The Annals of Statistics
2013, Vol. 41, No. 3, 1268–1298
DOI: 10.1214/13-AOS1110

MOMENT BOUNDS AND MEAN SQUARED PREDICTION ERRORS OF LONG-MEMORY TIME SERIES[1]

BY NGAI HANG CHAN, SHIH-FENG HUANG AND CHING-KANG ING

Chinese University of Hong Kong, National University of Kaohsiung, and Academia Sinica and National Taiwan University

A moment bound for the normalized conditional-sum-of-squares (CSS) estimate of a general autoregressive fractionally integrated moving average (ARFIMA) model with an arbitrary unknown memory parameter is derived in this paper. To achieve this goal, a uniform moment bound for the inverse of the normalized objective function is established. An important application of these results is to establish asymptotic expressions for the one-step and multi-step mean squared prediction errors (MSPE) of the CSS predictor. These asymptotic expressions not only explicitly demonstrate how the multi-step MSPE of the CSS predictor manifests with the model complexity and the dependent structure, but also offer means to compare the performance of the CSS predictor with the least squares (LS) predictor for integrated autoregressive models. It turns out that the CSS predictor can gain substantial advantage over the LS predictor when the integration order is high. Numerical findings are also conducted to illustrate the theoretical results.

1. Introduction. Long-memory behavior has been extensively documented in a spectrum of applications. For background information on long-memory time series and their applications, readers are referred to Doukhan, Oppenheim and Taqqu (2003), where important theories and applications of long-memory models in the areas of finance, insurance, the environment and telecommunications are surveyed. One distinctive feature of the long-memory phenomenon is that the autocorrelation function of a long-memory process decays at a polynomial rate, which is much slower than the exponential rate of a short-memory process. This feature not only enriches the modeling of time series data, but also offers new challenges. While considerable attention has been given in the literature to the derivation of the law of large numbers and the central limit theorem for the estimated parameters in many long-memory time series models [see, e.g., Dahlhaus (1989), Fox and Taqqu (1986), Giraitis and Surgailis (1990), Robinson and Hidalgo (1997) and Robinson (2006)], less attention has been devoted to their moment properties. On the other

Received December 2012; revised March 2013.

[1]Supported in part by Grants from HKSAR-RGC-GRF 400410, NSC 100-2118-M-390-001, NSC 99-2118-M-001-008-MY2 and Academia Sinica Investigator Award.

MSC2010 subject classifications. Primary 62J02; secondary 62M10, 62F12, 60F25.

Key words and phrases. ARFIMA model, integrated AR model, long-memory time series, mean squared prediction error, moment bound, multi-step prediction.

hand, moment properties of the estimated parameters in short-memory time series models have been widely studied. For example, Fuller and Hasza (1981) and Kunitomo and Yamamoto (1985) obtained moment bounds for the least squares (LS) estimators of stationary autoregressive (AR) models, which led to asymptotic expressions for the mean squared prediction error (MSPE) of the corresponding least squares predictors. Ing and Wei (2003) established a moment bound for the inverse of Fisher's information matrix of increasing dimension under a short-memory AR(∞) process, which enabled them to derive an asymptotic expression for the MSPE of the least squares predictor of increasing order. When the moving average (MA) part is taken into account, moment bounds for the estimated parameters are much more difficult to establish, however. Chan and Ing (2011) recently resolved this difficulty by establishing a *uniform* moment bound for the inverse of Fisher's information matrix of nonlinear stochastic regression models. Based on this bound, they analyzed the MSPE of the conditional-sum-of-squares (CSS) predictor (defined in Section 3) and explained how the final prediction error can be used as an effective tool in the model selection of autoregressive moving average (ARMA) models.

These aforementioned studies primarily deal with the stationary cases, which may be inapplicable in many important situations when nonstationary behaviors are often encountered. In view of the importance of incorporating long-memory, short-memory and nonstationary features simultaneously, we are led to consider the following general autoregressive fractionally integrated moving average (ARFIMA) model. Specifically, suppose the data $y_1, \ldots, y_n$ are generated by

$$\begin{aligned}(1.1)\qquad &(1 - \alpha_{0,1}B - \cdots - \alpha_{0,p_1}B^{p_1})(1-B)^{d_0} y_t \\ &\quad = (1 - \beta_{0,1}B - \cdots - \beta_{0,p_2}B^{p_2})\varepsilon_t,\end{aligned}$$

where $\boldsymbol{\eta}_0 = (\boldsymbol{\theta}_0^\top, d_0)^\top = (\alpha_{0,1}, \ldots, \alpha_{0,p_1}, \beta_{0,1}, \ldots, \beta_{0,p_2}, d_0)^\top$ is an unknown coefficient vector with $d_0 \in R$ and $1 - \sum_{j=1}^{p_1} \alpha_{0,j} z^j \neq 0$ and $1 - \sum_{j=1}^{p_2} \beta_{0,j} z^j \neq 0$ for $|z| \leq 1$, B is the back-shift operator and ε_t's are independent random disturbances with $\mathrm{E}(\varepsilon_t) = 0$ for all t. Throughout this paper, it will be assumed that $y_t = \varepsilon_t = 0$ for all $t \leq 0$. These types of initial conditions are commonly used in the nonstationary time series literature; see, for example, Chan and Wei (1988), Hualde and Robinson (2011) and Katayama (2008). Assume that

$$(1.2)\qquad \boldsymbol{\theta}_0 \times d_0 \in \Pi \times D,$$

where

$$(1.3)\qquad D = [L, U] \qquad \text{with } -\infty < L < U < \infty$$

and Π is a compact set in $R^{p_1+p_2}$ whose element $\boldsymbol{\theta} = (\alpha_1, \dots, \alpha_{p_1}, \beta_1, \dots, \beta_{p_2})^\top$ satisfies

$$A_{1,\boldsymbol{\theta}}(z) = 1 - \sum_{j=1}^{p_1} \alpha_j z^j \neq 0, \tag{1.4}$$
$$A_{2,\boldsymbol{\theta}}(z) = 1 - \sum_{j=1}^{p_2} \beta_j z^j \neq 0 \qquad \text{for all } |z| \leq 1;$$

$$A_{1,\boldsymbol{\theta}}(z) \quad \text{and} \quad A_{2,\boldsymbol{\theta}}(z) \text{ have no common zeros;} \tag{1.5}$$

$$|\alpha_{p_1}| + |\beta_{p_2}| > 0. \tag{1.6}$$

Note that in the current setting, D can be any general compact interval in R, which encompasses the important case of nonstationary long-memory models when $d \geq 0.5$.

Let $\varepsilon_t(\boldsymbol{\eta}) = A_{1,\boldsymbol{\theta}}(B) A_{2,\boldsymbol{\theta}}^{-1}(B)(1-B)^d y_t$, where $\boldsymbol{\eta} = (\eta_1, \dots, \eta_{\bar{p}})^\top = (\boldsymbol{\theta}^\top, d)^\top$ with $\bar{p} = p_1 + p_2 + 1$. Then, the CSS estimate of $\boldsymbol{\eta}_0$, $\hat{\boldsymbol{\eta}}_n = (\hat{\boldsymbol{\theta}}_n^\top, \hat{d}_n)^\top$, is given by $\hat{\boldsymbol{\eta}}_n = \arg\min_{\boldsymbol{\eta} \in \Pi \times D} S_n(\boldsymbol{\eta})$, where $S_n(\boldsymbol{\eta}) = \sum_{t=1}^n \varepsilon_t^2(\boldsymbol{\eta})$ is called the objective function. The main goal of this paper is to establish a moment bound for $n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)$, namely,

$$\mathrm{E}\|n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\|^q = O(1), \qquad q \geq 1, \tag{1.7}$$

where $\|\cdot\|$ denotes the Euclidean norm. We focus on model (1.1) instead of more general ones because of its specific and simple short-memory component, which makes our proof much more transparent. On the other hand, it is possible to extend our proof to a broader class of linear processes; see the discussion given at the of Section 2 for details.

Although it is assumed in (1.1) that $\mathrm{E}(y_t) = 0$, this condition is not an issue of overriding concern. To see this, assume that $y_t = \zeta(t) + A_{1,\boldsymbol{\theta}_0}^{-1}(B)(1 - B)^{-d_0} A_{2,\boldsymbol{\theta}_0}(B)\varepsilon_t$ has a mean $\zeta(t)$, where $\zeta(t)$ is a polynomial in t whose degree $k \in \{0, 1, 2, \dots\}$ is known and coefficients are unknown. Then, it is easy to see that $(1 - B)^{k+1} y_t$ is a zero-mean ARFIMA process with memory parameter $d_0 - k - 1$. Given that (1.7) is valid for any value of d_0, the CSS estimate of $\boldsymbol{\eta}_0^* = (\boldsymbol{\theta}_0, d_0 - k - 1)^\top$ based on $(1 - B)^{k+1} y_t$, say $\hat{\boldsymbol{\eta}}_n^*$, still satisfies $\mathrm{E}\|n^{1/2}(\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_0^*)\|^q = O(1), q \geq 1$.

