CONSISTENT ORDER SELECTION FOR ARFIMA PROCESSES

BY HSUEH-HAN HUANG[1], NGAI HANG CHAN[2], KUN CHEN[3], AND CHING-KANG ING[1]

[1]*National Tsing Hua University*

[2]*Chinese University of Hong Kong*

[3]*Southwestern University of Finance and Economics*

Estimating the orders of the autoregressive fractionally integrated moving average (ARFIMA) model has been a long-standing problem in time series analysis. This paper tackles this 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 an important 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 ∞.

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). Likelise, estimating $(p_0 + d_0, q_0)$ for ARIMA models, see Guo, Chen and Zhang (1989) and

AMS 2000 subject classifications: Primary 62M10; secondary 62M20

Keywords and phrases: ARFIMA models, Bayesian information criterion, conditional heteroscedastic errors, long-memory and non-stationary time series, order selection

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

$$d_0 \geq -0.5,\ d_0 \notin \{-0.5, 0.5, 1.5, 2.5, \ldots\}. \tag{1.1}$$

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,

$$(1 - \alpha_{0,1}B - \cdots - \alpha_{0,p_0}B^{p_0})(1-B)^{d_0} y_t = (1 - \beta_{0,1}B - \cdots - \beta_{0,q_0}B^{q_0})\varepsilon_t, \tag{1.2}$$

where p_0 and q_0 are unknown non-negative integers; B is the back-shift operator; $\{\varepsilon_t\}$ is a sequence of random disturbances with mean 0 and variance σ_ε^2; and $\alpha_{0,i}$, $\beta_{0,j}$, and d_0 are unknown coefficients satisfying $d_0 \in \mathbb{R}$,

$$1 - \sum_{j=1}^{p_0} \alpha_{0,j} z^j \neq 0,\ 1 - \sum_{j=1}^{q_0} \beta_{0,j} z^j \neq 0 \text{ for } |z| \leq 1, \tag{1.3}$$

in which $\sum_a^b \cdot = 0$ if $a > b$,

$$1 - \sum_{j=1}^{p_0} \alpha_{0,j} z^j \text{ and } 1 - \sum_{j=1}^{q_0} \beta_{0,j} z^j \text{ have no common zeros,} \tag{1.4}$$

and

$$|\alpha_{0,p_0}| > 0,\ |\beta_{0,q_0}| > 0, \tag{1.5}$$

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, $\hat{\boldsymbol{\eta}}_{n,pq}$, which is the minimizer of $\sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\eta}_{pq})$ over $\boldsymbol{\eta}_{pq} = (\boldsymbol{\theta}_{pq}^\top, d)^\top = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q, d)^\top$ in $\Pi_{pq} \times D \subset \mathbb{R}^{p+q+1}$, where

$$\varepsilon_t(\boldsymbol{\eta}_{pq}) = A_{1,\boldsymbol{\theta}_{pq}}(B) A_{2,\boldsymbol{\theta}_{pq}}^{-1}(B)(1-B)^d y_t, \tag{1.6}$$

with

$$A_{1,\boldsymbol{\theta}_{pq}}(z) = 1 - \sum_{j=1}^{p} \alpha_j z^j,\ A_{2,\boldsymbol{\theta}_{pq}}(z) = 1 - \sum_{j=1}^{q} \beta_j z^j,$$

and $\Pi_{pq} \times D$ is the parameter space to be specified in the next section. Note that $\boldsymbol{\theta}_{pq}$ vanishes if $p = q = 0$, and $\boldsymbol{\theta}_{pq} = (\alpha_1, \ldots, \alpha_p)^\top$ if $p \geq 1$ and $q = 0$, and $(\beta_1, \ldots, \beta_q)^\top$ if $q \geq 1$ and $p = 0$. Consider a BIC-type criterion,

$$\phi(p,q) = n \log \hat{\sigma}_{pq}^2 + (p+q)p(n), \tag{1.7}$$

where

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

is the CSS estimate of σ^2_ε when model (p,q) is postulated, and $p(n)$ is a penalty term obeying

$$\lim_{n\to\infty} p(n) = \infty \text{ and } \lim_{n\to\infty} \frac{p(n)}{n} = 0. \tag{1.8}$$

Note that $n\log\hat{\sigma}^2_{pq}$ in (1.7) is not exactly $-2\log L(\hat{\boldsymbol{\eta}}_{n,pq})$, where $L(\boldsymbol{\eta}_{pq}) = L(\boldsymbol{\eta}_{pq}|y_1,\ldots,y_n)$ is the Gaussian conditional likelihood. However, their difference, depending only on n, has no impact on order selection results. Let

$$(\hat{p}_n,\hat{q}_n) = \arg\min_{0\le p\le P, 0\le q\le Q} \phi(p,q). \tag{1.9}$$

The main goal of this paper is to show that

$$\lim_{n\to\infty} \mathbb{P}\{(\hat{p}_n,\hat{q}_n) = (p_0,q_0)\} = 1. \tag{1.10}$$

The major difficulty in proving (1.10) is to handle the asymptotic behavior of $\hat{\sigma}^2_{pq}$ 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

$$A_{2,\boldsymbol{\theta}_{pq}}(z)A^{-1}_{1,\boldsymbol{\theta}_{pq}}(z) = A_{2,\boldsymbol{\theta}_{p_0q_0}}(z)A^{-1}_{1,\boldsymbol{\theta}_{p_0q_0}}(z). \tag{1.11}$$

Therefore, $\hat{\boldsymbol{\eta}}_{n,pq}$ does not possess a probability limit, and hence $\hat{\sigma}^2_{pq}$ 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

$$d(\hat{\boldsymbol{\eta}}_{n,pq}, S^+_{0,pq}) = O_p(n^{-1/2}),\ p_0 < p \le P, q_0 < q \le Q, \tag{1.12}$$

where $S^+_{0,pq} \subset \mathbb{R}^{p+q+1}$, defined in Section 2, contains all points in the parameter space satisfying (1.11), and $d(x,S) := \inf_{w\in S}\|x-w\|$, with $\|\cdot\|$ denoting the Euclidean norm. Equation (1.12) essentially says that while $\hat{\boldsymbol{\eta}}_{n,pq}$ does not have a probability limit, its distance to a set of parameters equivalent to the true parameter converges to 0 at a rate of $1/\sqrt{n}$. This property enables us to show that for any two-sided overfitted candidate,

$$\hat{\sigma}^2_{pq} - \hat{\sigma}^2_{p_0q_0} = O_p(n^{-1}), \tag{1.13}$$

which, in turn, becomes the key ingredient in proving (1.10).

After obtaining (1.10) in ARFIMA models with independent errors, we focus on extending the result to ARFIMA models with conditional heteroscedastic errors. Our assumptions on the error terms are fairly general and easily satisfied by the generalized autoregressive conditional heteroskedasticity (GARCH) model (Bollerslev (1986)) as well as the GJR-GARCH model (Glosten, Jagannathan and Runkle (1993)). We show that (1.12) and (1.13) are still true, and hence (1.10) remains valid. Since we allow $d_0 \in \mathbb{R}$ and $\{\varepsilon_t\}$ to be conditionally heteroscedastic, this is one of the most comprehensive results to date on order selection consistency established for the ARFIMA model.

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. The proofs of all theorems are provided in Section 5. We conclude 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

(2.1) $$D = [L, U], \text{ for some } -\infty < L < U < \infty,$$

and Π_{PQ} is a compact set in $\mathbb{R}^{P+Q}$ with element $\boldsymbol{\theta}_{PQ} = (\alpha_1, \ldots, \alpha_P, \beta_1, \ldots, \beta_Q)^\top$ satisfying the stationarity condition,

(2.2) $$A_{1,\boldsymbol{\theta}_{PQ}}(z) \neq 0, A_{2,\boldsymbol{\theta}_{PQ}}(z) \neq 0 \text{ for all } |z| \leq 1.$$

For candidate model (p, q), the parameter space is $\Pi_{pq} \times D$, where

$$\Pi_{pq} \equiv \{\boldsymbol{\theta}_{pq} = (\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)^\top | (\alpha_1, \ldots, \alpha_p, 0, \ldots, 0, \beta_1, \ldots, \beta_q, 0, \ldots, 0)^\top \in \Pi_{PQ}\},$$

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

$$(\hat{\boldsymbol{\theta}}_{n,pq}^\top, \hat{d}_{n,pq})^\top = \hat{\boldsymbol{\eta}}_{n,pq} = \arg \min_{\boldsymbol{\eta}_{pq} \in \Pi_{pq} \times D} \sum_{t=1}^{n} \varepsilon_t^2(\boldsymbol{\eta}_{pq}).$$

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

Equation (1.10) is ensured by

(2.3) $$\lim_{n\to\infty} \mathbb{P}(\hat{p}_n < p_0 \text{ or } \hat{q}_n < q_0) = 0$$

and

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

Throughout the paper, we assume that

(2.5) $$(\alpha_{0,1}, \ldots, \alpha_{0,p_0}, 0, \ldots, 0, \beta_{0,1}, \ldots, \beta_{0,q_0}, 0, \ldots, 0, d_0)^\top \in \text{int}\, \Pi_{PQ} \times D,$$

and denote $\boldsymbol{\eta}_0 = (\boldsymbol{\theta}_0^\top, d_0)^\top$ with $\boldsymbol{\theta}_0^\top = (\alpha_{0,1}, \ldots, \alpha_{0,p_0}, \beta_{0,1}, \ldots, \beta_{0,q_0})$.

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 δ for which

(2.6) $$\hat{\boldsymbol{\eta}}_{n,pq} \notin B_\delta(\boldsymbol{\eta}_0),$$

where $B_\delta(\boldsymbol{\eta}_0)$ is the open ball of radius δ centered at $\boldsymbol{\eta}_0$, and with a slight abuse of notation, $\hat{\boldsymbol{\eta}}_{n,pq}$ and $\boldsymbol{\eta}_0$ 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 (5.11)–(5.13) that for $p < p_0$ or $q < q_0$,

(2.7) $$\lim_{n\to\infty} \mathbb{P}(\hat{\sigma}_{pq}^2 - \hat{\sigma}_{p_0 q_0}^2 \leq c) = 0,$$

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

$$\Pi_{pq}^+ = \{\boldsymbol{\theta}_{pq} \in \mathbb{R}^{p+q} | (\alpha_1, \ldots, \alpha_p, 0, \ldots, 0, \beta_1, \ldots, \beta_q, 0, \ldots, 0)^\top \text{ satisfies (2.2)}\}$$

and

(2.8) $$S_{0,pq}^+ = \{(\boldsymbol{\theta}_{pq}^\top, d)^\top \in \Pi_{pq}^+ \times \{d_0\} | \boldsymbol{\theta}_{pq} \text{ obeys (1.11)}\},$$

noting that $\Pi_{pq} \subset \Pi_{pq}^{+}$ and $S_{0,pq}^{+}$ 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

$$d(\hat{\boldsymbol{\eta}}_{n,pq}, S_{0,pq}^{+}) = O_p(n^{-1/2}),\ p_0 \le p \le P,\ q_0 \le q \le Q, \tag{2.9}$$

which is a slightly stronger version of (1.12). Indeed, for the case of $p = p_0$ and $q = q_0$, $p = p_0$ and $q > q_0$, or $p > p_0$ and $q = q_0$, $S_{0,pq}^{+}$ only contains one point and (2.9) has already been established by Hualde 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 $\hat{\boldsymbol{\eta}}_{n,pq}$ 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, $\boldsymbol{\eta}_{pq}^{\star} = F_0(\boldsymbol{\eta}_{pq})$, contains $\max\{p_0 + q, q_0 + p\} + 1$ identifiable components in the sense that the first $\max\{p_0 + q, q_0 + p\}$ components, $\boldsymbol{\theta}_{1,pq}^{\star}$, and the last component, d, of $\boldsymbol{\eta}_{pq}^{\star}$ are the same for all $\boldsymbol{\eta}_{pq} \in S_{0,pq}^{+}$. Note that $F_0(\cdot)$ is a one-to-one linear transformation depending on $\boldsymbol{\eta}_0$, which is defined in (5.18) of Section 5.2. Denote $\boldsymbol{\eta}_{1,pq}^{\star} \equiv (\boldsymbol{\theta}_{1,pq}^{\star\top}, d)^{\top}$ by $\boldsymbol{\eta}_{0,1,pq}^{\star}$ when $\boldsymbol{\eta}_{pq} \in S_{0,pq}^{+}$. Also, let the subvector of $\hat{\boldsymbol{\eta}}_{pq}^{\star} \equiv F_0(\hat{\boldsymbol{\eta}}_{n,pq})$ corresponding to $\boldsymbol{\eta}_{0,1,pq}^{\star}$ be denoted by $\hat{\boldsymbol{\eta}}_{1,pq}^{\star}$. The uniqueness of $\boldsymbol{\eta}_{0,1,pq}^{\star}$ allows one to obtain

$$\|\hat{\boldsymbol{\eta}}_{1,pq}^{\star} - \boldsymbol{\eta}_{0,1,pq}^{\star}\| = O_p(n^{-1/2}),\ p_0 \le p \le P,\ q_0 \le q \le Q, \tag{2.10}$$

by analyzing $\sum_{t=1}^{n} \nabla_1 \varepsilon_t^2(F_0^{-1}(\hat{\boldsymbol{\eta}}_{pq}^{\star})) - \sum_{t=1}^{n} \nabla_1 \varepsilon_t^2(F_0^{-1}(\boldsymbol{\eta}_{pq}^{\star}))$ based on the mean value theorem, and establishing uniform moment/probability bounds (see (5.30)–(5.33)) for

$$\begin{gathered} n^{-1/2} \sum_{t=1}^{n} \nabla_1 \varepsilon_t(F_0^{-1}(\boldsymbol{\eta}_{pq}^{\star}))\varepsilon_t,\ n^{-1/2} \sum_{t=1}^{n} \nabla_1^2 \varepsilon_t(F_0^{-1}(\boldsymbol{\eta}_{pq}^{\star}))\varepsilon_t, \\ n^{-1} \sum_{t=1}^{n} \|\nabla_1 \varepsilon_t(F_0^{-1}(\boldsymbol{\eta}_{pq}^{\star}))\|^2,\ n^{-1} \sum_{t=1}^{n} \mathrm{tr}\{[\nabla_1^2 \varepsilon_t(F_0^{-1}(\boldsymbol{\eta}_{pq}^{\star}))]^2\}, \end{gathered} \tag{2.11}$$

