/2}, \text{ for all } 0 \leq j \leq n-2, j \leq k \leq n,\]

(A.13)

\[
E| \sum_{t=1}^{n} \varepsilon_t^2 - E(\varepsilon_t^2) | = O(n^{1/2}),
\]

(A.14)

\[
E| \sum_{r=j+1}^{k} \sum_{l=r-j+1}^{n-j} \{\varepsilon_t \varepsilon_{l-r+j} - E(\varepsilon_t \varepsilon_{l-r+j})\} \]
\[\leq C(k-j)^{1/2}(n-j)^{1/2}, \text{ for all } 0 \leq j \leq n-2, j \leq k \leq n,\]

(A.15)

\[
E| \sum_{k=1}^{j} z_{n-k} |^2 \leq C_j, \text{ for all } 0 \leq j \leq n-1,
\]

(A.16)

\[
E| \sum_{k=0}^{j} \varepsilon_{s-k} |^2 \leq C_j, \text{ for all } 0 \leq j \leq s \leq n,
\]

(A.17)

\[
E| \sum_{j=1}^{n} \varepsilon_j^2 | = O(n),
\]

and for some small \(\xi, \epsilon_1 > 0\),

(A.18)

\[
\lim_{n \to \infty} P(\inf_{\eta \in \Pi} \sup_{d_0 \in [L, d_0-1/2-\xi]} \frac{1}{n^{d_0-d+1/2}} \sum_{t=1}^{n} \varepsilon_t(\eta_{pq}))^2 > \epsilon_1) = 1,
\]

noting that \(L\) is the prescribed lower bound of \(d\). Then, (6.15) follows from these equations, (6.16), and the argument used in the proof of Theorem 2.1 of Hualde and Robinson (2011).

Since the proofs of (A.11)–(A.17) are similar, we choose to present the proof of (A.12) in Section S3 of the supplementary material, while omitting the proofs of the others. The proof of (A.18) is given below.
Thus, (A.22) follows from (A.23)–(A.25).

(A.25) \[ E(\varepsilon_t \varepsilon_m | F_t) = 0 \quad \text{a.s. and } E(\varepsilon_t \varepsilon_m) = 0. \]

For \( l = m \), it follows from (3.1), (3.2), (3.6), Burkholder’s inequality, and Minkowski’s inequality that

\[
E|E(\varepsilon_t^2 | F_t) - E(\varepsilon_t^2)|^2 = E|E(\sum_{s=0}^{\infty} a_s^T w_{t-s} | F_t)|^2 =
\]

\[
= E| \sum_{s=1-t}^{\infty} a_s^T w_{t-s} |^2 \leq C \sum_{s=1-t}^{\infty} \| a_s \|^2 = O((l - t)^{-2a + 1}),
\]

which converges to 0 as \( l \) grows to \( \infty \). Thus, A(ii) is ensured by (A.19) and (A.20).

It remains to prove A(iii). Given any real vector \((a_1, \ldots, a_n)\), we have

(A.21) \[ \text{var}\left( \sum_{t=1}^{n} a_t \varepsilon_t \right) = E(\sum_{t=1}^{n} a_t \varepsilon_t)^2 = \sigma^2 \sum_{t=1}^{n} a_t^2. \]

In view of (4) and (5) of Hosoya (2005) and (A.21), A(iii) follows if

(A.22) \[ E|\sum_{t=1}^{n} a_t \varepsilon_t|^4 \leq C(\sum_{t=1}^{n} a_t^2)^2. \]

Simple algebraic manipulations give

(A.23) \[ E|\sum_{t=1}^{n} a_t \varepsilon_t|^4 = E|\sum_{t=1}^{n} a_t \varepsilon_t|^2 - E(\sum_{t=1}^{n} a_t \varepsilon_t)^2 + E(\sum_{t=1}^{n} a_t \varepsilon_t)^2|^2 \]

\[ \leq C[E\left| \sum_{t=1}^{n} a_t^2 (\varepsilon_t^2 - E(\varepsilon_t^2)) \right|^2 + E\left| \sum_{t_1=2}^{n} a_{t_1} a_{t_2} \varepsilon_{t_1} \varepsilon_{t_2} \right|^2 + (\sum_{t=1}^{n} a_t^2)^2]. \]

By (3.1), (3.2), (3.6), Burkholder’s inequality, and Minkowski’s inequality, we have

(A.24) \[ E\left| \sum_{t=1}^{n} a_t^2 (\varepsilon_t^2 - E(\varepsilon_t^2)) \right|^2 \leq \left| \sum_{t=1}^{n} \{ E[a_t^2 (\varepsilon_t^2 - E(\varepsilon_t^2)))]^2 \right|^{1/2} \leq C(\sum_{t=1}^{n} a_t^2)^2, \]

and

(A.25) \[ E\left| \sum_{t_1=2}^{n} \sum_{t_2=1}^{t_1-1} a_{t_1} a_{t_2} \varepsilon_{t_1} \varepsilon_{t_2} \right|^2 \leq C(\sum_{t=1}^{n} a_t^2)^2. \]

Thus, (A.22) follows from (A.23)–(A.25). \( \square \)

REFERENCES