An important and interesting consequence of (1.7) is that asymptotic expressions for the one-step and multi-step MSPEs of the CSS predictor can be established. These asymptotic expressions not only explicitly demonstrate how the multi-step MSPE of the CSS predictor manifests with the model complexity and the dependent structure, but also offer means to compare the performance of the CSS predictor with the LS predictor for integrated AR models. It is worth mentioning that Hualde and Robinson (2011) have shown that $n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)$ converges in distribution to a zero-mean multivariate normal distribution. However, their result cannot be applied to obtain (1.7) because convergence in distribution does not imply convergence of moments. While existence of moments of $\hat{\boldsymbol{\eta}}_n$ can be guaranteed easily by the compactness of $\Pi \times D$, this only yields a bound of $O(n^{q/2})$ for the left-hand side of (1.7), which is greatly improved by the bound on the right-hand side of (1.7). Equation (1.7) can also be used to investigate the higher-order bias and the higher-order efficiency of $\hat{\boldsymbol{\eta}}_n$. Because these types of problems require a separate treatment, they are not pursued in this paper.

Note that under (1.1) with $d_0 > 1/2$, Beran (1995) argued that the consistency and asymptotic normality of $\hat{\boldsymbol{\eta}}_n$ should hold. However, as pointed out by Hualde and Robinson (2011), the proof given in Beran (1995) appears to be incomplete because the property that $\hat{\boldsymbol{\eta}}_n$ lies in a small neighborhood of $\boldsymbol{\eta}_0$ with probability tending to 1 is applied with no justification. Indeed, this property, reliant on uniform probability bounds for $\{S_n(\boldsymbol{\eta}) < S_n(\boldsymbol{\eta}_0)\}$, is difficult to establish for a general d. To circumvent this difficulty, Hualde and Robinson (2011) partitioned the parameter space (after a small ball centered at $\boldsymbol{\eta}_0$ is removed) into four disjoint subsets according to the value of d, and devised different strategies to establish uniform probability bounds for $\{S_n(\boldsymbol{\eta}) < S_n(\boldsymbol{\eta}_0)\}$ over different subsets. Consequently, the consistency and asymptotic normality of $\hat{\boldsymbol{\eta}}_n$ are first rigorously established in Hualde and Robinson (2011) for model (1.1) with a general d. However, the uniform probability bounds given in Hualde and Robinson (2011), converging to zero without rates, are insufficient to establish (1.7). To prove (1.7), we would require rates of convergence of uniform probability bounds, which are in turn ensured by a uniform moment bound of the inverse of the normalized objective function, $a_n^{-1}(d) \sum_{t=1}^{n} (\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t(\boldsymbol{\eta}_0))^2$, where $a_n(d) = n I_{\{d \geq d_0 - 1/2\}} + n^{2(d_0 - d)} I_{\{d < d_0 - 1/2\}}$ with I_B denoting the indicator function of set B. This uniform moment bound, as stipulated and proved in Lemma 2.1, is based on an argument quite different from those in Chan and Ing (2011) and Hualde and Robinson (2011), and constitutes one of the major contributions of this article.

In Section 2, by making use of Lemma 2.1 and other uniform probability/moment bounds, (1.7) is proved in Theorem 2.1. The problem of extending (1.7) to a general linear process that encompasses (1.1) as a special case is also briefly discussed. In Section 3, Lemma 2.1 and Theorem 2.1 are applied to derive asymptotic expressions for the one-step and multi-step MSPEs of the CSS predictors; see Theorems 3.1 and 3.2. These expressions show that whereas the contribution of the estimated parameters to the MSPE, referred to as the second-order MSPE, in the one-step case only involves the number of the estimated parameters, the second-order MSPE in the multi-step case reflects more features of the underlying model, thereby shedding light about the intriguing multi-step prediction behaviors of the ARFIMA processes. Another important implication of Theorems 3.1 and 3.2 is that even for an integrated AR model, the CSS predictor can outperform the LS predictor when the order of integration is large. To facilitate the presentation, more technical proofs are deferred to the Appendix. By means of Monte Carlo simulations, we also demonstrate that the finite-sample behaviors of the one-step and multi-step MSPEs in ARFIMA models can be revealed by the asymptotic results obtained in Section 3. Details of this Monte Carlo study, along with the proof of (2.9), which is the long-memory counterpart of Theorem 3.1 of Chan and Ing (2011) and crucial in proving (1.7), are provided in the supplementary material [Chan, Huang and Ing (2013)] in light of space constraint.

2. Moment bounds. The major goal of this section is to prove (1.7). To this end, we need an assumption on ε_t.

(A1) There exist $0 < \delta_0 \le 1$, $0 < \alpha_0 \le 1$ and $0 < M_1 < \infty$ such that for any $0 < s - v \le \delta_0$, $\sup_{1\le t<\infty, \|\mathbf{v}_t\|=1} |F_{t,\mathbf{v}_t}(s) - F_{t,\mathbf{v}_t}(v)| \le M_1(s-v)^{\alpha_0}$, where $\mathbf{v}_t$ is a t-dimensional vector and $F_{t,\mathbf{v}_t}(\cdot)$ denotes the distribution of $\mathbf{v}_t^{\mathrm{T}}(\varepsilon_t, \ldots, \varepsilon_1)$.

Note that an assumption like (A1) has been used in the literature to deal with the moment properties of the LS estimates in the AR or ARMA context; see, for example, Findley and Wei (2002), Ing (2003), Schorfheide (2005) and Chan and Ing (2011). When ε_t's are normally distributed, (A1) is satisfied with $M_1 = (2\pi\sigma^2)^{-1/2}$ and $\alpha_0 = 1$ for any $\delta_0 > 0$. In addition, when ε_t's are i.i.d. with an integrable characteristic function, (A1) is satisfied with any $\delta_0 > 0$, $\alpha_0 = 1$ and some $M_1 > 0$. For a more detailed discussion of (A1), see Ing and Sin (2006).

The following two lemmas, which may be of independent interest, play a key role in proving (1.7). Let $B_\delta(\boldsymbol{\eta}_0) = \{\boldsymbol{\eta} \in R^{\bar{p}} : \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| < \delta\}$.

LEMMA 2.1. *Assume* (1.1)–(1.6) *and* (A1). *Then, for any* $\delta > 0$ *such that* $\Pi \times D - B_\delta(\boldsymbol{\eta}_0) \ne \varnothing$, *any* $q > 0$ *and any* $\theta > 0$, *we have*

$$\mathrm{E}\left[\left\{\inf_{\boldsymbol{\eta}\in\Pi\times D - B_\delta(\boldsymbol{\eta}_0)} a_n^{-1}(d)\sum_{t=1}^{n}\bigl(\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t(\boldsymbol{\eta}_0)\bigr)^2\right\}^{-q}\right] = O\bigl((\log n)^{\theta}\bigr). \tag{2.1}$$

To perceive the subtlety of Lemma 2.1, first express $a_n^{-1}(d)\sum_{t=1}^{n}(\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t(\boldsymbol{\eta}_0))^2$ as $n^{-1}\sum_{t=1}^{n} g_t^2(\boldsymbol{\eta})$, where $g_t(\boldsymbol{\eta}) = g_{t,n}(\boldsymbol{\eta}) = n^{1/2}(\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t(\boldsymbol{\eta}_0)) \times a_n^{-1/2}(d)$. Since $g_t(\boldsymbol{\eta})$ is a scalar-valued continuous function on $\Pi \times D - B_\delta(\boldsymbol{\eta}_0)$, in view of the proof of Theorem 2.1 of Chan and Ing (2011), (2.1) follows if we can show that $g_t(\boldsymbol{\eta})$ satisfy conditions (C2) and (C3) of the same paper with slight modifications to accommodate the triangular array feature of $g_t(\boldsymbol{\eta})$. However, for $d \le d_0 - 1/2$ and for all large n, the correlation between $g_t(\boldsymbol{\eta})$ and $g_s(\boldsymbol{\eta})$ is overwhelmingly large if $t, s \to \infty$ as $n \to \infty$ and $|t - s|$ is bounded by a positive constant. Therefore, even when (A1) is imposed, it is still difficult to find a positive constant b such that for all large t and n, the conditional distribution of $g_t(\boldsymbol{\eta})$ given $\{\varepsilon_s, s \le t-b\}$ is sufficiently smooth, which corresponds to (C2) of Chan and Ing (2011). Moreover, while $g_t(\boldsymbol{\eta})$ is continuous on $\Pi \times D - B_\delta(\boldsymbol{\eta}_0)$, it is not differentiable at $d = d_0 - 1/2$, making it quite cumbersome to prove that there exist $c_1 > 0$ and nonnegative random variables B_t's satisfying $\max_{1\le t\le n} \mathrm{E}(B_t) = O(1)$ such that for *all* $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in \Pi \times D - B_\delta(\boldsymbol{\eta}_0)$ with $\|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\| < c_1$, $|g_t(\boldsymbol{\eta}_1) - g_t(\boldsymbol{\eta}_2)| \le B_t\|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\|$ a.s., which corresponds to (C3) of Chan and Ing (2011). Indeed, this latter condition is particularly difficult to verify when $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ lie on *different sides* of the hyperplane $d = d_0 - 1/2$. As will be seen in the Appendix, the B_t's derive in (A.8) and (A.9) no longer satisfy $\max_{1\le t\le n} \mathrm{E}(B_t) = O(1)$, which also results in a slowly varying component on the right-hand side of (2.1).