where tr$(\cdot)$ stands for the trace operator and, for a twice differentiable function $f(\cdot)$ on $\mathbb{R}^{p+q+1}$, $\nabla_1 f(\boldsymbol{\eta}_{pq}^{\star}) = \frac{\partial}{\partial \boldsymbol{\eta}_{1,pq}^{\star}} f(\boldsymbol{\eta}_{pq}^{\star})$ and $\nabla_1^2 f(\boldsymbol{\eta}_{pq}^{\star}) = \frac{\partial^2}{\partial \boldsymbol{\eta}_{1,pq}^{\star} \partial \boldsymbol{\eta}_{1,pq}^{\star\top}} f(\boldsymbol{\eta}_{pq}^{\star})$. Equation (2.10) serves as an important vehicle for deriving (2.9). By making use of (2.9), we obtain

$$\begin{aligned} \hat{\sigma}_{pq}^2 - \hat{\sigma}_{p_0 q_0}^2 &= n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\hat{\boldsymbol{\eta}}_{n,pq}) - n^{-1} \sum_{t=1}^{n} \varepsilon_t^2(\hat{\boldsymbol{\eta}}_{n,p_0 q_0}) \\ &= O_p(n^{-1}),\ p_0 \le p \le P,\ q_0 \le q \le Q, \end{aligned} \tag{2.12}$$

leading immediately to (2.4). For the details on the proofs of (2.3), (2.4), (2.9), and (2.10), see Sections 5.1 and 5.2. 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), *ε_t's are independent, and*

$$\sup_{-\infty < t < \infty} \mathbb{E}|\varepsilon_t|^4 < \infty. \tag{2.13}$$

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 σ-fields. We further assume that $\{\varepsilon_t^2\}$ admits an infinite-order moving average representation,

$$\varepsilon_t^2 - \mathbb{E}(\varepsilon_1^2) = \sum_{s=0}^{\infty} \mathbf{a}_s^\top \boldsymbol{w}_{t-s}, \tag{3.1}$$

in which $\mathbf{a}_s$ are l-dimensional real vectors for some $l \geq 1$,

$$\|\mathbf{a}_s\| = O((s+1)^{-\iota}), \text{ for some } \iota > 1, \tag{3.2}$$

and $\{\boldsymbol{w}_t, \mathcal{F}_t\}$ 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,

$$\varepsilon_t = \sigma_t z_t,\ \sigma_t^2 = \varphi_{0,0} + \sum_{i=1}^{p_0'} \varphi_{0,i}\varepsilon_{t-i}^2 + \sum_{j=1}^{q_0'} \psi_{0,j}\sigma_{t-j}^2 + \sum_{k=1}^{p_0'} \zeta_{0,k}\varepsilon_{t-k}^2 I_{\{\varepsilon_{t-k}<0\}}, \tag{3.3}$$

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

$$\sum_{i=1}^{p_0'} \varphi_{0,i} + \sum_{j=1}^{q_0'} \psi_{0,j} + \sum_{k=1}^{p_0'} \frac{\zeta_{0,k}}{2} < 1. \tag{3.4}$$

According to (3.3), we express ε_t^2 as

$$\varepsilon_t^2 = \varphi_{0,0} + \sum_{i=1}^{\max\{p_0',q_0'\}} (\varphi_{0,i} + \psi_{0,i} + \frac{\zeta_{0,i}}{2})\varepsilon_{t-i}^2 + w_{1,t} - \sum_{j=1}^{q_0'} \psi_{0,j} w_{1,t-j} + \sum_{k=1}^{p_0'} \zeta_{0,k} w_{2,t-k}, \tag{3.5}$$

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 $\{(w_{1,t}, w_{2,t})^\top\}$ 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 σ-field generated by $\{z_t, z_{t-1}, \ldots\}$, it can be shown that (3.1) and (3.2) hold with $l = 2$, $\boldsymbol{w}_t = (w_{1,t}, w_{2,t})^T$, and $\mathbf{a}_s = (b_s, c_s)^T$, where b_j and c_j, respectively, satisfy

$$\sum_{j=0}^{\infty} b_j z^j = \frac{1 - \sum_{j=1}^{q_0'} \psi_{0,j} z^j}{1 - \sum_{i=1}^{\max\{p_0',q_0'\}} (\varphi_{0,i} + \psi_{0,i} + \frac{\zeta_{0,i}}{2}) z^j}$$

and

$$\sum_{j=0}^{\infty} c_j z^j = \frac{\sum_{j=1}^{p_0'} \zeta_{0,j} z^j}{1 - \sum_{i=1}^{\max\{p_0',q_0'\}} (\varphi_{0,i} + \psi_{0,i} + \frac{\zeta_{0,i}}{2}) z^j}.$$

Moreover, since $|b_j|$ and $|c_j|$ decay exponentially as j increases, (3.2) is valid for arbitrarily large values of ι. 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 $l = 1$, $\boldsymbol{w}_t = w_{1,t}$, and $\mathbf{a}_s = b_s$.

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*

$$\sup_{-\infty<t<\infty} \mathbb{E}\|\boldsymbol{w}_t\|^2 < \infty. \tag{3.6}$$

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 $\{\varepsilon_t, \mathcal{F}_t\}$ is a martingale difference sequence obeying

$$\mathbb{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \mathbb{E}(\varepsilon_1^2) = \sigma_\varepsilon^2 \text{ a.s.} \tag{3.7}$$

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

$$\mathbb{E}|\sigma_1|^4 < \infty, \tag{3.8}$$

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, (5.30)–(5.33), established for (2.11) in the case of independent errors are no longer applicable. To alleviate this difficulty, Lemma 5.2 extends the uniform moment bounds in Lemma B.1 of Chan and Ing (2011) to linear processes driven by conditionally heteroscedastic errors, thereby generalizing (5.30)–(5.33) 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
$$\hat{r}_n^{(1)} = \arg\min_{0\le r\le R} \phi(r,r),$$
where $R = \max\{P, Q\}$.

2: Estimate p_0 and q_0 using
$$\hat{p}_n^{(1)} = \arg\min_{0\le p\le \hat{r}_n^{(1)}} \phi(p, \hat{r}_n^{(1)})$$
and
$$\hat{q}_n^{(1)} = \arg\min_{0\le q\le \hat{r}_n^{(1)}} \phi(\hat{r}_n^{(1)}, q).$$

The consistency of RBIC is confirmed in the next corollary.

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

$$\lim_{n\to\infty} \mathbb{P}\{(\hat{p}_n^{(1)}, \hat{q}_n^{(1)}) = (p_0, q_0)\} = 1. \tag{3.9}$$

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 ε_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 $\hat{p}_{(l)}$, $\hat{q}_{(l)}$, and $\hat{d}_{(l)}$ the estimators of p_0, q_0, and d_0 obtained in the l-th simulation, where $1 \le l \le M$. Note that $\hat{d}_{(l)}$ is derived from model $(\hat{p}_{(l)}, \hat{q}_{(l)})$. We compute the following performance measures,

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

$$\text{Frequency of exact selection (Ext)} = \sum_{l=1}^{M} \mathbf{1}(\hat{p}_{(l)} = p_0, \hat{q}_{(l)} = q_0),$$

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

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

TABLE 1
Simulation results of RBIC in model (I) with independent errors.

n		d=-0.5	0	0.25	0.5	0.75	1	1.5
250	Over	65	34	36	56	26	28	60
	Ext	935	962	958	941	970	966	937
	Under	0	4	6	3	4	6	3
	d-MAE	0.076	0.060	0.065	0.072	0.054	0.061	0.076
500	Over	62	16	22	62	14	19	78
	Ext	938	984	978	938	986	981	922
	Under	0	0	0	0	0	0	0
	d-MAE	0.055	0.036	0.039	0.053	0.036	0.041	0.061

TABLE 2
Simulation results of RBIC in model (II) with independent errors.

n		d=-0.5	0	0.25	0.5	0.75	1	1.5
250	Over	4	3	3	3	5	6	0
	Ext	852	910	904	898	919	915	905
	Under	144	87	93	99	76	79	95
	d-MAE	0.198	0.150	0.154	0.157	0.146	0.152	0.161
500	Over	2	2	0	2	2	1	2
	Ext	976	985	987	978	984	986	976
	Under	22	13	13	20	14	13	22
	d-MAE	0.143	0.103	0.103	0.110	0.102	0.108	0.135

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,

$$\text{ARCH } (1) : \varepsilon_t = \sigma_t z_t,\ \sigma_t^2 = 0.4 + 0.5\varepsilon_{t-1}^2,$$

$$\text{GARCH } (1,1) : \varepsilon_t = \sigma_t z_t,\ \sigma_t^2 = 0.4 + 0.2\varepsilon_{t-1}^2 + 0.7\sigma_{t-1}^2,$$

$$\text{GARCH } (2,2) : \varepsilon_t = \sigma_t z_t,\ \sigma_t^2 = 0.4 + 0.3\varepsilon_{t-1}^2 + 0.2\varepsilon_{t-2}^2 + 0.2\sigma_{t-1}^2 + 0.1\sigma_{t-2}^2,$$

$$\text{GJR-GARCH } (1,1) : \varepsilon_t = \sigma_t z_t,\ \sigma_t^2 = 0.4 + 0.2\varepsilon_{t-1}^2 + 0.6\sigma_{t-1}^2 + 0.1\varepsilon_{t-1}^2 I_{\{\varepsilon_{t-1}<0\}},$$

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 791–875 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 is 500; we conclude that the finite-sample behavior of RBIC concurs with the asymptotic results developed in Section 3.

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

5.1. *Proof of Theorem 2.1.*

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

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

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

n		d=-0.5	0	0.25	0.5	0.75	1	1.5
250	Over	13	16	15	13	11	12	8
	Ext	808	869	858	854	875	875	868
	Under	179	115	127	133	114	113	124
	d-MAE	0.215	0.179	0.184	0.183	0.175	0.179	0.184
500	Over	4	7	6	8	7	8	5
	Ext	939	958	962	948	952	954	946
	Under	57	35	32	44	41	38	49
	d-MAE	0.162	0.121	0.120	0.129	0.124	0.130	0.154

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

n		d=-0.5	0	0.25	0.5	0.75	1	1.5
250	Over	17	17	15	14	18	19	15
	Ext	813	856	853	848	863	865	858
	Under	170	127	132	138	119	116	127
	d-MAE	0.218	0.180	0.185	0.187	0.177	0.180	0.189
500	Over	11	12	12	12	12	12	10
	Ext	933	954	955	944	954	950	946
	Under	56	34	33	44	34	38	44
	d-MAE	0.164	0.123	0.123	0.133	0.123	0.132	0.153

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

n		d=-0.5	0	0.25	0.5	0.75	1	1.5
250	Over	29	26	25	29	28	28	24
	Ext	791	818	806	800	820	828	821
	Under	180	156	169	171	152	144	155
	d-MAE	0.229	0.189	0.193	0.200	0.188	0.186	0.196
500	Over	27	33	29	30	38	29	30
	Ext	899	908	910	905	903	910	904
	Under	74	59	61	65	59	61	66
	d-MAE	0.178	0.138	0.142	0.146	0.138	0.146	0.170

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

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

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

$$\lim_{n\to\infty} \mathbb{P}(\tilde{\boldsymbol{\eta}}_{n,pq} \text{ exists}) = 1, \tag{5.1}$$

TABLE 7
Simulation results of RBIC in model (II) with GJR-GARCH(1,1) errors.

n		d=-0.5	0	0.25	0.5	0.75	1	1.5
250	Over	14	15	16	12	15	13	10
	Ext	809	853	837	840	853	857	856
	Under	177	132	147	148	132	130	134
	d-MAE	0.217	0.175	0.180	0.183	0.176	0.178	0.181
500	Over	11	15	11	13	14	16	9
	Ext	924	943	945	937	946	947	940
	Under	65	42	44	50	40	37	51
	d-MAE	0.172	0.129	0.130	0.139	0.130	0.134	0.158

which is ensured by Lemma 5.1 and the compactness of Π_{pq}. In the rest of the section, we abbreviate $\hat{\boldsymbol{\eta}}_{n,pq}$, $\tilde{\boldsymbol{\eta}}_{n,pq}$, $\boldsymbol{\eta}_{pq}$, $S^{+}_{0,pq}$, and $S_{\delta,pq}$ as $\hat{\boldsymbol{\eta}}_n$, $\tilde{\boldsymbol{\eta}}_n$, $\boldsymbol{\eta}$, S^{+}_{0}, and S_δ for notational simplicity.

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

$$\begin{aligned}\hat{\sigma}^2_{p_0q_0} - \hat{\sigma}^2_{pq} =& n^{-1}\sum_{t=1}^{n}\varepsilon_t^2(\hat{\boldsymbol{\eta}}_{n,p_0q_0}) - n^{-1}\sum_{t=1}^{n}\varepsilon_t^2(\boldsymbol{\eta}_0)\\ &- (n^{-1}\sum_{t=1}^{n}\varepsilon_t^2(\hat{\boldsymbol{\eta}}_n) - n^{-1}\sum_{t=1}^{n}\varepsilon_t^2(\tilde{\boldsymbol{\eta}}_n)).\end{aligned} \tag{5.2}$$