Throughout this paper, C represents a generic positive constant, independent of n, whose value may differ from one occurrence to another.

LEMMA 2.2. *Assume* (1.1)–(1.6), (A1) *and*

$$\sup_{t\ge 1} \mathrm{E}|\varepsilon_t|^{q_1} < \infty, \tag{2.2}$$

where $q_1 > q \ge 2$. *Let* δ *satisfy* $\Pi \times D - B_\delta(\boldsymbol{\eta}_0) \ne \varnothing$ *and* $v > 0$ *be a small constant. Define* $B_{0,v} = \{(\boldsymbol{\theta}^\top, d)^\top : (\boldsymbol{\theta}^\top, d)^\top \in \Pi \times D - B_\delta(\boldsymbol{\eta}_0)$ *with* $d_0 - \frac{1}{2} \le d \le U\}$ *and* $B_{j,v} = \{(\boldsymbol{\theta}^\top, d)^\top : (\boldsymbol{\theta}^\top, d)^\top \in \Pi \times D - B_\delta(\boldsymbol{\eta}_0)$ *with* $d_0 - 1/2 - jv \le d \le d_0 - 1/2 - (j-1)v\}$, $j \ge 1$. *Then, for any* $\theta > 0$,

$$\begin{aligned} &\mathrm{E}\left\{\frac{\sup_{\boldsymbol{\eta}\in B_{j,v}} |\sum_{t=1}^n (\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t(\boldsymbol{\eta}_0))\varepsilon_t|}{\inf_{\boldsymbol{\eta}\in B_{j,v}} \sum_{t=1}^n (\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t(\boldsymbol{\eta}_0))^2}\right\}^q \\ &\quad \le \begin{cases} C\left(\dfrac{(\log n)^{3/2}}{n^{1/2}}\right)^q (\log n)^\theta, & j = 0, \\ C\left(\dfrac{\log n}{n^{1/2+jv-2v}}\right)^q (\log n)^\theta, & j \ge 1. \end{cases} \end{aligned} \tag{2.3}$$

Lemma 2.2 implies

$$P(\hat{\boldsymbol{\eta}}_n \in B_{j,v}) \le P\Big(\inf_{\boldsymbol{\eta}\in B_{j,v}} S_n(\boldsymbol{\eta}) \le S_n(\boldsymbol{\eta}_0)\Big) = O\big(\{(\log n)^{3/2}/n^{1/2}\}^q (\log n)^\theta\big)$$

for $j = 0$, and $O(\{\log n/n^{1/2+jv-2v}\}^q (\log n)^\theta)$ for $j \ge 1$, suggesting that for $d < d_0 - 1/2$, the smaller the value of d, the less likely $\hat{d}_n$ will fall in a neighborhood of d. These probability bounds can suppress the orders of magnitude of $\|n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\|^q$ and $\sup_{\boldsymbol{\eta}\in B_{j,v}} n\{\varepsilon_{n+1}(\boldsymbol{\eta}) - \varepsilon_{n+1}(\boldsymbol{\eta}_0)\}^2$, thereby yielding that $\mathrm{E}\{\|n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\|^q \times I_{\{\hat{\boldsymbol{\eta}}_n\in\Pi\times D - B_\delta(\boldsymbol{\eta}_0)\}}\}$ and $n\mathrm{E}[\{\varepsilon_{n+1}(\hat{\boldsymbol{\eta}}_n) - \varepsilon_{n+1}(\boldsymbol{\eta}_0)\}^2 I_{\{\hat{\boldsymbol{\eta}}_n\in\Pi\times D - B_\delta(\boldsymbol{\eta}_0)\}}]$ are asymptotically negligible; see Corollary 2.1 and Lemma 3.1. As will become clear later, the first moment property is indispensable for proving (1.7), whereas the second one is important in analyzing the MSPE of the CSS predictor. It is also worth mentioning that the order of magnitude of $\sup_{\boldsymbol{\eta}\in B_{j,v}} n\{\varepsilon_{n+1}(\boldsymbol{\eta})-\varepsilon_{n+1}(\boldsymbol{\eta}_0)\}^2$ is $n(\log n)^3$ for $j=0$ and $n^{1+2vj}(\log n)^2$ for $j\geq 1$, which increases as j does; see (3.5) for more details. The next corollary is a direct application of Lemma 2.2.

COROLLARY 2.1. *Suppose that the assumptions of Lemma 2.2 hold. Then, for any* $\delta>0$ *such that* $\Pi\times D-B_\delta(\boldsymbol{\eta}_0)\neq\varnothing$,

$$\mathrm{E}\big\{\|n^{1/2}(\hat{\boldsymbol{\eta}}_n-\boldsymbol{\eta}_0)\|^q I_{\{\hat{\boldsymbol{\eta}}_n\in\Pi\times D-B_\delta(\boldsymbol{\eta}_0)\}}\big\}=o(1). \tag{2.4}$$

PROOF. Since both $\hat{\boldsymbol{\eta}}_n$ and $\boldsymbol{\eta}_0$ are in $\Pi\times D$, $\|\hat{\boldsymbol{\eta}}_n\|$ and $\|\boldsymbol{\eta}_0\|$ are bounded above by a finite constant. Therefore, it suffices for (2.4) to show that

$$P\big(\hat{\boldsymbol{\eta}}_n\in\Pi\times D-B_\delta(\boldsymbol{\eta}_0)\big)=o\big(n^{-q/2}\big). \tag{2.5}$$

Let $q_1>q_1^*>q$ and $0<v<\frac{1}{2}(1-q/q_1^*)$. Without loss of generality, assume that $L=d_0-(1/2)-Wv$ for some large integer $W>0$. Then, it follows from Lemma 2.2 (with $q=q_1^*$) and Chebyshev's inequality that for any $\theta>0$,

$$\begin{aligned}
P\big(\hat{\boldsymbol{\eta}}_n\in\Pi\times D-B_\delta(\boldsymbol{\eta}_0)\big)&\leq\sum_{j=0}^{W}P(\hat{\boldsymbol{\eta}}_n\in B_{j,v})\\
&\leq C\sum_{j=0}^{W}\mathrm{E}\bigg\{\frac{\sup_{\boldsymbol{\eta}\in B_{j,v}}|\sum_{t=1}^n(\varepsilon_t(\boldsymbol{\eta})-\varepsilon_t(\boldsymbol{\eta}_0))\varepsilon_t|}{\inf_{\boldsymbol{\eta}\in B_{j,v}}\sum_{t=1}^n(\varepsilon_t(\boldsymbol{\eta})-\varepsilon_t(\boldsymbol{\eta}_0))^2}\bigg\}^{q_1^*}\\
&\leq C\bigg(\frac{\log n}{n^{1/2-v}}\bigg)^{q_1^*}(\log n)^\theta\\
&=o\big(n^{-q/2}\big),
\end{aligned}$$

which gives (2.5). □

While Theorem 2.1 of Hualde and Robinson (2011) showed that $\hat{\boldsymbol{\eta}}_n\to_p\boldsymbol{\eta}_0$ under substantially weaker assumptions on ε_t, it seems tricky to extend their arguments to obtain a convergence rate like the one given in (2.5), which is critical to proving of (2.4). As a by-product of (2.5), we obtain $\hat{\boldsymbol{\eta}}_n\to\boldsymbol{\eta}_0$ a.s., which follows immediately from (2.5) with $q>2$ and the Borel–Cantelli lemma. The main result is given in the next theorem. First, some notation. For $1\leq m\leq\bar{p}$, define $\mathbf{J}(m,\bar{p})=\{(j_1,\dots,j_m)\colon j_1<\cdots<j_m, j_i\in\{1,\dots,\bar{p}\}, 1\leq i\leq m\}$, and for $\mathbf{j}=(j_1,\dots,j_m)\in\mathbf{J}(m,\bar{p})$ and a smooth function $w=w(\xi_1,\dots,\xi_{\bar{p}})$, let $\mathbf{D}_{\mathbf{j}}w=\partial^m w/\partial\xi_{j_1},\dots,\partial\xi_{j_m}$.

THEOREM 2.1. *Assume* (1.1)–(1.6), (A1),

$$\sup_{t\geq 1} \mathrm{E}|\varepsilon_t|^{4q_1} < \infty, \qquad q_1 > q \geq 1, \tag{2.6}$$

and

$$\boldsymbol{\eta}_0 \in \operatorname{int} \Pi \times D. \tag{2.7}$$

Then (1.7) *holds*.

PROOF. Since by (2.5) or Theorem 2.1 of Hualde and Robinson (2011), $\hat{\boldsymbol{\eta}}_n \to \boldsymbol{\eta}_0$ in probability, and since (2.7) is assumed, there exists $0 < \tau_1 < \{1 - (q/q_1)\}/2$ such that

$$B_{\tau_1}(\boldsymbol{\eta}_0) \subset \Pi \times D \quad \text{and} \quad \lim_{n\to\infty} P\big(\hat{\boldsymbol{\eta}}_n \in B_{\tau_1}(\boldsymbol{\eta}_0)\big) = 1. \tag{2.8}$$

Let $\nabla\varepsilon_t(\boldsymbol{\eta}) = (\nabla\varepsilon_t(\boldsymbol{\eta})_1, \ldots, \nabla\varepsilon_t(\boldsymbol{\eta})_{\bar{p}})^{\mathsf{T}} = (\partial\varepsilon_t(\boldsymbol{\eta})/\partial\eta_1, \ldots, \partial\varepsilon_t(\boldsymbol{\eta})/\partial\eta_{\bar{p}})^{\mathsf{T}}$ and $\nabla^2\varepsilon_t(\boldsymbol{\eta}) = (\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j}) = (\partial^2\varepsilon_t(\boldsymbol{\eta})/\partial\eta_i\,\partial\eta_j)$. Assume first that the following relations hold:

$$\mathrm{E}\left\{\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} \lambda_{\min}^{-\gamma}\left(n^{-1}\sum_{t=1}^{n}\nabla\varepsilon_t(\boldsymbol{\eta})\big(\nabla\varepsilon_t(\boldsymbol{\eta})\big)^{\mathsf{T}}\right)\right\} = O(1) \quad \text{for any } \gamma \geq 1, \tag{2.9}$$

$$\max_{1\leq i,j\leq\bar{p}} \mathrm{E}\left\{\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)}\left|n^{-1/2}\sum_{t=2}^{n}\varepsilon_t\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j}\right|^{q_1}\right\} = O(1), \tag{2.10}$$

$$\max_{1\leq i,j\leq\bar{p},2\leq t\leq n} \mathrm{E}\left\{\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)}\left|\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j}\right|^{4q_1}\right\} = O(1), \tag{2.11}$$

$$\max_{1\leq i\leq\bar{p},2\leq t\leq n} \mathrm{E}\left\{\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)}\left|\nabla\varepsilon_t(\boldsymbol{\eta})_{i}\right|^{4q_1}\right\} = O(1), \tag{2.12}$$

$$P\left\{\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} \lambda_{\min}^{-1}\left(n^{-1}\sum_{t=2}^{n}\nabla\varepsilon_t(\boldsymbol{\eta})\big(\nabla\varepsilon_t(\boldsymbol{\eta})\big)^{\mathsf{T}}\right) > \bar{M}\right\} = O\big(n^{-q}\big) \quad \text{for some } \bar{M} > 0, \tag{2.13}$$

$$P\left\{\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} n^{-1}\sum_{t=2}^{n}\big\|\nabla\varepsilon_t(\boldsymbol{\eta})\big\|^2 > \bar{M}\right\} = O\big(n^{-q}\big) \quad \text{for some } \bar{M} > 0, \tag{2.14}$$

$$\max_{1\leq i,j\leq\bar{p}} P\left\{\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} n^{-1}\sum_{t=2}^{n}\big(\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j}\big)^2 > \bar{M}\right\} = O\big(n^{-q}\big) \quad \text{for some } \bar{M} > 0. \tag{2.15}$$

Then, making use of (2.8)–(2.15) and an argument given in the proof of Theorem 2.2 of Chan and Ing (2011), we obtain

$$\mathrm{E}\big(\|n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\|^q I_{O_{1n}}\big) = O(1), \tag{2.16}$$

where $O_{1n} = \{\hat{\boldsymbol{\eta}}_n \in B_{\tau_1^*}(\boldsymbol{\eta}_0)\}$ with $0 < \tau_1^* < \min\{\tau_1, 3^{-1}\bar{p}^{-1}\bar{M}^{-2}\}$. Moreover, it follows from Corollary 2.1 that

$$\mathrm{E}\big(\|n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\|^q I_{O_{2n}}\big) = O(1), \tag{2.17}$$

where $O_{2n} = \{\hat{\boldsymbol{\eta}}_n \in \Pi \times D - B_{\tau_1^*}(\boldsymbol{\eta}_0)\}$. Combining (2.16) and (2.17) gives the desired conclusion (1.7). To complete the proof, it remains to show that (2.9)–(2.15) are true.

A proof of (2.9), which is similar to that of Theorem 3.1 of Chan and Ing (2011), but needs to be modified with the long-memory effect of $\nabla\varepsilon_t(\boldsymbol{\eta})$, $\boldsymbol{\eta} \in B_{\tau_1}(\boldsymbol{\eta}_0)$, is deferred to the supplementary material [Chan, Huang and Ing (2013)]. To prove (2.15), write

$$\begin{aligned}\varepsilon_t(\boldsymbol{\eta}) &= (1-B)^{d-d_0} A_{2,\boldsymbol{\theta}_0}(B) A_{1,\boldsymbol{\theta}_0}^{-1}(B) A_{1,\boldsymbol{\theta}}(B) A_{2,\boldsymbol{\theta}}^{-1}(B)\varepsilon_t \\ &= \sum_{s=0}^{t-1} b_s(\boldsymbol{\eta})\varepsilon_{t-s},\end{aligned}$$

where $b_0(\boldsymbol{\eta}) = 1$. Then, with $c_{s,ij}(\boldsymbol{\eta}) = \partial^2 b_s(\boldsymbol{\theta})/\partial\eta_i\,\partial\eta_j$ and $b_{s,i}(\boldsymbol{\eta}) = \partial b_s(\boldsymbol{\eta})/\partial\eta_i$, $\nabla\varepsilon_t(\boldsymbol{\eta})_i = \sum_{s=1}^{t-1} b_{s,i}(\boldsymbol{\eta})\varepsilon_{t-s}$ and $\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j} = \sum_{s=1}^{t-1} c_{s,ij}(\boldsymbol{\eta})\varepsilon_{t-s}$. It is clear that $b_{s,i}(\boldsymbol{\eta})$ and $c_{s,ij}(\boldsymbol{\eta})$ have continuous partial derivatives, $\mathbf{D}_{\mathbf{j}} b_{s,i}(\boldsymbol{\eta})$ and $\mathbf{D}_{\mathbf{j}} c_{s,ij}(\boldsymbol{\eta})$, on $B_{\tau_1}(\boldsymbol{\eta}_0)$. Moreover, it follows from arguments similar to those in the proofs of Theorem 4.1 of Ling (2007) and Lemma 4 of Hualde and Robinson (2011) that for any $s \ge 1$,

$$\max_{1\le i,j\le\bar{p}} \sup_{\boldsymbol{\eta}\in B_{\tau_1(\eta_0)}} |c_{s,ij}(\boldsymbol{\eta})| \le C\big(\log(s+1)\big)^2 s^{-1+\tau_1} \tag{2.18}$$

and

$$\max_{1\le i,j\le\bar{p}} \max_{\mathbf{j}\in\mathbf{J}(m,\bar{p}),1\le m\le\bar{p}} \sup_{\boldsymbol{\eta}\in B_{\tau_1(\eta_0)}} |D_{\mathbf{j}} c_{s,ij}(\boldsymbol{\eta})| \le C\big(\log(s+1)\big)^3 s^{-1+\tau_1}. \tag{2.19}$$