Let $\nabla\varepsilon_t(\boldsymbol{\eta}) = \partial\varepsilon_t(\boldsymbol{\eta})/\partial\boldsymbol{\eta} = (\nabla\varepsilon_t(\boldsymbol{\eta})_1, \dots, \nabla\varepsilon_t(\boldsymbol{\eta})_{\bar{r}})^\top$, with $\bar{r} = p+q+1$. By the mean value theorem,

$$\begin{aligned}&|\sum_{t=1}^{n}\varepsilon_t^2(\hat{\boldsymbol{\eta}}_n) - \sum_{t=1}^{n}\varepsilon_t^2(\tilde{\boldsymbol{\eta}}_n)|\\ &\le 2|\{\int_0^1\sum_{t=1}^{n}\varepsilon_t\nabla\varepsilon_t(\tilde{\boldsymbol{\eta}}_n + r(\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n))dr\}^\top(\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n)|\\ &+ 2|\{\int_0^1\sum_{t=1}^{n} r(\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n)^\top\nabla\varepsilon_t(\boldsymbol{\eta}^\star_{t,r})\nabla^\top\varepsilon_t(\tilde{\boldsymbol{\eta}}_n + r(\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n))dr\}(\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n)|,\end{aligned} \tag{5.3}$$

where $\boldsymbol{\eta}^\star_{t,r}$ satisfies $\|\boldsymbol{\eta}^\star_{t,r} - \tilde{\boldsymbol{\eta}}_n\| \le r\|\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n\|$. Given $M > 0$, define $\Lambda_n(M) = \{d(\hat{\boldsymbol{\eta}}_n, S_0^+) < v_n\}$, where $v_n = \min\{M^{1/4}n^{-1/2}, \delta_1\}$, for some $0 < \delta_1 < 1/2$. It follows from (5.3), Jensen's inequality, and the Cauchy–Schwarz inequality that

$$\begin{aligned}&\mathbb{P}(|\sum_{t=1}^{n}\varepsilon_t^2(\hat{\boldsymbol{\eta}}_n) - \sum_{t=1}^{n}\varepsilon_t^2(\tilde{\boldsymbol{\eta}}_n)| > M, \Omega_n\cap\Lambda_n(M))\\ &\le \mathbb{P}(\bar{r}^{1/2}\max_{1\le i\le\bar{r}}\sup_{\boldsymbol{\eta}\in S_{v_n}}|\sum_{t=1}^{n}\varepsilon_t\nabla\varepsilon_t(\boldsymbol{\eta})_i|\|\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n\| > 4^{-1}M, \Omega_n\cap\Lambda_n(M))\\ &\quad+\mathbb{P}(\bar{r}\max_{1\le i\le\bar{r}}\sup_{\boldsymbol{\eta}\in S_{v_n}}\{\sum_{t=1}^{n}(\nabla\varepsilon_t(\boldsymbol{\eta})_i)^2\}\|\hat{\boldsymbol{\eta}}_n - \tilde{\boldsymbol{\eta}}_n\|^2 > 4^{-1}M, \Omega_n\cap\Lambda_n(M))\\ &\le \mathbb{P}(\max_{1\le i\le\bar{r}}\sup_{\boldsymbol{\eta}\in S_{v_n}}|n^{-1/2}\sum_{t=1}^{n}\varepsilon_t\nabla\varepsilon_t(\boldsymbol{\eta})_i| > 4^{-1}\bar{r}^{-1/2}M^{3/4}, \Omega_n)\end{aligned} \tag{5.4}$$

$$+\mathbb{P}(\max_{1\le i\le\bar{r}}\sup_{\boldsymbol{\eta}\in S_{v_n}}\{n^{-1}\sum_{t=1}^{n}(\nabla\varepsilon_t(\boldsymbol{\eta})_i)^2\}>4^{-1}\bar{r}^{-1}M^{1/2},\Omega_n).$$

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

$$\mathbb{E}(\max_{1\le i\le\bar{r}}\sup_{\boldsymbol{\eta}\in S_{v_n}}|n^{-1/2}\sum_{t=1}^{n}\varepsilon_t\nabla\varepsilon_t(\boldsymbol{\eta})_i|^2)=O(1) \tag{5.5}$$

and

$$\mathbb{P}(\max_{1\le i\le\bar{r}}\sup_{\boldsymbol{\eta}\in S_{v_n}}\{n^{-1}\sum_{t=1}^{n}(\nabla\varepsilon_t(\boldsymbol{\eta})_i)^2\}>\bar{M})=o(1), \tag{5.6}$$

for some $\bar{M}>0$. Combining (5.4)–(5.6) yields that for any $\epsilon>0$, there exist $M_1,N_1>0$ such that for all $n>N_1$,

$$\mathbb{P}(|\sum_{t=1}^{n}\varepsilon_t^2(\hat{\boldsymbol{\eta}}_n)-\sum_{t=1}^{n}\varepsilon_t^2(\tilde{\boldsymbol{\eta}}_n)|>M_1,\Omega_n\cap\Lambda_n(M_1))<\epsilon/2. \tag{5.7}$$

Moreover, Lemma 5.1 ensures that for any $\epsilon>0$, there exist $M_2,N_2>0$ such that for all $n>N_2$,

$$\mathbb{P}(\Lambda_n^c(M_2))<\epsilon/2. \tag{5.8}$$

Thus, by (5.7) and (5.8), for any $\epsilon>0$, there exists $M=\max\{M_1,M_2\}$ such that for all $n>\max\{N_1,N_2\}$,

$$\mathbb{P}(|\sum_{t=1}^{n}\varepsilon_t^2(\hat{\boldsymbol{\eta}}_n)-\sum_{t=1}^{n}\varepsilon_t^2(\tilde{\boldsymbol{\eta}}_n)|>M,\Omega_n)<\epsilon. \tag{5.9}$$

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

$$\mathbb{P}(|\sum_{t=1}^{n}\varepsilon_t^2(\hat{\boldsymbol{\eta}}_{n,p_0q_0})-\sum_{t=1}^{n}\varepsilon_t^2(\boldsymbol{\eta}_0)|>M_3)<\epsilon. \tag{5.10}$$

Combining (5.1), (5.2), (5.9), and (5.10) 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$(p^\star,d_0,q^\star)$ model,

$$(1-\alpha_{0,1}B-\cdots-\alpha_{0,p^\star}B^{p^\star})(1-B)^{d_0}y_t=(1-\beta_{0,1}B-\cdots-\beta_{0,q^\star}B^{q^\star})\varepsilon_t, \tag{5.11}$$

where $p^\star=\max\{p,p_0\}$, $q^\star=\max\{q,q_0\}$, and $\alpha_{0,i}=\beta_{0,j}=0$ for $i>p_0$ and $j>q_0$. Define $\boldsymbol{\eta}_{0,p^\star q^\star}=(\alpha_{0,1},\dots,\alpha_{0,p_0},0,\dots,0,\beta_{0,1},\dots,\beta_{0,q_0},0,\dots,0,d_0)\in\mathbb{R}^{p^\star+q^\star+1}$, which is $\boldsymbol{\eta}_0$ extended by adding zeros to the overfitted entries. It follows from (1.4), (5.11), and the proofs of (2.8), (2.9), and (2.18) of Hualde and Robinson (2011) that for any $\delta>0$, there exists a small number $c_\delta^\star>0$ such that

$$\mathbb{P}(\inf_{\boldsymbol{\eta}_{p^\star q^\star}\in\Pi_{p^\star q^\star}\times D-B_\delta(\boldsymbol{\eta}_{0,p^\star q^\star})}n^{-1}\sum_{t=1}^{n}\varepsilon_t^2(\boldsymbol{\eta}_{p^\star q^\star})-n^{-1}\sum_{t=1}^{n}\varepsilon_t^2(\boldsymbol{\eta}_{0,p^\star q^\star})\le c_\delta^\star)=o(1). \tag{5.12}$$

Using (1.5) and (5.12) with a small enough δ, we obtain for some small constant $c>0$ that

(5.13)

$$\begin{aligned}&\mathbb{P}(n^{-1}\sum_{t=1}^{n}\varepsilon_t^2(\hat{\boldsymbol{\eta}}_n)-n^{-1}\sum_{t=1}^{n}\varepsilon_t^2(\boldsymbol{\eta}_0)\le c)\\ \le&\mathbb{P}(\inf_{\boldsymbol{\eta}_{p^\star q^\star}\in\Pi_{p^\star q^\star}\times D-B_\delta(\boldsymbol{\eta}_{0,p^\star q^\star})}n^{-1}\sum_{t=1}^{n}\varepsilon_t^2(\boldsymbol{\eta}_{p^\star q^\star})-n^{-1}\sum_{t=1}^{n}\varepsilon_t^2(\boldsymbol{\eta}_{0,p^\star q^\star})\le c)=o(1).\end{aligned}$$

As a result, (2.3) follows from (5.10), (5.13), and $\hat{\sigma}^2_{p_0q_0} - \hat{\sigma}^2_{pq} = n^{-1}\sum_{t=1}^n \varepsilon_t^2(\hat{\boldsymbol{\eta}}_{n,p_0q_0}) - n^{-1}\sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\eta}_0) - (n^{-1}\sum_{t=1}^n \varepsilon_t^2(\hat{\boldsymbol{\eta}}_n) - n^{-1}\sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\eta}_0))$. Equation (1.10) now is a straightforward consequence of (2.3) and (2.4). □

5.2. *Proof of* (2.9) *in the case of independent errors.*

LEMMA 5.1. *Under the assumptions of Theorem 2.1,* (2.9) *holds.*

PROOF OF LEMMA 5.1. For given $p_0 \le p \le P$ and $q_0 \le q \le Q$, denote Π_{pq}, Π^+_{pq}, $\hat{\boldsymbol{\theta}}_{n,pq}$, $\hat{d}_{n,pq}$, and $\boldsymbol{\theta}_{pq}$ by Π, Π^+, $\hat{\boldsymbol{\theta}}_n$, $\hat{d}_n$, and $\boldsymbol{\theta}$. First, we show that for any $\delta > 0$,

$$\mathbb{P}(\hat{\boldsymbol{\eta}}_n \in \Pi \times D - S_\delta) = o(1), \tag{5.14}$$

which is, in turn, ensured by

$$\lim_{n\to\infty} \mathbb{P}(\inf_{\boldsymbol{\eta}\in\Pi\times D - S_\delta} n^{-1}\sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\eta}) - n^{-1}\sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\eta}_0) \le c_\delta) = 0, \tag{5.15}$$

where c_δ is a small positive constant depending on δ. Note that for $\boldsymbol{\eta} \in \Pi \times D - S_\delta$,

$$(1-z)^{d-d_0} A_{1,\boldsymbol{\theta}}(z) A^{-1}_{2,\boldsymbol{\theta}}(z) A^{-1}_{1,\boldsymbol{\theta}_0}(z) A_{2,\boldsymbol{\theta}_0}(z) \ne 1. \tag{5.16}$$

Thus, (5.15) follows from the same arguments as those in the proofs of (2.8), (2.9), and (2.18) in Hualde and Robinson (2011) except that the open ball centered at the true parameter $\boldsymbol{\eta}_0$ is replaced by S_δ.

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 $\mathbf{A}$, depending only on $\alpha_{0,i}$ and $\beta_{0,j}$, such that

$$\mathbf{A}(\boldsymbol{\theta} - \boldsymbol{\theta}_{0,pq}) = \mathbf{A}\{\begin{pmatrix}\alpha_1\\ \vdots\\ \alpha_p\\ \beta_1\\ \vdots\\ \beta_q\end{pmatrix} - \begin{pmatrix}\alpha_{0,1}\\ \vdots\\ \alpha_{0,p}\\ \beta_{0,1}\\ \vdots\\ \beta_{0,q}\end{pmatrix}\} = \begin{pmatrix}\boldsymbol{\theta}_1^\star\\ \boldsymbol{\theta}_2^\star\end{pmatrix} \tag{5.17}$$

where $\boldsymbol{\theta}_1^\star$ is the $\max\{p+q_0, q+p_0\}$-dimensional vector formed by the coefficients of the polynomial $A_{1,\boldsymbol{\theta}}(z)A_{2,\boldsymbol{\theta}_0}(z) - A_{1,\boldsymbol{\theta}_0}(z)A_{2,\boldsymbol{\theta}}(z)$, and $\boldsymbol{\theta}_2^\star$ is a $\min\{p-p_0, q-q_0\}$-dimensional vector. Let $\boldsymbol{\theta}^\star = (\boldsymbol{\theta}_1^{\star\top}, \boldsymbol{\theta}_2^{\star\top})^\top$, $s_1^\star = \max\{p+q_0, q+p_0\}$, $s_2^\star = \min\{p-p_0, q-q_0\}$, $\boldsymbol{\eta}_1^\star = (\boldsymbol{\theta}_1^{\star\top}, d)^\top$, and F_0 be the one-to-one linear transformation depending on $\mathbf{A}$ and $\boldsymbol{\theta}_{0,pq}$ such that

$$\boldsymbol{\eta}^\star := F_0(\boldsymbol{\eta}) = (\boldsymbol{\theta}^{\star\top}, d)^\top. \tag{5.18}$$

Denote by $\hat{\boldsymbol{\theta}}^\star_{2n}$, $\hat{\boldsymbol{\eta}}^\star_{1n}$, and $\hat{\boldsymbol{\eta}}^\star_n$ the vectors corresponding to $\boldsymbol{\theta}_2^\star$, $\boldsymbol{\eta}_1^\star$, and $\boldsymbol{\eta}^\star$ when $\boldsymbol{\eta}$ is replaced by $\hat{\boldsymbol{\eta}}_n$. Equation (5.14) immediately leads to

$$\hat{\boldsymbol{\eta}}^\star_{1n} - \boldsymbol{\eta}^\star_{0,1} = o_p(1), \tag{5.19}$$

where $\boldsymbol{\eta}^\star_{0,1} = (\mathbf{0}^\top, d_0)^\top$.