Define

$$\begin{aligned}A_s(i,j) &= \max_{\mathbf{j}\in\mathbf{J}(m,\bar{p}),1\le m\le\bar{p}} \sup_{\boldsymbol{\eta}\in B_{\tau_1}(\eta_0)} \big(\mathbf{D}_{\mathbf{j}} c_{s,ij}(\boldsymbol{\eta})\big)^2, \\ S_{r,s}(i,j) &= \max_{\mathbf{j}\in\mathbf{J}(m,\bar{p}),1\le m\le\bar{p}} \sup_{\boldsymbol{\eta}\in B_{\tau_1}(\eta_0)} \big|D_{\mathbf{j}}\{c_{r,ij}(\boldsymbol{\eta})c_{s,ij}(\boldsymbol{\eta})\}\big|\end{aligned}$$

and $B_s(i,j) = \sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} c_{s,ij}^2(\boldsymbol{\eta})$. Equations (2.18) and (2.19) yield that for any $1 \le i, j \le \bar{p}$,

$$\sum_{l=1}^{\infty} c_{l,ij}^2(\boldsymbol{\eta}_0) \le C,$$
$$\left\{\sum_{l=1}^{\infty} S_{l,l}(i,j)\right\}^2 \le C\left[\left\{\sum_{l=1}^{\infty} A_l(i,j)\right\}^2 + \left\{\sum_{l=1}^{\infty} B_l(i,j)\right\}^2\right] \le C. \tag{2.20}$$

On the other hand, by (B.6) of Chan and Ing (2011), Chebyshev's inequality and (2.6), we have for $\bar{M} > 2\sigma^2 \max_{1\le i,j\le r} \sum_{s=1}^{\infty} \sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} c_{s,ij}^2(\boldsymbol{\eta})$ and any $1 \le i$, $j \le \bar{p}$,

$$\begin{aligned}
&P\left\{\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} n^{-1}\sum_{t=2}^{n}(\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j})^2 > \bar{M}\right\} \\
&\quad\le P\left\{\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)}\left|n^{-1}\sum_{t=2}^{n}\left[(\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j})^2 - \mathrm{E}(\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j})^2\right]\right|^{2q_1} > \left(\frac{\bar{M}}{2}\right)^{2q_1}\right\} \\
&\quad\le Cn^{-2q}\left\{\sum_{s=1}^{n-1}\left(\sum_{l=1}^{n-s} c_{l,ij}^2(\boldsymbol{\eta}_0)\right)^2 + \sum_{s=1}^{n-1}\left(\sum_{l=1}^{n-s} S_{l,l}(i,j)\right)^2\right\}^{q_1} \\
&\qquad + Cn^{-q_1-1}\sum_{r=2}^{n-1}\left\{\left[\sum_{s=1}^{r-1}\left(\sum_{l=1}^{n-s} S_{l+r-s,l}(i,j)\right)^2\right]^{q_1}\right. \\
&\qquad\qquad \left. + \left[\sum_{s=1}^{r-1}\left(\sum_{l=1}^{n-s}\left|c_{l+r-s,ij}(\boldsymbol{\eta}_0)c_{l,ij}(\boldsymbol{\eta}_0)\right|\right)^2\right]^{q_1}\right\}.
\end{aligned} \tag{2.21}$$

Since (2.18) and (2.19) also ensure that for some $\tau_1 < \tau_2 < \{1 - (q/q_1)\}/2$, any $1 \le i, j \le \bar{p}$ and any $r > s$, $\{\sum_{l=1}^{\infty} S_{l+r-s,l}(i,j)\}^2 \le C(r-s)^{-1+2\tau_2}$ and $(\sum_{l=1}^{\infty}|c_{l+r-s,ij}(\boldsymbol{\eta}_0)c_{l,ij}(\boldsymbol{\eta}_0)|)^2 \le C(r-s)^{-1+2\tau_2}$, we conclude from these, (2.20) and (2.21) that for any $1 \le i, j \le \bar{p}$,

$$P\left(\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} n^{-1}\sum_{t=2}^{n}(\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j})^2 > \bar{M}\right) = O\left(n^{-q_1(1-2\tau_2)}\right) = o\left(n^{-q}\right).$$

Thus, (2.15) follows.

By analogy with (2.18) and (2.19), we have

$$\max_{1\le i\le\bar{p}} \sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} \left|b_{s,i}(\boldsymbol{\eta})\right| \le C\left(\log(s+1)\right)s^{-1+\tau_1} \tag{2.22}$$

and

$$\max_{1\le i\le\bar{p}} \max_{\mathbf{j}\in\mathbf{J}(m,\bar{p}),1\le m\le\bar{p}} \sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} \left|D_{\mathbf{j}}b_{s,i}(\boldsymbol{\eta})\right| \le C\left(\log(s+1)\right)^2 s^{-1+\tau_1}. \tag{2.23}$$

In addition, (0.3) and (0.4) in the supplementary material [Chan, Huang and Ing (2013)] ensure that there exists $\underline{c} > 0$ such that for all large n,

$$\inf_{\eta \in B_{\tau_1}(\eta_0)} \lambda_{\min}\left(n^{-1}\sum_{t=2}^{n} \mathrm{E}\{\nabla\varepsilon_t(\boldsymbol{\eta})(\nabla\varepsilon_t(\boldsymbol{\eta}))^{\top}\}\right) > \underline{c}. \tag{2.24}$$

Denote $\nabla\varepsilon_t(\boldsymbol{\eta})(\nabla\varepsilon_t(\boldsymbol{\eta}))^{\top}$ and $\mathrm{E}\{\nabla\varepsilon_t(\boldsymbol{\eta})(\nabla\varepsilon_t(\boldsymbol{\eta}))^{\top}\}$ by $W_t(\boldsymbol{\eta})$ and $\bar{W}_t(\boldsymbol{\eta})$, respectively. By making use of

$$\begin{aligned}\inf_{\eta \in B_{\tau_1}(\eta_0)} \lambda_{\min}\left(n^{-1}\sum_{t=2}^{n} W_t(\boldsymbol{\eta})\right) \geq{}& \inf_{\eta \in B_{\tau_1}(\eta_0)} \lambda_{\min}\left(n^{-1}\sum_{t=2}^{n} \bar{W}_t(\boldsymbol{\eta})\right) \\ &- \sup_{\eta \in B_{\tau_1}(\eta_0)}\left\| n^{-1}\sum_{t=2}^{n}[W_t(\boldsymbol{\eta}) - \bar{W}_t(\boldsymbol{\eta})]\right\|,\end{aligned}$$

(B.6) of Chan and Ing (2011), (2.22)–(2.24) and (2.6), we get from an argument similar to that used to prove (2.15) that for $\bar{M} > 2/\underline{c}$,

$$P\left\{\sup_{\eta \in B_{\tau_1}(\eta_0)} \lambda_{\min}^{-1}\left(n^{-1}\sum_{t=2}^{n} W_t(\boldsymbol{\eta})\right) > \bar{M}\right\} = o(n^{-q}),$$

which gives (2.13). As the proof of (2.14) is similar to (2.13), details are omitted.

To prove (2.10), first note that by (2.6) and (B.5) of Chan and Ing (2011), we have for any $1 \leq i, j \leq \bar{p}$,

$$\begin{aligned}&\mathrm{E}\left\{\sup_{\eta \in B_{\tau_1}(\eta_0)}\left| n^{-1/2}\sum_{t=2}^{n} \varepsilon_t \nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j}\right|^{q_1}\right\} \\ &\quad\leq C\left[\left\{\sum_{s=1}^{n-1} c_{s,ij}^2(\boldsymbol{\eta}_0)\right\}^{q_1/2} + \left\{\sum_{s=1}^{n-1} A_s(i,j)\right\}^{q_1/2}\right].\end{aligned} \tag{2.25}$$

Combining (2.25) with (2.20) gives the desired conclusion. Finally, by (2.6), (2.18), (2.19), (2.22), (2.23), Lemma 2 of Wei (1987) and an argument used in the proof of (A.12) in Appendix A, we have for any $1 \leq, i, j \leq \bar{p}$,

$$\begin{aligned}&\mathrm{E}\left\{\sup_{\eta \in B_{\tau_1}(\eta_0)} |\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j}|^{4q_1}\right\} \\ &\quad\leq C\left[\left\{\sum_{s=1}^{\infty} \max_{\mathbf{j} \in \mathbf{J}(m,\bar{p}), 1 \leq m \leq \bar{p}} \sup_{\eta \in B_{\tau_1}(\eta_0)} (\mathbf{D}_{j} c_{s,ij}(\boldsymbol{\eta}))^2\right\}^{2q_1} \right. \\ &\qquad\left. + \left\{\sum_{s=1}^{\infty} c_{s,ij}^2(\boldsymbol{\eta}_0)\right\}^{2q_1}\right] \leq C\end{aligned} \tag{2.26}$$