Define $\Pi^\star = \{\boldsymbol{\theta}^\star \in \mathbb{R}^{p+q} | \boldsymbol{\theta} \in \Pi\}$ and $\Pi_1^\star = \{\boldsymbol{\theta}_1^\star \in \mathbb{R}^{s_1^\star} : \boldsymbol{\theta}^\star \in \Pi^\star\}$. It follows from (2.5) and (5.17) that

$$\boldsymbol{\eta}^\star_{0,1} \in \operatorname{int} \Pi_1^\star \times D. \tag{5.20}$$

Relation (5.20) ensures that there is a small constant $0 < \delta_1^\star < \min\{\tilde{\delta}_1, r_1^{\star -1}(M_2^\star)^{-2}\}$ such that $B_{\delta_1^\star}(\boldsymbol{\eta}_{0,1}^\star) \subset \text{int}\, \Pi_1^\star \times D$, where $r_1^\star = s_1^\star + 1$ and $\tilde{\delta}_1$ and $M_2^\star$ are positive constants to be specified later. Also, define

$$S_n(\boldsymbol{\eta}^\star) = \sum_{t=1}^n \tilde{\varepsilon}_t^2(\boldsymbol{\eta}^\star) \text{ and } \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star) = \varepsilon_t(F_0^{-1}(\boldsymbol{\eta}^\star)),$$

noting that $\varepsilon_t(F_0^{-1}(\boldsymbol{\eta}^\star)) = (1-B)^{d-d_0} A_{1,G_0^{-1}(\boldsymbol{\theta}^\star)}(B) A^{-1}_{2,G_0^{-1}(\boldsymbol{\theta}^\star)}(B) A^{-1}_{1,\boldsymbol{\theta}_0}(B) A_{2,\boldsymbol{\theta}_0}(B)\varepsilon_t$, with $G_0^{-1}(\boldsymbol{\theta}^\star) = \mathbf{A}^{-1}\boldsymbol{\theta}^\star + \boldsymbol{\theta}_{0,pq}$. For ease of exposition, we write $S_n(\boldsymbol{\eta}^\star) = S_n(\boldsymbol{\eta}_1^\star, \boldsymbol{\theta}_2^\star)$ and $\tilde{\varepsilon}_t(\boldsymbol{\eta}^\star) = \tilde{\varepsilon}_t(\boldsymbol{\eta}_1^\star, \boldsymbol{\theta}_2^\star)$. For $\delta > 0$, define $S_\delta^\star = \{\boldsymbol{\eta}^\star \in \Pi^\star \times D | \boldsymbol{\eta}_1^\star \in B_\delta(\boldsymbol{\eta}_{0,1}^\star)\}$. Then, by the mean value theorem for vector-valued functions, one obtains on set $A_n = \{\hat{\boldsymbol{\eta}}_n^\star \in S_{\delta_1^\star}^\star\}$,

$$\begin{aligned} \mathbf{0} &= \nabla_1 S_n(\hat{\boldsymbol{\eta}}_{1n}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star) = \nabla_1 S_n(\boldsymbol{\eta}_{0,1}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star) \\ &+ \{\int_0^1 \nabla_1^2 S_n(\boldsymbol{\eta}_{0,1}^\star + r(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star), \hat{\boldsymbol{\theta}}_{2n}^\star) dr\}(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star), \end{aligned} \tag{5.21}$$

where the integral of a matrix is to be understood component-wise. In view of (5.21), $\nabla_1 S_n(\boldsymbol{\eta}^\star) = 2\sum_{t=1}^n \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)\nabla_1\tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)$, and $\nabla_1^2 S_n(\boldsymbol{\eta}^\star) = 2\sum_{t=1}^n \nabla_1\tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)(\nabla_1\tilde{\varepsilon}_t(\boldsymbol{\eta}^\star))^\top + 2\sum_{t=1}^n \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)\nabla_1^2\tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)$, it holds that

$$\sum_{t=1}^n \varepsilon_t \nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{0,1}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star) = -\{L(\hat{\boldsymbol{\eta}}_{1n}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star) + Q(\hat{\boldsymbol{\eta}}_{1n}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star)\}(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star) \text{ on } A_n, \tag{5.22}$$

where

$$\begin{aligned} L(\hat{\boldsymbol{\eta}}_{1n}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star) &= \int_0^1 \sum_{t=1}^n \nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{0,1}^\star + r(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star), \hat{\boldsymbol{\theta}}_{2n}^\star) \\ &\qquad \times (\nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{0,1}^\star + r(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star), \hat{\boldsymbol{\theta}}_{2n}^\star))^\top \, dr, \\ Q(\hat{\boldsymbol{\eta}}_{1n}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star) &= \int_0^1 \sum_{t=1}^n \tilde{\varepsilon}_t(\boldsymbol{\eta}_{0,1}^\star + r(\hat{\boldsymbol{\eta}}_1^\star - \boldsymbol{\eta}_{0,1}^\star), \hat{\boldsymbol{\theta}}_2^\star)\nabla_1^2 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{0,1}^\star + r(\hat{\boldsymbol{\eta}}_1^\star - \boldsymbol{\eta}_{0,1}^\star), \hat{\boldsymbol{\theta}}_2^\star)\, dr. \end{aligned}$$

Define $S_{\delta,+}^\star = S_\delta^\star \cup \{\lambda\boldsymbol{\eta}^\star + (1-\lambda)(\mathbf{0}^\top, \boldsymbol{\theta}_2^{\star\top}, d_0)^\top | \boldsymbol{\eta}^\star \in S_\delta^\star,\ 0 \le \lambda \le 1\}$. Choose a small enough $\tilde{\delta}_1 \in (0, 1/2)$. Then, by the compactness of $\bar{S}_{\tilde{\delta}_1,+}^\star$ (the closure of $S_{\tilde{\delta}_1,+}^\star$), there exists a set of finite points $\{\boldsymbol{\eta}_1^\star, \ldots, \boldsymbol{\eta}_l^\star\} \subset \bar{S}_{\tilde{\delta}_1,+}^\star$ and a small positive number $0 < \tilde{\delta}_2 < 1/2 - \tilde{\delta}_1$, depending possibly on $\Pi^\star$ (and Π), such that

$$\bar{S}_{\tilde{\delta}_1,+}^\star \subset \bigcup_{k=1}^l B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star). \tag{5.23}$$

Moreover, for all $|z| \le 1$ and all $\boldsymbol{\theta}^\star$ for which $(\boldsymbol{\theta}^{\star\top}, d)^\top \in \bigcup_{k=1}^l B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)$,

$$A_{1,G_0^{-1}(\boldsymbol{\theta}^\star)}(z) \neq 0 \text{ and } A_{2,G_0^{-1}(\boldsymbol{\theta}^\star)}(z) \neq 0. \tag{5.24}$$

Direct algebraic manipulation gives

$$\lambda_{\min}(L(\hat{\boldsymbol{\eta}}_{1n}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star)) \ge \inf_{\boldsymbol{\eta}^\star \in \bigcup_{k=1}^l B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} \lambda_{\min}(\sum_{t=1}^n \nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)(\nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star))^\top) \text{ on } A_n, \tag{5.25}$$

where $\lambda_{\min}(\boldsymbol{M})$ denotes the minimum eigenvalue of matrix $\boldsymbol{M}$. It is shown in Appendix A that for some $M_1^\star > 0$,

$$\mathbb{P}\{\sup_{\boldsymbol{\eta}^\star \in \bigcup_{k=1}^{l} B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} \lambda_{\min}^{-1}(\hat{\boldsymbol{\Gamma}}_1(\boldsymbol{\eta}^\star)) > M_1^\star\} = o(1), \tag{5.26}$$

where $\hat{\boldsymbol{\Gamma}}_1(\boldsymbol{\eta}^\star) = n^{-1}\sum_{t=1}^{n} \nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)(\nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star))^\top$. In view of (5.25) and (5.26), we can assume without loss of generality that $L^{-1}(\hat{\boldsymbol{\eta}}_1^\star, \hat{\boldsymbol{\theta}}_2^\star)$ exists on A_n. Therefore, by (5.22),

(5.27)

$$\begin{aligned}
&\|n^{1/2}(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star)\| I_{A_n} \le \|nL^{-1}(\hat{\boldsymbol{\eta}}_{1n}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star)\| \|n^{-1/2}\sum_{t=1}^{n} \varepsilon_t \nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{0,1}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star)\| I_{A_n} \\
&+ \|nL^{-1}(\hat{\boldsymbol{\eta}}_{1n}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star)\| \| \int_0^1 n^{-1/2}\sum_{t=1}^{n} \varepsilon_t \nabla_1^2 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{0,1}^\star + r(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star), \hat{\boldsymbol{\theta}}_{2n}^\star) dr\| \\
&\times \|\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star\| I_{A_n} \\
&+ \|nL^{-1}(\hat{\boldsymbol{\eta}}_{1n}^\star, \hat{\boldsymbol{\theta}}_{2n}^\star)\| \| \int_0^1 n^{-1}\sum_{t=1}^{n} r(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star)^\top \nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{1,t,r}^\star) \\
&\qquad\qquad\qquad\qquad \times \nabla_1^2 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{0,1}^\star + r(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star), \hat{\boldsymbol{\theta}}_{2n}^\star) dr\| \\
&\times \|n^{1/2}(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star)\| I_{A_n},
\end{aligned}$$

where $\boldsymbol{\eta}_{1,t,r}^\star$ satisfies $\|\boldsymbol{\eta}_{1,t,r}^\star - \boldsymbol{\eta}_{0,1}^\star\| \le r\|\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star\|$.

It follows from the Cauchy–Schwarz inequality and Jensen's inequality that on the set A_n,

$$\begin{aligned}
&\| \int_0^1 n^{-1/2}\sum_{t=1}^{n} \varepsilon_t \nabla_1^2 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{0,1}^\star + r(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star), \hat{\boldsymbol{\theta}}_{2n}^\star)\, dr\| \\
&\le r_1^\star \max_{1 \le i,j \le r_1^\star} \sup_{\boldsymbol{\eta}^\star \in S_{\delta_1^\star}^\star} |n^{-1/2}\sum_{t=1}^{n} \varepsilon_t (\nabla_1^2 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)_{i,j})|
\end{aligned} \tag{5.28}$$

and

$$\begin{aligned}
&\| \int_0^1 n^{-1}\sum_{t=1}^{n} r(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star)^\top \nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{t,r}^\star) \nabla_1^2 \tilde{\varepsilon}_t(\boldsymbol{\eta}_{0,1}^\star + r(\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star), \hat{\boldsymbol{\theta}}_{2n}^\star)\, dr\| \\
&\le r_1^\star \|\hat{\boldsymbol{\eta}}_{1n}^\star - \boldsymbol{\eta}_{0,1}^\star\| (\sup_{\boldsymbol{\eta}^\star \in S_{\delta_1^\star}^\star} n^{-1}\sum_{t=1}^{n} \|\nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)\|^2)^{1/2} \\
&\quad \times \{\max_{1 \le i,j \le r_1^\star} \sup_{\boldsymbol{\eta}^\star \in S_{\delta_1^\star}^\star} n^{-1}\sum_{t=1}^{n} (\nabla^2 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)_{i,j})^2\}^{1/2}.
\end{aligned} \tag{5.29}$$

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 $M_2^\star \ge M_1^\star > 0$ such that

$$\mathbb{P}(\sup_{\boldsymbol{\eta}^\star \in \bigcup_{k=1}^{l} B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} n^{-1}\sum_{t=1}^{n} \|\nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)\|^2 > M_2^\star) = o(1), \tag{5.30}$$

$$\mathbb{P}(\max_{1 \le i,j \le r_1^\star} \sup_{\boldsymbol{\eta}^\star \in \bigcup_{k=1}^{l} B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} n^{-1}\sum_{t=1}^{n} (\nabla_1^2 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)_{i,j})^2 > M_2^\star) = o(1), \tag{5.31}$$

$$\mathbb{E}\Big(\sup_{\boldsymbol{\eta}^\star\in\bigcup_{k=1}^{l}B_{\tilde\delta_2}(\boldsymbol{\eta}_k^\star)}\|n^{-1/2}\sum_{t=1}^{n}\varepsilon_t\nabla_1\tilde\varepsilon_t(\boldsymbol{\eta}^\star)\|\Big)=O(1), \tag{5.32}$$

and

$$\mathbb{E}\Big(\max_{1\le i,j\le r_1^\star}\sup_{\boldsymbol{\eta}^\star\in\bigcup_{k=1}^{l}B_{\tilde\delta_2}(\boldsymbol{\eta}_k^\star)}|n^{-1/2}\sum_{t=1}^{n}\varepsilon_t(\nabla_1^2\tilde\varepsilon_t(\boldsymbol{\eta}^\star)_{i,j})|\Big)=O(1), \tag{5.33}$$

noting that (A.6)–(A.10) are given in Appendix A. Using (5.19) and (5.26)–(5.33), we obtain that for any $\epsilon>0$, there exist $M^\star, N^\star>0$ such that for all $n>N^\star$,

$$\mathbb{P}(\|n^{1/2}(\hat{\boldsymbol{\eta}}_{1n}^\star-\boldsymbol{\eta}_{0,1}^\star)\|>M^\star)<\epsilon. \tag{5.34}$$

Let $\bar{\boldsymbol{\theta}}_n=G_0^{-1}((\mathbf{0}^\top,\hat{\boldsymbol{\theta}}_{2n}^{\star\top})^\top)$. Since when $\tilde{\boldsymbol{\eta}}_n$ (defined in (5.1)) exists,

$$\begin{aligned} d(\hat{\boldsymbol{\eta}}_n,S_0^+)&\le(\|\hat{\boldsymbol{\theta}}_n-\bar{\boldsymbol{\theta}}_n\|^2+|\hat d_n-d_0|^2)^{1/2}\\ &\le(\|\mathbf{A}^{-1}\|^2\|(\hat{\boldsymbol{\theta}}_{1n}^{\star\top},\hat{\boldsymbol{\theta}}_{2n}^{\star\top})^\top-(\mathbf{0}^\top,\hat{\boldsymbol{\theta}}_{2n}^{\star\top})^\top\|^2+|\hat d_n-d_0|^2)^{1/2}\\ &\le\max\{\|\mathbf{A}^{-1}\|,1\}\|\hat{\boldsymbol{\eta}}_{1n}^\star-\boldsymbol{\eta}_{0,1}^\star\|, \end{aligned} \tag{5.35}$$

the desired conclusion (2.9) is ensured by (5.1), (5.34), and (5.35). □

5.3. *Proof of Theorem 3.1.* The proof of Theorem 3.1 relies heavily on Lemmas 5.2 and 5.3, whose proofs are given in the supplementary material and Appendix A, respectively. Lemma 5.2 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 5.2, for any $1\le m\le k$, define $\mathbf{J}(m,k)=\{(j_1,\dots,j_m)|j_1<\cdots<j_m, j_i\in\{1,\dots,k\},1\le i\le m\}$. Moreover, for $\mathbf{j}=(j_1,\dots,j_m)\in\mathbf{J}(m,k)$ and a smooth function $w=w(\xi_1,\dots,\xi_k)$, let $\mathbf{D}_{\mathbf{j}}w=\partial^m w/\partial\xi_{j_1},\dots,\partial\xi_{j_m}$.