and

$$
\begin{aligned}
&\mathrm{E}\Big\{\sup_{\eta\in B_{\tau_1}(\eta_0)}\big|\nabla\varepsilon_t(\boldsymbol{\eta})_i\big|^{4q_1}\Big\}\\
&\quad\le C\Bigg[\Bigg\{\sum_{s=1}^{\infty}\max_{\mathbf{j}\in\mathbf{J}(m,\bar{p}),1\le m\le\bar{p}}\sup_{\eta\in B_{\tau_1}(\eta_0)}\big(\mathbf{D}_j b_{s,i}(\boldsymbol{\eta})\big)^2\Bigg\}^{2q_1}\\
&\qquad\qquad+\Bigg\{\sum_{s=1}^{\infty}b_{s,i}^2(\boldsymbol{\eta}_0)\Bigg\}^{2q_1}\Bigg]\le C,
\end{aligned}
\tag{2.27}
$$

and hence (2.11) and (2.12) hold. This completes the proof of Theorem 2.1. □

We close this section with a brief discussion of generalizing (1.7) to the linear process

$$
y_t = m_t(\boldsymbol{\eta}_0) + \varepsilon_t, \tag{2.28}
$$

where ε_t's obey (A1), $\boldsymbol{\eta}_0 = (\boldsymbol{\theta}_0, d_0)^\top$ is an unknown $\bar{p}$-dimensional vector with $d_0 \in D$ and $\boldsymbol{\theta}_0$ lying in a given compact set $V \subset R^{\bar{p}-1}$, and $m_t(\boldsymbol{\eta}) = m_t(\boldsymbol{\eta}, y_{t-1}, \dots, y_1)$ admits a linear representation $\sum_{s=1}^{t-1}\tilde{c}_s(\boldsymbol{\eta})\varepsilon_{t-s}$ with $\tilde{c}_s(\boldsymbol{\eta})$'s being twice differentiable on $V \times D$. Assume that $\boldsymbol{\eta}_0 \in \operatorname{int} V \times D$ and $\tilde{c}_s(\boldsymbol{\eta})$'s satisfy some identifiability conditions leading to (A.19) in the Appendix and (0.1) in the supplementary material [Chan, Huang and Ing (2013)], and some smoothness conditions similar to (2.18), (2.19), (2.22), (2.23) and (A.12). Then the same argument used in the proof of Theorem 2.1 shows that (1.7) is still valid under (2.28). Note that these identifiability and smoothness conditions are readily fulfilled not only by (1.1), but also by (1.1) with the ARMA component being replaced by the exponential-spectrum model of Bloomfield (1973). Moreover, when the ARMA component of (1.1) is replaced by the more general one given in (1.3) of Hualde and Robinson (2011), these conditions can also be ensured by their assumptions A1 and A3, with A1(ii), A2(ii) and A2(iii) suitably modified.

3. Mean squared prediction errors. One important and intriguing application of Theorem 2.1 is the analysis of mean squared prediction errors. Assume that $y_1, \dots, y_n$ are generated by model (1.1). To predict y_{n+h}, $h \ge 1$, based on $y_1, \dots, y_n$, we first adopt the one-step CSS predictor, $\hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n) = y_{n+1} - \varepsilon_{n+1}(\hat{\boldsymbol{\eta}}_n)$, to forecast y_{n+1}, noting that $\hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n)$ depends solely on $y_1, \dots, y_n$. Define $p_t(\boldsymbol{\eta}, y_{t-1}, \dots, y_1) := y_t - \varepsilon_t(\boldsymbol{\eta}) = (1 - (1-B)^d A_{1,\boldsymbol{\theta}}(B) A_{2,\boldsymbol{\theta}}^{-1}(B)) y_t$. Then y_{n+h}, $h \ge 2$, can be predicted recursively by the h-step CSS predictor,

$$
\hat{y}_{n+h}(\hat{\boldsymbol{\eta}}_n) := p_{n+h}\big(\hat{\boldsymbol{\eta}}_n, \hat{y}_{n+h-1}(\hat{\boldsymbol{\eta}}_n), \dots, \hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n), y_n, \dots, y_1\big). \tag{3.1}
$$

When restricted to the short-memory AR case where $p_t(\boldsymbol{\eta}) = (1 - A_{1,\boldsymbol{\theta}}(B)) y_t$, $\hat{y}_{n+h}(\hat{\boldsymbol{\eta}}_n)$ is called the plug-in predictor in Ing (2003). Sections 3.1 and 3.2 provide an asymptotic expression for the MSPE of $\hat{y}_{n+h}(\hat{\boldsymbol{\eta}}_n)$, $\mathrm{E}\{y_{n+h} - \hat{y}_{n+h}(\hat{\boldsymbol{\eta}}_n)\}^2$, with $h = 1$ and $h > 1$, respectively.

3.1. *One-step prediction.* In this section, we apply Theorem 2.1 to analyze $\mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n)\}^2$. In particular, it is shown in Theorem 3.1 that the contribution of the estimated parameters to the one-step MSPE, $\mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n)\}^2 - \sigma^2$, is proportional to the number of parameters. We start with the following auxiliary lemma.

LEMMA 3.1. *Assume* (1.1)–(1.6), (A1) *and*

$$\sup_{t\geq 1} \mathrm{E}|\varepsilon_t|^{\gamma} < \infty \qquad \text{for some } \gamma > 4. \tag{3.2}$$

Then, for any $\delta > 0$ *such that* $\Pi \times D - B_\delta(\boldsymbol{\eta}_0) \neq \varnothing$,

$$n\mathrm{E}\big[\{\varepsilon_{n+1}(\hat{\boldsymbol{\eta}}_n) - \varepsilon_{n+1}(\boldsymbol{\eta}_0)\}^2 I_{\{\hat{\boldsymbol{\eta}}_n \in \Pi\times D - B_\delta(\boldsymbol{\eta}_0)\}}\big] = o(1). \tag{3.3}$$

PROOF. Let $4 < \gamma_1 < \gamma$ and $0 < v < (\gamma_1 - 4)/(2\gamma_1 + 8)$. Also let $B_{j,v}$, $j \geq 0$, be defined as in Lemma 2.2 and W be defined as in the proof of Corollary 2.1. By Cauchy–Schwarz's inequality, the left-hand side of (3.3) is bounded above by

$$n \sum_{j=0}^{W} \mathrm{E}^{1/2}\Big\{\sup_{\boldsymbol{\eta}\in B_{j,v}} \big(\varepsilon_{n+1}(\boldsymbol{\eta}) - \varepsilon_{n+1}(\boldsymbol{\eta}_0)\big)^4\Big\} P^{1/2}(\hat{\boldsymbol{\eta}}_n \in B_{j,v}). \tag{3.4}$$

By the compactness of $B_{j,v}$, (A.32) and (A.33) in the Appendix and an argument similar to that used to prove (2.27), it follows that

$$\begin{aligned}
&\mathrm{E}^{1/2}\Big\{\sup_{\boldsymbol{\eta}\in B_{j,v}} \big(\varepsilon_{n+1}(\boldsymbol{\eta}) - \varepsilon_{n+1}(\boldsymbol{\eta}_0)\big)^4\Big\} \\
&\quad \leq C\Big\{\sup_{\boldsymbol{\eta}\in B_{j,v}} \sum_{s=1}^{n} b_s^2(\boldsymbol{\eta}) + \sum_{s=1}^{n} \max_{\mathbf{j}\in J(m,\bar{p}), 1\leq m\leq \bar{p}} \sup_{\boldsymbol{\eta}\in B_{j,v}} \big(D_{\mathbf{j}} b_s(\boldsymbol{\eta})\big)^2\Big\} \qquad (3.5)\\
&\quad = \begin{cases} O\big((\log n)^3\big), & j = 0, \\ O\big(n^{2vj}(\log n)^2\big), & j \geq 1. \end{cases}
\end{aligned}$$

Moreover, Lemma 2.2 yields that for any $\theta > 0$,

$$\begin{aligned}
P^{1/2}(\hat{\boldsymbol{\eta}}_n \in B_{j,v}) &\leq C\bigg\{\mathrm{E}\bigg(\frac{\sup_{\boldsymbol{\eta}\in B_{j,v}} |\sum_{t=1}^{n}(\varepsilon_t(\boldsymbol{\eta}) - \varepsilon(\boldsymbol{\eta}_0))\varepsilon_t|}{\inf_{\boldsymbol{\eta}\in B_{j,v}} \sum_{t=1}^{n}(\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t(\boldsymbol{\eta}_0))^2}\bigg)^{\gamma_1}\bigg\}^{1/2} \qquad (3.6)\\
&= \begin{cases} O\bigg(\bigg(\dfrac{(\log n)^{3/2}}{n^{1/2}}\bigg)^{\gamma_1/2} (\log n)^{\theta}\bigg), & j = 0, \\[2ex] O\bigg(\bigg(\dfrac{\log n}{n^{1/2+jv-2v}}\bigg)^{\gamma_1/2} (\log n)^{\theta}\bigg), & j \geq 1. \end{cases}
\end{aligned}$$

Combining (3.4)–(3.6), we obtain for some $s > 0$, $n\mathrm{E}[\{\varepsilon_{n+1}(\hat{\boldsymbol{\eta}}_n) - \varepsilon_{n+1}(\boldsymbol{\eta}_0)\}^2 \times I_{\{\hat{\boldsymbol{\eta}}_n\in\Pi\times D - B_\delta(\boldsymbol{\eta}_0)\}}] = O(n^{1+2v-(\gamma_1/2)(1/2-v)}(\log n)^s) = o(1)$, where the last equality is ensured by the ranges of γ_1 and v given above. As a result, (3.3) is proved. □

Equipped with Lemma 3.1, we are now in a position to state and prove Theorem 3.1.