LEMMA 5.2. *Assume (3.1) and (3.2). Let $\boldsymbol{\theta}_a=(\theta_{a,1},\dots,\theta_{a,k})^\top$ be some point in $\mathbb{R}^k$, $k\ge1$, and $\bar\delta$ be a positive number. For $t\ge2$, define $K_t(\boldsymbol{\theta})=\sum_{i=1}^{t-1}c_i(\boldsymbol{\theta})\varepsilon_{t-i}$ and $Q_t(\boldsymbol{\theta})=\sum_{i=1}^{t-1}d_i(\boldsymbol{\theta})\varepsilon_{t-i}$, where $c_i(\boldsymbol{\theta})$ and $d_i(\boldsymbol{\theta})$ are real-valued functions on $B_{\bar\delta}(\boldsymbol{\theta}_a)$. Assume for any $i\ge1$, $\mathbf{j}=(j_1,\dots,j_m)^\top\in\mathbf{J}(m,k)$, and $1\le m\le k$, $\mathbf{D}_{\mathbf{j}}c_i(\boldsymbol{\theta})$ are continuous on $B_{\bar\delta}(\boldsymbol{\theta}_a)$ and*

$$\sup_{-\infty<i<\infty}\mathbb{E}\|\boldsymbol{w}_i\|^{m_1}<\infty, \tag{5.36}$$

for some $m_1\ge2$. Then, there exists $C>0$ such that for all $n\ge2$,

$$\begin{aligned} &\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in B_{\bar\delta}(\boldsymbol{\theta}_a)}|n^{-1/2}\sum_{t=2}^{n}K_t(\boldsymbol{\theta})\varepsilon_t|^{m_1}\Big)\\ \le&C\Big[\{n^{-1}\sum_{t=2}^{n}\sum_{i=1}^{t-1}\max_{\mathbf{j}\in\mathbf{J}(m,k),1\le m\le k}\sup_{\boldsymbol{\theta}\in B_{\bar\delta}(\boldsymbol{\theta}_a)}(\mathbf{D}_{\mathbf{j}}c_i(\boldsymbol{\theta}))^2\}^{m_1/2}\\ &+\{n^{-1}\sum_{t=2}^{n}\sum_{i=1}^{t-1}c_i^2(\boldsymbol{\theta}_a)\}^{m_1/2}\Big]. \end{aligned} \tag{5.37}$$

Furthermore, if for any $i,j\ge1$, $\mathbf{j}=(j_1,\dots,j_m)^\top\in\mathbf{J}(m,k)$, and $1\le m\le k$, $\mathbf{D}_{\mathbf{j}}\{c_i(\boldsymbol{\theta})d_j(\boldsymbol{\theta})\}$ are continuous on $B_{\bar\delta}(\boldsymbol{\theta}_a)$, then there exists $C>0$ such that for all $n\ge3$,

$$\mathbb{E}\Big(\sup_{\boldsymbol{\theta}\in B_{\bar\delta}(\boldsymbol{\theta}_a)}|n^{-1/2}\sum_{t=2}^{n}K_t(\boldsymbol{\theta})Q_t(\boldsymbol{\theta})-\mathbb{E}(K_t(\boldsymbol{\theta})Q_t(\boldsymbol{\theta}))|^{m_1}\Big)$$

$$
\begin{aligned}
&\le C[\{n^{-1}\sum_{i=1}^{n-1}\{\sum_{t=i+1}^{n}|c_{t-i}(\boldsymbol{\theta}_a)d_{t-i}(\boldsymbol{\theta}_a)|\}^2\}^{m_1/2} \\
&+\{n^{-1}\sum_{l=2}^{n-1}\sum_{i=1}^{l-1}\{\sum_{t=l+1}^{n}|c_{t-i}(\boldsymbol{\theta}_a)d_{t-l}(\boldsymbol{\theta}_a)+c_{t-l}(\boldsymbol{\theta}_a)d_{t-i}(\boldsymbol{\theta}_a)|\}^2\}^{m_1/2} \\
&+\{n^{-1}\sum_{i=1}^{n-1}[\sum_{t=i+1}^{n}\max_{\mathbf{j}\in\mathbf{J}(m,k),1\le m\le k}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}|\mathbf{D}_{\mathbf{j}}\{c_{t-i}(\boldsymbol{\theta})d_{t-i}(\boldsymbol{\theta})\}]^2\}^{m_1/2} \\
&+\{n^{-1}\sum_{l=2}^{n-1}\sum_{i=1}^{l-1}[\sum_{t=l+1}^{n}\max_{\mathbf{j}\in\mathbf{J}(m,k),1\le m\le k}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}|\mathbf{D}_{\mathbf{j}}\{c_{t-i}(\boldsymbol{\theta})d_{t-l}(\boldsymbol{\theta}) \\
&+c_{t-l}(\boldsymbol{\theta})d_{t-i}(\boldsymbol{\theta})\}|]^2\}^{m_1/2}].
\end{aligned}
\tag{5.38}
$$

Lemma 5.3 plays the same role as Lemma 5.1 in the proof of Theorem 2.1.

LEMMA 5.3. *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 (5.5), (5.6), and (5.8), whereas (2.3) is ensured by (5.12). By making use of Lemma 5.2, we prove (5.5) and (5.6) in Section S1 of the supplementary material. Moreover, (5.8) and (5.12) are immediate consequences of Lemma 5.3 and (5.15), respectively. Note that the proof of (5.15) under the assumptions of Theorem 3.1 is given in Appendix A. Consequently, the desired conclusion follows. □

6. 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 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.

APPENDIX A: PROOFS OF (5.26) AND LEMMA 5.3

PROOF OF (5.26). We first show that for all large n and $1 \le k \le l$, there exists $c_k > 0$ such that

$$
\inf_{\boldsymbol{\eta}^\star\in B_{\bar{\delta}_2}(\boldsymbol{\eta}_k^\star)} \lambda_{\min}(\boldsymbol{\Gamma}_1(\boldsymbol{\eta}^\star)) > c_k, \tag{A.1}
$$

where $\mathbf{\Gamma}_1(\boldsymbol{\eta}^\star) = \mathbb{E}(\hat{\mathbf{\Gamma}}_1(\boldsymbol{\eta}^\star))$. To prove (A.1), consider

$$\begin{aligned} C_{\boldsymbol{\eta}^\star}(z) :=& (1-z)^{d-d_0} A_{1,G_0^{-1}(\boldsymbol{\theta}^\star)}(z) A^{-1}_{2,G_0^{-1}(\boldsymbol{\theta}^\star)}(z) A^{-1}_{1,\boldsymbol{\theta}_0}(z) A_{2,\boldsymbol{\theta}_0}(z) \\ =& (1-z)^{d-d_0}(z)[\{A_{1,G_0^{-1}(\boldsymbol{\theta}^\star)}(z) A_{2,\boldsymbol{\theta}_0}(z) - A_{2,G_0^{-1}(\boldsymbol{\theta}^\star)}(z) A_{1,\boldsymbol{\theta}_0}(z)\} \\ & \times \{A^{-1}_{2,G_0^{-1}(\boldsymbol{\theta}^\star)}(z) A^{-1}_{2,\boldsymbol{\theta}_0}(z)\} + A_{1,\boldsymbol{\theta}_0}(z) A^{-1}_{2,\boldsymbol{\theta}_0}(z)] A^{-1}_{1,\boldsymbol{\theta}_0}(z) A_{2,\boldsymbol{\theta}_0}(z). \end{aligned} \tag{A.2}$$

According to the definition of $\boldsymbol{\theta}_1^\star$, we have for $i = 1, \ldots, s_1^\star$,

$$\frac{\partial}{\partial \eta^\star_{1,i}} C_{\boldsymbol{\eta}^\star}(z)\Big|_{\boldsymbol{\eta}_1^\star = \boldsymbol{\eta}^\star_{0,1}, \boldsymbol{\theta}_2^\star = \boldsymbol{\theta}_2^\star} = A^{-1}_{1,\boldsymbol{\theta}_0}(z) A^{-1}_{2,G_0^{-1}(\mathbf{0}^\top, \boldsymbol{\theta}_2^{\star\top})}(z) z^i \tag{A.3}$$

and

$$\frac{\partial}{\partial \eta^\star_{1,r_1^\star}} C_{\boldsymbol{\eta}^\star}(z)\Big|_{\boldsymbol{\eta}_1^\star = \boldsymbol{\eta}^\star_{0,1}, \boldsymbol{\theta}_2^\star = \boldsymbol{\theta}_2^\star} = \log(1-z), \tag{A.4}$$

where $(\eta^\star_{1,1}, \ldots, \eta^\star_{1,r_1^\star})^\top = \boldsymbol{\eta}_1^\star$. 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

$$\tilde{\varepsilon}_t(\boldsymbol{\eta}^\star) = \sum_{s=0}^{t-1} b_s(\boldsymbol{\eta}^\star) \varepsilon_{t-s}, \tag{A.5}$$

where $b_0(\boldsymbol{\eta}^\star) = 1$. Then $\nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)_i = \sum_{s=1}^{t-1} b_{s,i}(\boldsymbol{\eta}^\star)\varepsilon_{t-s}$, where $b_{s,i}(\boldsymbol{\eta}^\star) = \partial b_s(\boldsymbol{\eta}^\star)/\partial \eta_i^\star$. Recall the definition of $\mathbf{D_j}$ given in Section 5.3. It is clear that $b_{s,i}(\boldsymbol{\eta}^\star)$ has continuous partial derivatives $\mathbf{D_j} b_{s,i}(\boldsymbol{\eta}^\star)$. Combining (A.1) with

$$\inf_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} \lambda_{\min}(\hat{\mathbf{\Gamma}}_1(\boldsymbol{\eta}^\star)) \geq \inf_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} \lambda_{\min}(\mathbf{\Gamma}_1(\boldsymbol{\eta}^\star)) - \sup_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} \|\mathbf{\Gamma}_1(\boldsymbol{\eta}^\star) - \hat{\mathbf{\Gamma}}_1(\boldsymbol{\eta}^\star)\|,$$

(2.13), Lemma B.1 of Chan and Ing (2011), and Markov's inequality, one obtains for $M^\star_{1,k} > 2/c_k$,

$$\begin{aligned} &\mathbb{P}\{\sup_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} \lambda^{-1}_{\min}(\hat{\mathbf{\Gamma}}_1(\boldsymbol{\eta}^\star)) > M^\star_{1,k}\} \leq \mathbb{P}(\sup_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} \|\mathbf{\Gamma}_1(\boldsymbol{\eta}^\star) - \hat{\mathbf{\Gamma}}_1(\boldsymbol{\eta}^\star)\| > c_k/2) \\ \leq& (4 r_1^{\star 4}/c_k^2) \max_{1 \leq i,j \leq r_1^\star} \mathbb{E}[\sup_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} |n^{-1} \sum_{t=1}^{n} (\nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)_i)(\nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)_j) \\ & - \mathbb{E}\{(\nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)_i)(\nabla_1 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)_j)\}|^2] \\ \leq& C n^{-2} \max_{1 \leq i,j \leq r_1^\star} [\sum_{u=1}^{n-1} (\sum_{w=1}^{n-u} S^{(k)}_{w,w}(i,j))^2 + \sum_{u=1}^{n-1} (\sum_{u=1}^{n-u} V^{(k)}_{w,w}(i,j))^2 \\ & + \sum_{u=2}^{n-1} \{\sum_{v=1}^{u-1} (\sum_{w=1}^{n-u} S^{(k)}_{w+u-v,w}(i,j))^2 + \sum_{v=1}^{u-1} (\sum_{w=1}^{n-u} S^{(k)}_{w,w+u-v}(i,j))^2 \\ & + \sum_{v=1}^{u-1} (\sum_{w=1}^{n-u} V^{(k)}_{w+u-v,w}(i,j))^2 + \sum_{v=1}^{u-1} (\sum_{w=1}^{n-u} V^{(k)}_{w,w+u-v}(i,j))^2\}], \end{aligned} \tag{A.6}$$

where

$$S^{(k)}_{u,v}(i,j) = \max_{\mathbf{j} \in \mathbf{J}(m, r_1^\star), 1 \leq m \leq r_1^\star} \sup_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} |\mathbf{D_j}\{b_{u,i}(\boldsymbol{\eta}^\star) b_{v,j}(\boldsymbol{\eta}^\star)\}|,$$

$$V^{(k)}_{u,v}(i,j) = |b_{u,i}(\boldsymbol{\eta}_k^\star) b_{v,j}(\boldsymbol{\eta}_k^\star)|.$$

By (5.23), (5.24), the boundedness of $\|\mathbf{A}^{-1}\|$, and arguments similar to those in the proofs of Theorem 4.1 of Ling (2007) and Lemma 4 of Hualde and Robinson (2011), we obtain for any $s \geq 1$ and $1 \leq k \leq l$,

$$\max_{1\leq i\leq r_1^\star} \sup_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} |b_{s,i}(\boldsymbol{\eta}^\star)| \leq C(\log(s+1))s^{-1+\tilde{\delta}_1+\tilde{\delta}_2}, \tag{A.7}$$

and

$$\max_{1\leq i\leq r_1^\star} \max_{\mathbf{j}\in \mathbf{J}(m,r_1^\star),1\leq m\leq r_1^\star} \sup_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} |\mathbf{D}_{\mathbf{j}} b_{s,i}(\boldsymbol{\eta}^\star)| \leq C(\log(s+1))^2 s^{-1+\tilde{\delta}_1+\tilde{\delta}_2}. \tag{A.8}$$

In view of (A.7) and (A.8), it follows that for all $1 \leq i,j \leq r_1^\star$, $(\sum_{w=1}^{\infty} S_{w,w}^{(k)}(i,j))^2$ and $(\sum_{w=1}^{\infty} V_{w,w}^{(k)}(i,j))^2$ are bounded by some constant C, and $(\sum_{w=1}^{\infty} S_{w+u-v,w}^{(k)}(i,j))^2$, $(\sum_{w=1}^{\infty} S_{w,w+u-v}^{(k)}(i,j))^2$, $(\sum_{w=1}^{\infty} V_{w+u-v,w}^{(k)}(i,j))^2$, and $(\sum_{w=1}^{\infty} V_{w,w+u-v}^{(k)}(i,j))^2$ are bounded by $C(u-v)^{-1+2(\tilde{\delta}_1+\tilde{\delta}_2)}$. These bounds together with (A.6) and $\tilde{\delta}_1 + \tilde{\delta}_2 < 1/2$ yield

$$\mathbb{P}\{\sup_{\boldsymbol{\eta}^\star \in B_{\delta_2}(\boldsymbol{\eta}_k^\star)} \lambda_{\min}^{-1}(\hat{\boldsymbol{\Gamma}}_1(\boldsymbol{\eta}^\star)) > M_{1,k}^\star\} \leq Cn^{-1+2(\tilde{\delta}_1+\tilde{\delta}_2)} = o(1).$$

Thus (5.26) holds with $M_1^\star = \max_{1\leq k\leq l} M_{1,k}^\star$. □

REMARK A.1. Equation (A.5) implies that $\nabla_1^2 \tilde{\varepsilon}_t(\boldsymbol{\eta}^\star)_{i,j} = \sum_{s=1}^{t-1} c_{s,ij}(\boldsymbol{\eta}^\star)\varepsilon_{t-s}$, where $c_{s,ij}(\boldsymbol{\eta}^\star) = \partial^2 b_s(\boldsymbol{\eta}^\star)/\partial\eta_i^\star \partial\eta_j^\star$. By an argument similar to that used to prove (A.7) and (A.8), we have

$$\max_{1\leq i,j\leq r_1^\star} \sup_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} |c_{s,ij}(\boldsymbol{\eta}^\star)| \leq C(\log(s+1))^2 s^{-1+\tilde{\delta}_1+\tilde{\delta}_2}, \tag{A.9}$$

and

$$\max_{1\leq i,j\leq r_1^\star} \max_{\mathbf{j}\in \mathbf{J}(m,r_1^\star),1\leq m\leq r_1^\star} \sup_{\boldsymbol{\eta}^\star \in B_{\tilde{\delta}_2}(\boldsymbol{\eta}_k^\star)} |\mathbf{D}_{\mathbf{j}} c_{s,ij}(\boldsymbol{\eta}^\star)| \leq C(\log(s+1))^3 s^{-1+\tilde{\delta}_1+\tilde{\delta}_2}. \tag{A.10}$$

PROOF OF LEMMA 5.3. Let (p,q) satisfy $p_0 \leq p \leq P$ and $q_0 \leq q \leq Q$. If (5.15), (5.26), and (5.30)–(5.33) hold under the assumptions of Theorem 3.1, then the desired conclusion follows from the same argument as that in the proof of Lemma 5.1. Using Lemma 5.2 to replace Lemma B.1 of Chan and Ing (2011) in the proofs of (5.26) and (5.30)–(5.33), one can easily establish these equations under the assumptions of Theorem 3.1. The details are omitted. Therefore, it remains to prove (5.15).

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

$$\mathbb{E}|\sum_{t=1}^{n} z_t^2 - \mathbb{E}(z_t^2)| = O(n^{1/2}), \tag{A.11}$$

$$\begin{aligned}&\mathbb{E}|\sum_{r=j+1}^{k} \sum_{l=r-j+1}^{n-j} \{z_l z_{l-r+j} - \mathbb{E}(z_l z_{l-r+j})\}| \\ &\leq C(k-j)^{1/2}(n-j)^{1/2}, \text{ for all } 0 \leq j \leq n-2, j \leq k \leq n,\end{aligned} \tag{A.12}$$

$$\mathbb{E}|\sum_{t=1}^{n} \varepsilon_t^2 - \mathbb{E}(\varepsilon_t^2)| = O(n^{1/2}), \tag{A.13}$$

(A.14)
$$\mathbb{E}|\sum_{r=j+1}^{k}\sum_{l=r-j+1}^{n-j}\{\varepsilon_l\varepsilon_{l-r+j}-\mathbb{E}(\varepsilon_l\varepsilon_{l-r+j})\}| \le C(k-j)^{1/2}(n-j)^{1/2},\ \text{for all } 0\le j\le n-2, j\le k\le n,$$

(A.15)
$$\mathbb{E}|\sum_{k=1}^{j} z_{n-k}|^2 \le Cj,\ \text{for all } 0\le j\le n-1,$$

(A.16)
$$\mathbb{E}|\sum_{k=0}^{j} \varepsilon_{s-k}|^2 \le Cj,\ \text{for all } 0\le j\le s\le n,$$

(A.17)
$$\mathbb{E}|\sum_{j=1}^{n} z_j^2| = O(n),$$

and for some small $\xi, \epsilon_1 > 0$,

(A.18)
$$\lim_{n\to\infty}\mathbb{P}(\inf_{\boldsymbol{\eta}_{pq}\in\Pi_{pq}\times[L,d_0-1/2-\xi]}(\frac{1}{n^{d_0-d+1/2}}\sum_{t=1}^{n}\varepsilon_t(\boldsymbol{\eta}_{pq}))^2 > \epsilon_1) = 1,$$

noting that L is the prescribed lower bound of d. Then, (5.15) follows from these equations, (5.16), 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.

In view of (2.46) of Hualde and Robinson (2011), (A.18) holds if A(i), A(ii), and A(iii) of Hosoya (2005) are fulfilled by error terms obeying (3.1), (3.2), and (3.6). Note first that A(i) clearly holds. In addition, for $l, m > t \ge 1$, with $l \ne m$,

(A.19)
$$\mathbb{E}(\varepsilon_l\varepsilon_m|\mathcal{F}_t) = 0 \text{ a.s. and } \mathbb{E}(\varepsilon_l\varepsilon_m) = 0.$$

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

(A.20)
$$\mathbb{E}|\mathbb{E}({\varepsilon_l}^2|\mathcal{F}_t)-\mathbb{E}({\varepsilon_l}^2)|^2 = \mathbb{E}|\mathbb{E}(\sum_{s=0}^{\infty}\mathbf{a}_s^\top\boldsymbol{w}_{l-s}|\mathcal{F}_t)|^2 = \\ =\mathbb{E}|\sum_{s=l-t}^{\infty}\mathbf{a}_s^\top\boldsymbol{w}_{l-s}|^2 \le C\sum_{s=l-t}^{\infty}\|\mathbf{a}_s\|^2 = O((l-t)^{-2\iota+1}).$$

Thus, A(ii) is ensured by (A.19) and (A.20).

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

(A.21)
$$\mathrm{Var}(\sum_{t=1}^{n}a_t\varepsilon_t) = \mathbb{E}(\sum_{t=1}^{n}a_t\varepsilon_t)^2 = \sigma_\varepsilon^2\sum_{t=1}^{n}a_t^2.$$

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

(A.22)
$$\mathbb{E}|\sum_{t=1}^{n}a_t\varepsilon_t|^4 \le C(\sum_{t=1}^{n}a_t^2)^2.$$

Simple algebraic manipulations give

$$\begin{aligned}\mathbb{E}|\sum_{t=1}^{n} a_t\varepsilon_t|^4 &= \mathbb{E}|(\sum_{t=1}^{n} a_t\varepsilon_t)^2 - \mathbb{E}(\sum_{t=1}^{n} a_t\varepsilon_t)^2 + \mathbb{E}(\sum_{t=1}^{n} a_t\varepsilon_t)^2|^2 \\ &\le C[\mathbb{E}|\sum_{t=1}^{n} a_t^2(\varepsilon_t^2 - \mathbb{E}(\varepsilon_t^2))|^2 + \mathbb{E}|\sum_{t_1=2}^{n}\sum_{t_2=1}^{t_1-1} a_{t_1}a_{t_2}\varepsilon_{t_1}\varepsilon_{t_2}|^2 + (\sum_{t=1}^{n} a_t^2)^2].\end{aligned} \tag{A.23}$$

By (3.1), (3.2), (3.6), Minkowski's inequality, Burkholder's inequality, and the Cauchy–Schwarz inequality, we have

$$\mathbb{E}|\sum_{t=1}^{n} a_t^2(\varepsilon_t^2 - \mathbb{E}(\varepsilon_t^2))|^2 \le [\sum_{t=1}^{n}\{\mathbb{E}|a_t^2(\varepsilon_t^2 - \mathbb{E}(\varepsilon_t^2))|^2\}^{1/2}]^2 \le C(\sum_{t=1}^{n} a_t^2)^2, \tag{A.24}$$

and

$$\mathbb{E}|\sum_{t_1=2}^{n}\sum_{t_2=1}^{t_1-1} a_{t_1}a_{t_2}\varepsilon_{t_1}\varepsilon_{t_2}|^2 \le C(\sum_{t=1}^{n} a_t^2)^2. \tag{A.25}$$

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

Acknowledgements. We would like to thank the Editor, an Associate Editor and an anonymous referee for their critical comments and thoughtful suggestions, which led to an improved version of this paper. Further, we would like to thank Professor Ruey S. Tsay for his insightful comments and encouragement on an earlier draft of this paper. Huang and Ing's research was supported by grant 109-2118-M-007-007-MY3 from the Ministry of Science and Technology, Taiwan. Chan's research was supported, in part, by the General Research Fund of HKSAR-RGC-GRF Nos. 14308218 and 14307921, HKSAR-RGCCRF:CityU8/CRG/12G, and the Theme-based Research Scheme of HKSAR-RGC-TBS T32-101/15-R. Chen's research was supported by the National Natural Science Foundation of China under Contract No. 12001444, and the MOE (Ministry of Education in China) Project of Humanities and Social Sciences (20YJC910001).

SUPPLEMENTARY MATERIAL

Supplement to "Consistent Order Selection for ARFIMA Processes" (DOI:). The supplementary material contains the proofs of (5.5), (5.6), Lemma 5.2, and (A.12).
().

REFERENCES

BAILLIE, R. T., KAPETANIOS, G. and PAPAILIAS, F. (2014). Modified information criteria and selection of long memory time series models. *Computational Statistics and Data Analysis* **76** 116–131.

BARDET, J.-M., KAMILA, K. and KENGNE, W. (2020). Consistent model selection criteria and goodness-of-fit test for common time series models. *Electronic Journal of Statistics* **14** 2009–2052.

BERAN, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.

BERAN, J., BHANSALI, R. J. and OCKER, D. (1998). On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes. *Biometrika* **85** 921–934.

BOLLERSLEV, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31** 307–327.

CHAN, N. H., HUANG, S. F. and ING, C. K. (2013). Moment bounds and mean squared prediction errors of long-memory time series. *The Annals of Statistics* **41** 1268–1298.

CHAN, N. H. and ING, C. K. (2011). Uniform moment bounds of Fisher's information with applications to time series. *The Annals of Statistics* **39** 1526–1550.

GLOSTEN, L. R., JAGANNATHAN, R. and RUNKLE, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* **48** 1779–1801.

GUO, L., CHEN, H.-F. and ZHANG, J.-F. (1989). Consistent order estimation for linear stochastic feedback control systems (CARMA model). *Automatica* **25** 147–151.

HANNAN, E. J. (1980). The estimation of the order of an ARMA process. *The Annals of Statistics* **8** 1071–1081.

HANNAN, E. J. and KAVALIERIS, L. (1984). A method for autoregressive-moving average estimation. *Biometrika* **71** 273–280.

HANNAN, E. J. and RISSANEN, J. (1982). Recursive estimation of mixed autoregressive-moving average order. *Biometrika* **69** 81–94.

HOSOYA, Y. (2005). Fractional invariance principle. *Journal of Time Series Analysis* **26** 463–486.

HUALDE, J. and ROBINSON, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *The Annals of Statistics* **39** 3152–3181.

HUANG, D. and GUO, L. (1990). Estimation of nonstationary ARMAX models based on the Hannan-Rissanen method. *The Annals of Statistics* **18** 1729–1756.

ING, C. K., SIN, C.-Y. and YU, S. H. (2012). Model selection for integrated autoregressive processes of infinite order. *Journal of Multivariate Analysis* **106** 57–71.

ING, C. K. and WEI, C. Z. (2005). Order selection for same-realization predictions in autoregressive processes. *The Annals of Statistics* **33** 2423–2474.

LING, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *The Annals of Statistics* **35** 1213–1237.

LING, S. and MCALEER, M. (2002a). Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models. *Economic Theory* **18** 722–729.

LING, S. and MCALEER, M. (2002b). Stationarity and the existence of moments of a family of GARCH processes. *Journal of Econometrics* **106** 109–117.

PAULSEN, J. (1984). Order determination of multivariate autoregressive time series with unit roots. *Journal of Time Series Analysis* **5** 115–127.

SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *The Annals of Statistics* **8** 147–164.

TSAY, R. S. (1984). Order selection in nonstationary autoregressive models. *The Annals of Statistics* **12** 1425–1433.

WEI, C.-Z. (1992). On predictive least squares principles. *The Annals of Statistics* **20** 1–42.

SUPPLEMENT TO "CONSISTENT ORDER SELECTION FOR ARFIMA PROCESSES"

BY HSUEH-HAN HUANG[1], NGAI HANG CHAN[2], KUN CHEN[3], AND CHING-KANG ING[1]

[1]*National Tsing Hua University*

[2]*Chinese University of Hong Kong*

[3]*Southwestern University of Finance and Economics*

This supplement contains the proofs of (5.5), (5.6), Lemma 5.2, and (A.12).

S1. Proofs of (5.5) and (5.6) for independent and conditional heteroscedastic errors.

PROOF OF (5.5). *The case of independent errors:* Let $\bar{S}_{\delta_1}$ be the closure of S_{δ_1}. By the compactness of $\bar{S}_{\delta_1}$, there exists a set of finite l points $\{\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_l\} \subset \bar{S}_{\delta_1}$ and a small positive number $0 < \bar{\delta}_1 < 1/2 - \delta_1$, depending possibly on Π, such that

$$\bar{S}_{\delta_1} \subset \bigcup_{k=1}^{l} B_{\bar{\delta}_1}(\boldsymbol{\eta}_k), \tag{S1.1}$$

and for each $\boldsymbol{\eta} \in B_{\bar{\delta}_1}(\boldsymbol{\eta}_k)$ and $1 \le k \le l$,

$$A_{1,\boldsymbol{\theta}}(z) \neq 0, A_{2,\boldsymbol{\theta}}(z) \neq 0, \ |z| \le 1. \tag{S1.2}$$

Write

$$\varepsilon_t(\boldsymbol{\eta}) = \sum_{s=0}^{t-1} \bar{b}_s(\boldsymbol{\eta})\varepsilon_{t-s}, \tag{S1.3}$$

and let $\bar{b}_{s,i}(\boldsymbol{\eta}) = \partial \bar{b}_s(\boldsymbol{\eta})/\partial \eta_i$ and $\nabla \varepsilon_t(\boldsymbol{\eta})_i = \sum_{s=1}^{t-1} \bar{b}_{s,i}(\boldsymbol{\eta})\varepsilon_{t-s}$. Recall the definition of $\mathbf{D_j}$ given in Section 5.3. It is clear that $\bar{b}_{s,i}(\boldsymbol{\eta})$ has continuous partial derivatives, $\mathbf{D_j}\bar{b}_{s,i}(\boldsymbol{\eta})$, on each $B_{\bar{\delta}_1}(\boldsymbol{\eta}_k)$. By arguments similar to those in the proofs of Theorem 4.1 of Ling (2007) and Lemma 4 of Hualde and Robinson (2011), we have for any $s \ge 1$ and $1 \le k \le l$,

$$\max_{1 \le i \le \bar{r}} \sup_{\boldsymbol{\eta} \in B_{\bar{\delta}_1}(\boldsymbol{\eta}_k)} |\bar{b}_{s,i}(\boldsymbol{\eta})| \le C(\log(s+1))s^{-1+\delta_1+\bar{\delta}_1}, \tag{S1.4}$$

$$\max_{1 \le i \le \bar{r}} \max_{\mathbf{j} \in \mathbf{J}(m,\bar{r}), 1 \le m \le \bar{r}} \sup_{\boldsymbol{\eta} \in B_{\bar{\delta}_1}(\boldsymbol{\eta}_k)} |\mathbf{D_j}\bar{b}_{s,i}(\boldsymbol{\eta})| \le C(\log(s+1))^2 s^{-1+\delta_1+\bar{\delta}_1}, \tag{S1.5}$$

where C, here and hereafter, represents a generic positive constants independent of n. Then, it follows from (S1.1)–(S1.5), (2.13), and Lemma B.1 of Chan and Ing (2011) that

$$
\begin{aligned}
&\mathbb{E}(\max_{1\le i\le \bar r}\sup_{\boldsymbol{\eta}\in S_{v_n}}|n^{-1/2}\sum_{t=1}^{n}\varepsilon_t\nabla\varepsilon_t(\boldsymbol{\eta})_i|^2)\\
\text{(S1.6)}\qquad &\le l\bar r C[\{\sum_{s=1}^{n-1}\max_{1\le i\le \bar r}\sup_{\boldsymbol{\eta}\in\bigcup_{k=1}^{l}B_{\delta_2}(\boldsymbol{\eta}_k)}\bar b_{s,i}^2(\boldsymbol{\eta})\}\\
&\quad+\{\sum_{s=1}^{n-1}\max_{1\le i\le \bar r}\max_{\mathbf{j}\in\mathbf{J}(m,\bar r),1\le m\le \bar r}\sup_{\boldsymbol{\eta}\in\bigcup_{k=1}^{l}B_{\delta_2}(\boldsymbol{\eta}_k)}(\mathbf{D_j}\bar b_{s,i}(\boldsymbol{\eta}))^2\}]\\
&=O(1).
\end{aligned}
$$

Thus the desired conclusion follows.