THEOREM 3.1. *Suppose that the assumptions of Theorem* 2.1 *hold except that* (2.6) *is replaced by*

$$\sup_{t\geq 1} \mathrm{E}|\varepsilon_t|^{\gamma} < \infty \qquad \textit{for some } \gamma > 10. \tag{3.7}$$

Assume also that ε_t*'s are i.i.d. random variables. Then*

$$\lim_{n\to\infty} n\big[\mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n)\}^2 - \sigma^2\big] = \bar{p}\sigma^2. \tag{3.8}$$

PROOF. Let $0 < \tau_1 < 1/2$ satisfy $B_{\tau_1}(\boldsymbol{\eta}_0) \subset \Pi \times D$. Define $\mathcal{D}_n = \{\hat{\boldsymbol{\eta}}_n \in B_{\tau_1}(\boldsymbol{\eta}_0)\}$ and $\mathcal{D}_n^c = \{\hat{\boldsymbol{\eta}}_n \in \Pi \times D - B_{\tau_1}(\boldsymbol{\eta}_0)\}$. By Taylor's theorem,

$$\begin{aligned} n^{1/2}&\big(y_{n+1} - \hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n) - \varepsilon_{n+1}\big) \\ &= n^{1/2}\big(\nabla\varepsilon_t(\boldsymbol{\eta}_0)\big)^{\top}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) I_{\mathcal{D}_n} \\ &\quad + \frac{n^{1/2}}{2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)^{\top}\nabla^2\varepsilon_{n+1}\big(\boldsymbol{\eta}^*\big)(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) I_{\mathcal{D}_n} \\ &\quad + n^{1/2}\big(\varepsilon_{n+1}(\hat{\boldsymbol{\eta}}_n) - \varepsilon_{n+1}(\boldsymbol{\eta}_0)\big) I_{\mathcal{D}_n^c}, \end{aligned} \tag{3.9}$$

where $\|\boldsymbol{\eta}^* - \boldsymbol{\eta}_0\| \leq \|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|$. Since Lemma 3.1 ensures that the second moment of the third term on the right-hand side of (3.9) converges to 0, the desired conclusion (3.8) follows immediately from

$$\lim_{n\to\infty} \mathrm{E}\big[n\big\{\big(\nabla\varepsilon_{n+1}(\boldsymbol{\eta}_0)\big)^{\top}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\big\}^2 I_{\mathcal{D}_n}\big] = \bar{p}\sigma^2 \tag{3.10}$$

and

$$\mathrm{E}\big[n\big\{(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)^{\top}\nabla^2\varepsilon_{n+1}\big(\boldsymbol{\eta}^*\big)(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\big\}^2 I_{\mathcal{D}_n}\big] = o(1). \tag{3.11}$$

Note first that by Theorem 2.2 of Hualde and Robinson (2011),

$$n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) \Rightarrow \mathbf{Q}, \tag{3.12}$$

where $\mathbf{Q}$ is distributed as $N(\mathbf{0}, \sigma^2\Gamma^{-1}(\boldsymbol{\eta}_0))$ with $\Gamma(\boldsymbol{\eta}_0)_{i,j} = \lim_{t\to\infty} \mathrm{E}(\nabla\varepsilon_t(\boldsymbol{\eta}_0)_i \times \nabla\varepsilon_t(\boldsymbol{\eta}_0)_j) = \sigma^2 \sum_{s=1}^{\infty} b_{s,i}(\boldsymbol{\eta}_0) b_{s,i}(\boldsymbol{\eta}_0)$ for $1 \leq i, j \leq \bar{p}$ and $\Rightarrow$ denotes convergence in distribution. [Note that $\Gamma(\boldsymbol{\eta}_0)$ is independent of d_0 and $\Gamma(\boldsymbol{\eta}_0)_{\bar{p},\bar{p}} = \pi^2\sigma^2/6$.] Define $\nabla\varepsilon_{n+1,m}(\boldsymbol{\eta}_0) = (\sum_{s=1}^{m} b_{s,i}(\boldsymbol{\eta}_0)\varepsilon_{n+1-s})_{1\leq i\leq\bar{p}}$. Then by (3.12) and the independence between $\nabla\varepsilon_{n+1,m}(\boldsymbol{\eta}_0)$ and $n^{1/2}(\hat{\boldsymbol{\eta}}_{n-m} - \boldsymbol{\eta}_0)$,

$$Z_{n,m} = n^{1/2}\big(\nabla\varepsilon_{n+1,m}(\boldsymbol{\eta}_0)\big)^{\top}(\hat{\boldsymbol{\eta}}_{n-m} - \boldsymbol{\eta}_0) \Rightarrow \mathbf{F}_m^{\top}\mathbf{Q} \qquad \text{as } n \to \infty \tag{3.13}$$

and

$$\mathbf{F}_m^{\top}\mathbf{Q} \Rightarrow \mathbf{F}^{\top}\mathbf{Q} \qquad \text{as } m \to \infty, \tag{3.14}$$

where $\mathbf{F}$ and $\mathbf{F}_m$, independent of $\mathbf{Q}$, have the same distribution as those of $(\sum_{s=1}^{\infty} b_{s,i}(\boldsymbol{\eta}_0)\varepsilon_s,)_{1\le i\le \bar{p}}$ and $\nabla\varepsilon_{m+1,m}(\boldsymbol{\eta}_0)$, respectively. By making use of (2.13), (2.27), (3.12), $\hat{\boldsymbol{\eta}}_{n-m} \to_p \boldsymbol{\eta}_0$ as $n\to\infty$, and $\nabla S_{n-m}(\hat{\boldsymbol{\eta}}_{n-m}) = \mathbf{0}$ on $\{\hat{\boldsymbol{\eta}}_{n-m} \in B_{\tau_1}(\boldsymbol{\eta}_0)\}$, we obtain that for any $\epsilon > 0$,

$$\lim_{m\to\infty}\limsup_{n\to\infty} P\{|n^{1/2}(\nabla\varepsilon_{n+1}(\boldsymbol{\eta}_0))^{\mathsf{T}}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) - Z_{n,m}| > \epsilon\} = 0,$$

which, together with $\lim_{n\to\infty} P(\mathcal{D}_n) = 1$, Theorem 4.2 of Billingsley (1968), (3.13), (3.14) and the continuous mapping theorem, yields

$$n\{(\nabla\varepsilon_{n+1}(\boldsymbol{\eta}_0))^{\mathsf{T}}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\}^2 I_{\mathcal{D}_n} \Rightarrow (\mathbf{F}^{\mathsf{T}}\mathbf{Q})^2. \tag{3.15}$$

Let $5 < v < \gamma/2$ and $\theta = (\gamma/v) - 2$. It follows from (3.7), Theorem 2.1 and Hölder's inequality that

$$\begin{aligned} &\mathrm{E}|(\nabla\varepsilon_{n+1}(\boldsymbol{\eta}_0))^{\mathsf{T}} n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)|^{2+\theta} \\ &\quad\le \{\mathrm{E}\|\nabla\varepsilon_{n+1}(\boldsymbol{\eta}_0)\|^{\gamma}\}^{1/v}\{\mathrm{E}\|n^{1/2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|^{\gamma/(v-1)}\}^{(v-1)/v} \\ &\quad= O(1) \end{aligned} \tag{3.16}$$

and hence $n\{(\nabla\varepsilon_{n+1}(\boldsymbol{\eta}_0))^{\mathsf{T}}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\}^2 I_{\mathcal{D}_n}$ is uniformly integrable. This, (3.15) and $\mathrm{E}(\mathbf{F}^{\mathsf{T}}\mathbf{Q})^2 = \bar{p}\sigma^2$ together imply (3.10).