The case of conditional heteroscedastic errors: In the above argument, using (3.1), (3.2), and (3.6) to replace the assumptions on $\{\varepsilon_t\}$ in Section 2, and using Lemma 5.2 to replace Lemma B.1 of Chan and Ing (2011), we can still obtain (S1.6), and hence the desired conclusion follows. □

PROOF OF (5.6). *The case of independent errors:* It follows from (S1.1)–(S1.5), (2.13), Lemma B.1 of Chan and Ing (2011), and Markov's inequality that for $\bar M > 2\sigma_\varepsilon^2\sum_{s=1}^{\infty}\sup_{\boldsymbol{\eta}\in\bigcup_{k=1}^{l}B_{\bar\delta_1}(\boldsymbol{\eta}_k)}\bar b_{s,i}^2(\boldsymbol{\eta})$,

$$
\begin{aligned}
&\mathbb{P}(\max_{1\le i\le \bar r}\sup_{\boldsymbol{\eta}\in S_{v_n}}\{n^{-1}\sum_{t=1}^{n}(\nabla\varepsilon_t(\boldsymbol{\eta})_i)^2\}>\bar M)\\
&\le C\max_{1\le i\le \bar r}\mathbb{P}(\sup_{\boldsymbol{\eta}\in S_{\delta_1}}|n^{-1}\sum_{t=1}^{n}(\nabla\varepsilon_t(\boldsymbol{\eta})_i)^2-\mathbb{E}((\nabla\varepsilon_t(\boldsymbol{\eta})_i)^2)|^2>(\frac{\bar M}{2})^2)\\
\text{(S1.7)}\qquad &\le Cn^{-2}\max_{1\le i\le \bar r}[\sum_{u=1}^{n-1}(\sum_{w=1}^{n-u}\bar S_{w,w}(i,i))^2+\sum_{u=1}^{n-1}(\sum_{u=1}^{n-u}\bar V_{w,w}(i,i))^2\\
&\quad+\sum_{u=2}^{n-1}\{\sum_{v=1}^{u-1}(\sum_{w=1}^{n-u}\bar S_{w+u-v,w}(i,i))^2+\sum_{v=1}^{u-1}(\sum_{w=1}^{n-u}\bar S_{w,w+u-v}(i,i))^2\\
&\quad+\sum_{v=1}^{u-1}(\sum_{w=1}^{n-u}\bar V_{w+u-v,w}(i,i))^2+\sum_{v=1}^{u-1}(\sum_{w=1}^{n-u}\bar V_{w,w+u-v}(i,i))^2\}]\\
&\le Cn^{-1+2(\delta_1+\bar\delta_1)}=o(1),
\end{aligned}
$$

where

$$
\bar S_{u,v}(i,j)=\max_{\mathbf{j}\in\mathbf{J}(m,\bar r),1\le m\le \bar r}\sup_{\boldsymbol{\eta}\in\bigcup_{k=1}^{l}B_{\bar\delta_1}(\boldsymbol{\eta}_k)}|\mathbf{D_j}\{\bar b_{u,i}(\boldsymbol{\eta})\bar b_{v,j}(\boldsymbol{\eta})\}|,
$$

and

$$
\bar V_{u,v}(i,j)=\max_{1\le k\le l}|\bar b_{u,i}(\boldsymbol{\eta}_k)\bar b_{v,j}(\boldsymbol{\eta}_k)|.
$$

Thus the desired conclusion follows.
The case of conditional heteroscedastic errors: In the above argument, using (3.1), (3.2), and (3.6) to replace the assumptions on $\{\varepsilon_t\}$ in Section 2, and using Lemma 5.2 to replace Lemma B.1 of Chan and Ing (2011), we can still obtain (S1.7), and hence the desired conclusion follows. □

S2. Proof of Lemma 5.2. Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)^\top \in B_{\bar{\delta}}(\boldsymbol{\theta}_a)$. By (3.10) of Lai (1994), the convexity of $|x|^{m_1}$, and Jensen's inequality, it holds that for all $t \geq 2$,

(S2.1)
$$\begin{aligned}
&|\sum_{t=2}^{n} \varepsilon_t\{K_t(\boldsymbol{\theta}) - K_t(\boldsymbol{\theta}_a)\}|^{m_1} \\
=&|\sum_{m=1}^{k} \sum_{\mathbf{j}\in\mathbf{J}(m,k)} \int\cdots\int_{Q_{\mathbf{j}}(\boldsymbol{\theta}_a,\boldsymbol{\theta})} \sum_{t=2}^{n} (\mathbf{D}_{\mathbf{j}} K_t|_{\xi_j=\theta_{a,j},j\notin\mathbf{j}})\varepsilon_t\, d\xi_{j_1}\cdots d\xi_{j_m}|^{m_1} \\
\leq&2^{k(m_1-1)} \sum_{m=1}^{k} \sum_{\mathbf{j}\in\mathbf{J}(m,k)} |\int\cdots\int_{Q_{\mathbf{j}}(\boldsymbol{\theta}_a,\boldsymbol{\theta})} \sum_{t=2}^{n} (\mathbf{D}_{\mathbf{j}} K_t|_{\xi_j=\theta_{a,j},j\notin\mathbf{j}})\varepsilon_t\, d\xi_{j_1}\cdots d\xi_{j_m}|^{m_1} \\
\leq&2^{k(m_1-1)} \sum_{m=1}^{k} \sum_{\mathbf{j}\in\mathbf{J}(m,k)} \mathrm{vol}^{m_1-1}(B_{\bar{\delta}}(\boldsymbol{\theta}_a, \mathbf{j})) \\
&\times \int\cdots\int_{B_{\bar{\delta}}(\boldsymbol{\theta}_a,\mathbf{j})} |\sum_{t=2}^{n} (\mathbf{D}_{\mathbf{j}} K_t|_{\xi_j=\theta_{a,j},j\notin\mathbf{j}})\varepsilon_t|^{m_1}\, d\xi_{j_1}\cdots d\xi_{j_m},
\end{aligned}$$

where $Q_{\mathbf{j}}(\boldsymbol{\theta}_a, \boldsymbol{\theta})$ denotes the rectangle formed by $(\theta_{a,j_1}, \ldots, \theta_{a,j_m})^\top$ and $(\theta_{j_1}, \ldots, \theta_{j_m})^\top$, $B_{\bar{\delta}}(\boldsymbol{\theta}_a, \mathbf{j})$ denotes the m-dimensional sphere $\{(\xi_{j_1}, \ldots, \xi_{j_m})|(\theta_{a,1}, \ldots, \theta_{a,j_1-1}, \xi_{j_1}, \theta_{a,j_1+1}, \ldots, \theta_{a,j_2-1}, \xi_{j_2}, \ldots, \theta_{a,j_m-1}, \xi_{j_m}, \theta_{a,j_m+1}, \ldots, \theta_k) \in B_{\bar{\delta}}(\boldsymbol{\theta}_a)\}$, and $\mathrm{vol}(\cdot)$ denotes the Euclidean volume. From (S2.1), we have

(S2.2)
$$\begin{aligned}
&\mathbb{E}(\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)} |\sum_{t=2}^{n} \varepsilon_t\{K_t(\boldsymbol{\theta}) - K_t(\boldsymbol{\theta}_a)\}|^{m_1}) \\
\leq&C \max_{\mathbf{j}\in\mathbf{J}(m,k),1\leq m\leq k} \sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)} \mathbb{E}|\sum_{t=2}^{n} \varepsilon_t \mathbf{D}_{\mathbf{j}} K_t(\boldsymbol{\theta})|^{m_1}.
\end{aligned}$$

Moreover, it follows from (3.1), (3.2), (5.36), Burkholder's inequality, Minkowski's inequality, and the Cauchy–Schwarz inequality that for any $\mathbf{j} \in \mathbf{J}(m, k),\ 1 \leq m \leq k$ and $\boldsymbol{\theta} \in B_{\bar{\delta}}(\boldsymbol{\theta}_a)$,

$$\begin{aligned}
&\mathbb{E}|n^{-1/2} \sum_{t=2}^{n} \varepsilon_t \mathbf{D}_{\mathbf{j}} K_t(\boldsymbol{\theta})|^{m_1} \leq C\mathbb{E}|\sum_{t=2}^{n} \varepsilon_t^2 (n^{-1/2} \sum_{i=1}^{t-1} \mathbf{D}_{\mathbf{j}} c_i(\boldsymbol{\theta})\varepsilon_{t-i})^2|^{m_1/2} \\
&\leq C[\sum_{t=2}^{n} \{\mathbb{E}(|\varepsilon_t|^{m_1} |n^{-1/2} \sum_{i=1}^{t-1} \mathbf{D}_{\mathbf{j}} c_i(\boldsymbol{\theta})\varepsilon_{t-i}|^{m_1})\}^{2/m_1}]^{m_1/2} \\
\text{(S2.3)}&\leq C\{\sum_{t=2}^{n} (\mathbb{E}|\varepsilon_t|^{2m_1})^{1/m_1} (\mathbb{E}|n^{-1/2} \sum_{i=1}^{t-1} \mathbf{D}_{\mathbf{j}} c_i(\boldsymbol{\theta})\varepsilon_{t-i}|^{2m_1})^{1/m_1}\}^{m_1/2}
\end{aligned}$$

$$\le C[\sum_{t=2}^{n}\{\mathbb{E}|\sum_{i=1}^{t-1}n^{-1}(\mathbf{D_j}c_i(\boldsymbol{\theta}))^2\varepsilon_{t-i}^2|^{m_1}\}^{1/m_1}]^{m_1/2}$$

$$\le C[\sum_{t=2}^{n}\{\sum_{i=1}^{t-1}n^{-1}(\mathbf{D_j}c_i(\boldsymbol{\theta}))^2(\mathbb{E}|\varepsilon_{t-i}|^{2m_1})^{1/m_1}\}]^{m_1/2}$$

$$\le C(n^{-1}\sum_{t=2}^{n}\sum_{i=1}^{t-1}(\mathbf{D_j}c_i(\boldsymbol{\theta}))^2)^{m_1/2}.$$

An argument similar to (S2.3) also yields

$$\mathbb{E}|n^{-1/2}\sum_{t=2}^{n}\varepsilon_t K_t(\boldsymbol{\theta}_a)|^{m_1} \le C\{n^{-1}\sum_{t=2}^{n}\sum_{i=1}^{t-1}c_i^2(\boldsymbol{\theta}_a)\}^{m_1/2}. \tag{S2.4}$$

Consequently, (5.37) follows from (S2.2)–(S2.4).

To show (5.38), define $r_t(\boldsymbol{\theta}) = K_t(\boldsymbol{\theta})Q_t(\boldsymbol{\theta}) - E(K_t(\boldsymbol{\theta})Q_t(\boldsymbol{\theta}))$. Then, by the convexity of $|x|^{m_1}$,

(S2.5)

$$\mathbb{E}(n^{-1/2}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}|\sum_{t=2}^{n}r_t(\boldsymbol{\theta})|^{m_1})$$

$$\le 2^{m_1-1}\{\mathbb{E}(n^{-1/2}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}|\sum_{t=2}^{n}\{r_t(\boldsymbol{\theta})-r_t(\boldsymbol{\theta}_a)\}|^{m_1}) + \mathbb{E}|n^{-1/2}\sum_{t=2}^{n}r_t(\boldsymbol{\theta}_a)|^{m_1}\}$$

$$:= 2^{m_1-1}\{(I)+(II)\}.$$

By an argument similar to (S2.2), it follows that

$$(I) \le C\max_{\mathbf{j}\in\mathbf{J}(m,k),1\le m\le k}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}\mathbb{E}|n^{-1/2}\sum_{t=2}^{n}\mathbf{D_j}r_t(\boldsymbol{\theta})|^{m_1}. \tag{S2.6}$$

Straightforward calculations give

$$\sum_{t=2}^{n}\mathbf{D_j}r_t(\boldsymbol{\theta}) = \sum_{i=1}^{n-1}g_{n,i}(\boldsymbol{\theta})(\varepsilon_i^2 - \mathbb{E}(\varepsilon_i^2)) + \sum_{l=2}^{n-1}\{\sum_{i=1}^{l-1}h_{n,il}(\boldsymbol{\theta})\varepsilon_i\}\varepsilon_l,$$

where $g_{n,i}(\boldsymbol{\theta}) = \sum_{t=i+1}^{n}\mathbf{D_j}\{c_{t-i}(\boldsymbol{\theta})d_{t-i}(\boldsymbol{\theta})\}$ and $h_{n,il}(\boldsymbol{\theta}) = \sum_{t=l+1}^{n}\mathbf{D_j}\{c_{t-i}(\boldsymbol{\theta})d_{t-l}(\boldsymbol{\theta}) + c_{t-l}(\boldsymbol{\theta})d_{t-i}(\boldsymbol{\theta})\}$. We will show later that

(S2.7)

$$\max_{\mathbf{j}\in\mathbf{J}(m,k),1\le m\le k}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}\mathbb{E}|n^{-1/2}\sum_{i=1}^{n-1}g_{n,i}(\boldsymbol{\theta})(\varepsilon_i^2-\mathbb{E}(\varepsilon_i^2))|^{m_1}$$

$$\le C\{n^{-1}\sum_{i=1}^{n-1}[\sum_{t=i+1}^{n}\max_{\mathbf{j}\in\mathbf{J}(m,k),1\le m\le k}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}|\mathbf{D_j}\{c_{t-i}(\boldsymbol{\theta})d_{t-i}(\boldsymbol{\theta})\}|]^2\}^{m_1/2},$$

and

$$\max_{\mathbf{j}\in\mathbf{J}(m,k),1\le m\le k}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}\mathbb{E}|n^{-1/2}\sum_{l=2}^{n-1}\{\sum_{i=1}^{l-1}h_{n,il}(\boldsymbol{\theta})\varepsilon_i\}\varepsilon_l|^{m_1}$$

$$\leq C\{n^{-1}\sum_{l=2}^{n-1}\sum_{i=1}^{l-1}[\sum_{t=l+1}^{n}\max_{\mathbf{j}\in\mathbf{J}(m,k),1\leq m\leq k}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}|\mathbf{D}_{\mathbf{j}}\{c_{t-i}(\boldsymbol{\theta})d_{t-l}(\boldsymbol{\theta}) + c_{t-l}(\boldsymbol{\theta})d_{t-i}(\boldsymbol{\theta})\}|]^2\}^{m_1/2}. \tag{S2.8}$$

Hence,

$$\begin{aligned}(I) \leq & C[\{n^{-1}\sum_{i=1}^{n-1}[\sum_{t=i+1}^{n}\max_{\mathbf{j}\in\mathbf{J}(m,k),1\leq m\leq k}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}|\mathbf{D}_{\mathbf{j}}\{c_{t-i}(\boldsymbol{\theta})d_{t-i}(\boldsymbol{\theta})\}]^2\}^{m_1/2} \\ & + \{n^{-1}\sum_{l=2}^{n-1}\sum_{i=1}^{l-1}[\sum_{t=l+1}^{n}\max_{\mathbf{j}\in\mathbf{J}(m,k),1\leq m\leq k}\sup_{\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)}|\mathbf{D}_{\mathbf{j}}\{c_{t-i}(\boldsymbol{\theta})d_{t-l}(\boldsymbol{\theta}) \\ & + c_{t-l}(\boldsymbol{\theta})d_{t-i}(\boldsymbol{\theta})\}|]^2\}^{m_1/2}].\end{aligned} \tag{S2.9}$$

By an argument similar to that used in proving (S2.9), it can be shown that

$$\begin{aligned}(II) \leq & C[\{n^{-1}\sum_{i=1}^{n-1}\{\sum_{t=i+1}^{n}|c_{t-i}(\boldsymbol{\theta}_a)d_{t-i}(\boldsymbol{\theta}_a)|\}^2\}^{m_1/2}. \\ & + \{n^{-1}\sum_{l=2}^{n-1}\sum_{i=1}^{l-1}\{\sum_{t=l+1}^{n}|c_{t-i}(\boldsymbol{\theta}_a)d_{t-l}(\boldsymbol{\theta}_a) + c_{t-l}(\boldsymbol{\theta}_a)d_{t-i}(\boldsymbol{\theta}_a)|\}^2\}^{m_1/2}].\end{aligned} \tag{S2.10}$$

The desired result, (5.38), now follows from (S2.5), (S2.9), and (S2.10).

PROOF OF (S2.7). By (3.1), (3.2), (5.36), Minkowski's inequality, and Burkholder's inequality, one has for any $\mathbf{j}\in\mathbf{J}(m,k),\ 1\leq m\leq k$, and $\boldsymbol{\theta}\in B_{\bar{\delta}}(\boldsymbol{\theta}_a)$,

$$\begin{aligned} & = \mathbb{E}|n^{-1/2}\sum_{i=1}^{n-1}g_{n,i}(\boldsymbol{\theta})(\varepsilon_i^2 - \mathbb{E}(\varepsilon_i^2))|^{m_1} \\ & = \mathbb{E}|n^{-1/2}\sum_{i=1}^{n-1}g_{n,i}(\boldsymbol{\theta})\sum_{s=0}^{\infty}\mathbf{a}_s^\top\boldsymbol{w}_{i-s}|^{m_1} \\ & \leq \{\sum_{s=0}^{\infty}(\mathbb{E}|\sum_{i=1}^{n-1}n^{-1/2}g_{n,i}(\boldsymbol{\theta})\mathbf{a}_s^\top\boldsymbol{w}_{i-s}|^{m_1})^{1/m_1}\}^{m_1} \\ & \leq C[\sum_{s=0}^{\infty}\{\mathbb{E}|\sum_{i=1}^{n-1}(n^{-1/2}g_{n,i}(\boldsymbol{\theta})\mathbf{a}_s^\top\boldsymbol{w}_{i-s})^2|^{m_1/2}\}^{1/m_1}]^{m_1} \\ & \leq C[\sum_{s=0}^{\infty}\{\sum_{i=1}^{n-1}(\mathbb{E}|n^{-1/2}g_{n,i}(\boldsymbol{\theta})\mathbf{a}_s^\top\boldsymbol{w}_{i-s}|^{m_1})^{2/m_1}\}^{1/2}]^{m_1} \\ & \leq C\{\sum_{s=0}^{\infty}\|\mathbf{a}_s\|(n^{-1}\sum_{i=1}^{n-1}g_{n,i}^2(\boldsymbol{\theta}))^{1/2}\}^{m_1}\end{aligned} \tag{S2.11}$$

$$\leq C(n^{-1}\sum_{i=1}^{n-1} g_{n,i}^2(\boldsymbol{\theta}))^{m_1/2}.$$

Thus (S2.7) is proved. □

PROOF OF (S2.8). By (3.1), (3.2), (5.36), Minkowski's inequality, Burkholder's inequality, and the Cauchy–Schwarz inequality, we have for any $\mathbf{j} \in \mathbf{J}(m,k),\ 1 \leq m \leq k$, and $\boldsymbol{\theta} \in B_{\bar{\delta}}(\boldsymbol{\theta}_a)$,

$$\begin{aligned}
&\mathbb{E}|n^{-1/2}\sum_{l=2}^{n-1}(\sum_{i=1}^{l-1} h_{n,il}(\boldsymbol{\theta})\varepsilon_i)\varepsilon_l|^{m_1}\\
&\leq C\mathbb{E}|\sum_{l=2}^{n-1}(n^{-1/2}\sum_{i=1}^{l-1} h_{n,il}(\boldsymbol{\theta})\varepsilon_i\varepsilon_l)^2|^{m_1/2}\\
&\leq C\{\sum_{l=2}^{n-1}(\mathbb{E}|n^{-1/2}\sum_{i=1}^{l-1} h_{n,il}(\boldsymbol{\theta})\varepsilon_i\varepsilon_l|^{m_1})^{2/m_1}\}^{m_1/2}\\
&\leq C\{\sum_{l=2}^{n-1}(\mathbb{E}|n^{-1/2}\sum_{i=1}^{l-1} h_{n,il}(\boldsymbol{\theta})\varepsilon_i|^{2m_1})^{1/m_1}(\mathbb{E}|\varepsilon_l|^{2m_1})^{1/m_1}\}^{m_1/2} \qquad \text{(S2.12)}\\
&\leq C[\sum_{l=2}^{n-1}\{\mathbb{E}|\sum_{i=1}^{l-1}(n^{-1/2}h_{n,il}(\boldsymbol{\theta})\varepsilon_i)^2|^{m_1}\}^{1/m_1}]^{m_1/2}\\
&\leq C\{\sum_{l=2}^{n-1}\sum_{i=1}^{l-1}(\mathbb{E}|n^{-1/2}h_{n,il}(\boldsymbol{\theta})\varepsilon_i|^{2m_1})^{1/m_1}]^{m_1/2}\\
&\leq C(n^{-1}\sum_{l=2}^{n-1}\sum_{i=1}^{l-1} h_{n,il}^2(\boldsymbol{\theta}))^{m_1/2},
\end{aligned}$$

which immediately leads to (S2.8). □

S3. Proof of (A.12). By (1.2) and (1.3), we have

$$z_i = \sum_{s=0}^{\infty} \tilde{a}_s \varepsilon_{i-s}, \qquad \text{(S3.1)}$$

where

$$|\tilde{a}_s| \leq C_1 \exp(-C_2 s), \qquad \text{(S3.2)}$$

for some positive constants C_1 and C_2. Therefore, by Minkowski's inequality,

(S3.3)

$$\begin{aligned}
&\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{r=j+1}^{k}\sum_{l=r-j+1}^{n-j}\{z_l z_{l-r+j}-\mathbb{E}(z_l z_{l-r+j})\}|^2\\
=&\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{r=j+1}^{l+j-1\wedge k}\{z_l z_{l-r+j}-\mathbb{E}(z_l z_{l-r+j})\}|^2\\
=&\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{r=j+1}^{l+j-1\wedge k}\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}\tilde{a}_u\tilde{a}_v(\varepsilon_{l-u}\varepsilon_{l-r+j-v}-\mathbb{E}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}))|^2\\
\le&\{\sum_{u=0}^{\infty}\sum_{v=0}^{\infty}|\tilde{a}_u\tilde{a}_v|(\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{r=j+1}^{l+j-1\wedge k}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}-\mathbb{E}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}))|^2)^{1/2}\}^2,
\end{aligned}$$

where $x\wedge y=\min\{x,y\}$. When $v\ge u$, by (3.1), (3.2), (3.6), Burkholder's inequality, Minkowski's inequality, and the Cauchy–Schwarz inequality,

(S3.4)

$$\begin{aligned}
&\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{r=j+1}^{(l+j-1)\wedge k}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}-\mathbb{E}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}))|^2\\
\le&C\mathbb{E}|\sum_{l=2}^{n-j}(\sum_{r=j+1}^{(l+j-1)\wedge k}\frac{1}{\sqrt{(k-j)(n-j)}}\varepsilon_{l-u}\varepsilon_{l-r+j-v})^2|\\
\le&C\sum_{l=2}^{n-j}\mathbb{E}|\sum_{r=j+1}^{(l+j-1)\wedge k}\frac{1}{\sqrt{(k-j)(n-j)}}\varepsilon_{l-u}\varepsilon_{l-r+j-v}|^2\\
\le&C\sum_{l=2}^{n-j}(\mathbb{E}|\varepsilon_{l-u}|^4)^{1/2}(\mathbb{E}|\sum_{r=j+1}^{(l+j-1)\wedge k}\frac{1}{\sqrt{(k-j)(n-j)}}\varepsilon_{l-r+j-v}|^4)^{1/2}\\
\le&C\sum_{l=2}^{n-j}\{\mathbb{E}|\sum_{r=j+1}^{(l+j-1)\wedge k}(\frac{1}{\sqrt{(k-j)(n-j)}}\varepsilon_{l-r+j-v})^2|^2\}^{1/2}\\
\le&C\sum_{l=2}^{n-j}\sum_{r=j+1}^{(l+j-1)\wedge k}(\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\varepsilon_{l-r+j-v}|^4)^{1/2}\le C.
\end{aligned}$$

When $u > v$,

(S3.5)

$$\begin{aligned}
&\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{r=j+1}^{(l+j-1)\wedge k}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}-\mathbb{E}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}))|^2\\
\leq & C\{\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{r=j+1+u-v}^{(l+j-1)\wedge k}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}-\mathbb{E}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}))|^2\\
&+\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{r=j+1}^{(j+u-v-1)\wedge k}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}-\mathbb{E}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}))|^2\\
&+\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}(\varepsilon_{l-u}^2-\mathbb{E}(\varepsilon_{l-u}^2))|^2\},
\end{aligned}$$

noting that $\sum_a^b \cdot = 0$ if $a > b$. By an argument similar to (S3.4), it can be shown that

$$\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{r=j+1+u-v}^{(l+j-1)\wedge k}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}-\mathbb{E}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}))|^2 \leq C. \tag{S3.6}$$

For $u - v = m \in \mathbb{N}$, with $k - j \geq m \geq 2$, we have

(S3.7)

$$\begin{aligned}
&\mathbb{E}|\sum_{l=2}^{n-j}\sum_{r=j+1}^{j+u-v-1}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}-\mathbb{E}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}))|^2\\
\leq & C\{\mathbb{E}|\sum_{s=3-m}^{0}\sum_{i=2-m}^{s-1}\varepsilon_{s-v}\varepsilon_{i-v}|^2\\
&+\mathbb{E}|\sum_{s=1}^{n-j-m+1}\sum_{i=s+1-m}^{s-1}\varepsilon_{s-v}\varepsilon_{i-v}|^2\\
&+\mathbb{E}|\sum_{s=n-j-m+2}^{n-j-1}\sum_{i=s+1-m}^{n-j-m}\varepsilon_{s-v}\varepsilon_{i-v}|^2\}\\
:= & C\{(I)+(II)+(III)\}.
\end{aligned}$$

By an argument similar to (S3.4), it can be shown that

$$(I) \leq Cm^2, \tag{S3.8}$$

$$(II) \leq C(n-j)m, \tag{S3.9}$$

and

$$(III) \leq Cm^2. \tag{S3.10}$$

Similarly, it can be readily shown that for $u - v > k - j$,

$$\mathbb{E}|\sum_{l=2}^{n-j}\sum_{r=j+1}^{k}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}-\mathbb{E}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}))|^2 \leq C(n-j)(k-j). \tag{S3.11}$$

Combining (S3.7)–(S3.11) yields
(S3.12)

$$\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{r=j+1}^{(j+u-v-1)\wedge k}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}-\mathbb{E}(\varepsilon_{l-u}\varepsilon_{l-r+j-v}))|^2\leq C.$$

By (3.1) and (3.2), we have

$$\begin{aligned}&\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}(\varepsilon_{l-u}^2-\mathbb{E}(\varepsilon_{l-u}^2))|^2\\ &=\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{s=0}^{\infty}\mathbf{a}_s^\top\boldsymbol{w}_{l-u-s}|^2.\end{aligned} \tag{S3.13}$$

An argument similar to (S3.4) also leads to

$$\mathbb{E}|\frac{1}{\sqrt{(k-j)(n-j)}}\sum_{l=2}^{n-j}\sum_{s=0}^{\infty}\mathbf{a}_s^\top\boldsymbol{w}_{l-u-s}|^2\leq C. \tag{S3.14}$$

Now, the desired conclusion, (A.12), is an immediate consequence of (S3.2)–(S3.6) and (S3.12)–(S3.14).

REFERENCES

CHAN, N. H. and ING, C. K. (2011). Uniform moment bounds of Fisher's information with applications to time series. *The Annals of Statistics* **39** 1526–1550.

HUALDE, J. and ROBINSON, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *The Annals of Statistics* **39** 3152–3181.

LAI, T. L. (1994). Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *The Annals of Statistics* **22** 1917–1930.

LING, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *The Annals of Statistics* **35** 1213–1237.