On the other hand, since on $\mathcal{D}_n$, $\|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\| < \tau_1$, we have for any $0 < \theta < 2$, $\|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|^4 I_{\mathcal{D}_n} \le K\|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|^{2+\theta} I_{\mathcal{D}_n}$, where K is some positive constant depending only on θ and τ_1. Let $0 < \theta < \min\{4^{-1}(\gamma - 2) - 2, 2\}$. Then, it follows from Theorem 2.1, (3.7) and Hölder's inequality that

$$\begin{aligned} &\mathrm{E}\{n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)^{\mathsf{T}}\nabla^2\varepsilon_{n+1}(\boldsymbol{\eta}^*)(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) I_{\mathcal{D}_n}\}^2 \\ &\quad\le K\mathrm{E}(n\|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|^{2+\theta}\|\nabla^2\varepsilon_{n+1}(\boldsymbol{\eta}^*)\|^2) \\ &\quad\le Kn^{-\theta/2}(\mathrm{E}\|n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)\|^{(2+\theta)\gamma/(\gamma-2)})^{(\gamma-2)/\gamma} \\ &\qquad\times\Big\{\mathrm{E}\Big(\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)}\|\nabla^2\varepsilon_{n+1}(\boldsymbol{\eta})\|^{\gamma}\Big)\Big\}^{2/\gamma} \\ &\quad= O(n^{-\theta/2})\Big\{\mathrm{E}\Big(\sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)}\|\nabla^2\varepsilon_{n+1}(\boldsymbol{\eta})\|^{\gamma}\Big)\Big\}^{2/\gamma}. \end{aligned} \tag{3.17}$$

An argument similar to that used to prove (2.26) also yields that the expectation on the right-hand side of (3.17) is bounded above by a finite constant, and hence (3.11) holds true. This completes the proof of the theorem. □

Theorem 3.1 asserts that the second-order MSPE of $\hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n)$, $\bar{p}\sigma^2 n^{-1} + o(n^{-1})$, only depending the number of estimated parameters, has nothing to do with dependent structure of the underlying process. This result is particularly interesting when compared with the second-order MSPE of the LS predictor in integrated AR models. To see this, assume first that there is a forecaster who believes that the true model is possibly an integrated AR(p_1) model,

$$\begin{aligned}&(1-\tilde{\alpha}_1 B-\cdots-\tilde{\alpha}_{p_1}B^{p_1})y_t\\&\quad=(1-B)^{v_0}(1-\theta_1 B-\cdots\theta_{p_1-v_0}B^{p_1-v_0})y_t=\varepsilon_t,\end{aligned}\tag{3.18}$$

where $v_0\in\{0,1,\ldots,p_1\}$ is unknown and $1-\theta_1 z-\cdots\theta_{p_1-d}z^{p_1-v_0}\neq 0$ for all $|z|\leq 1$. Then it is natural for this forecaster to predict y_{n+1} using the LS predictor $\tilde{y}_{n+1}$, in which $\tilde{y}_{n+1}=\mathbf{y}_n^{\top}(p_1)\tilde{\boldsymbol{\alpha}}_n(p_1)$ with $\mathbf{y}_t^{\top}(p_1)=(y_t,\ldots,y_{t-p_1+1})$ and $\tilde{\boldsymbol{\alpha}}_n(p_1)$ satisfies $\sum_{t=p_1}^{n-1}\mathbf{y}_t(p_1)\mathbf{y}_t^{\top}(p_1)\tilde{\boldsymbol{\alpha}}_n(p_1)=\sum_{t=p_1}^{n-1}\mathbf{y}_t(p_1)y_{t+1}$. On the other hand, another forecaster who has doubts on whether the v_0 in (3.18) is really an integer, chooses a more flexible alternative as follows:

$$(1-\alpha_1 B-\cdots-\alpha_{p_1}B^{p_1})(1-B)^{d_0}y_t=\varepsilon_t,\tag{3.19}$$

where $L_1\leq 0\leq d_0\leq p_1\leq U_1$ with $-\infty<L_1<U_1<\infty$ being some prescribed numbers, and $1-\sum_{j=1}^{p_1}\alpha_j z^j$ satisfies (1.4). Clearly, model (3.19), including model (3.18) as a particular case, is itself a special case of model (1.1) with $p_2=0$, and hence the CSS predictor, $\hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n)$, obtained from (3.19) is adopted naturally by the second forecaster.

If the data are truly generated by (3.18), then Theorem 2 of Ing, Sin and Yu (2010) shows that under certain regularity conditions,

$$\lim_{n\to\infty}n[\mathrm{E}\{y_{n+1}-\tilde{y}_{n+1}\}^2-\sigma^2]=(p_1+v_0^2)\sigma^2.\tag{3.20}$$

In addition, by Theorem 3.1 (which is still valid in the case of $p_2=0$), we have

$$\lim_{n\to\infty}n[\mathrm{E}\{y_{n+1}-\hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n)\}^2-\sigma^2]=(p_1+1)\sigma^2.\tag{3.21}$$

As shown in (3.20) and (3.21), while the second-order MSPE of the LS predictor $\tilde{y}_{n+1}$ increases as the strength of dependence in the data does (i.e., v_0 increases), the second-order MSPE of the CSS predictor $\hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n)$ does not vary with v_0. These equalities further indicate the somewhat surprising fact that for an integrated AR model, even the most popular LS predictor can be inferior to the CSS predictor, if the integration order is large. To further illustrate (3.20) and (3.21), we conduct a simulation study to compare the empirical estimates of $n[\mathrm{E}\{y_{n+1}-\tilde{y}_{n+1}\}^2-\sigma^2]$ and $n[\mathrm{E}\{y_{n+1}-\hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n)\}^2-\sigma^2]$ for (3.18) with $p_1=3$ and $v_0=0$, 1 and 2. These estimates, obtained based on 5000 replications for $n=1000$, are summarized in Table 1. As observed in Table 1, the empirical estimates of $n[\mathrm{E}\{y_{n+1}-\hat{y}_{n+1}(\hat{\boldsymbol{\eta}}_n)\}^2-\sigma^2]$ are quite close to 4 for all three models, whereas those of $n[\mathrm{E}\{y_{n+1}-\tilde{y}_{n+1}\}^2-\sigma^2]$ are not distant from 7, 4 and 3 for $v_0=2$, 1 and 0, respectively. Hence all these estimates align with their corresponding limiting values given in (3.20) and (3.21). This "dependency-free" feature of the CSS predictor in the one-step case, however, vanishes in the multi-step case, as will be seen in the next section.

TABLE 1
The empirical estimates of the second-order MSPEs of the CSS predictor (with $p_1 = 3$ and $p_2 = 0$) and the LS predictors with $p_1 = 3$

True model	**CSS predictor**	**LS predictor**
$(1+0.5B)(1-B)^2 y_t = \varepsilon_t$	4.0689	6.8409
$(1-0.25B^2)(1-B)y_t = \varepsilon_t$	4.3732	4.1975
$(1-0.2B-0.25B^2+0.5B^3)y_t = \varepsilon_t$	4.1828	3.1686

3.2. *Multi-step prediction.* Note that under (1.1), $y_t = \sum_{j=0}^{t-1} \underline{c}_s(\boldsymbol{\eta}_0)\varepsilon_{t-s}$, where for $\boldsymbol{\eta} = (\boldsymbol{\theta}^\top, d)^\top = (\alpha_1, \dots, \alpha_{p_1}, \beta_1, \dots, \beta_{p_2}, d)^\top \in \Pi \times D$, $\underline{c}_0(\boldsymbol{\eta}) = 1$ and $\underline{c}_s(\boldsymbol{\eta})$'s, $s \geq 0$, satisfy $\sum_{s=0}^{\infty} \underline{c}_s(\boldsymbol{\eta}) z^s = A_{2,\boldsymbol{\theta}}(z) A_{1,\boldsymbol{\theta}}^{-1}(z)(1-z)^{-d}$, $|z| < 1$. In addition, let $\bar{c}_0(d) = 1$ and $\bar{c}_s(d)$'s, $s \geq 0$, satisfy $\sum_{s=0}^{\infty} \bar{c}_s(d) z^s = (1-z)^{-d}$, $|z| < 1$. With $v_t(d) = (1-B)^d y_t$, define $\mathbf{u}_n(\boldsymbol{\eta}) = (-v_n(d), \dots, -v_{n-p_1+1}(d), \varepsilon_n(\boldsymbol{\eta}), \dots, \varepsilon_{n-p_2+1}(\boldsymbol{\eta}))^\top$. Now the h-step CSS predictor of y_{n+h} is given by $\hat{y}_{n+h}(\hat{\boldsymbol{\eta}}_n) = G_{\hat{d}_n}(B) y_n + \sum_{s=0}^{h-1} \bar{c}_s(\hat{d}_n) \hat{v}_{n+h-s}$, where $G_d(B) = B^{-h} - (1-B)^d \sum_{k=0}^{h-1} \bar{c}_k(d) \times B^{k-h} = (1-B)^d \sum_{k=h}^{\infty} \bar{c}_k(d) B^{k-h}$, $\hat{v}_{n+l} = -\mathbf{u}_n^\top(\hat{\boldsymbol{\eta}}_n) \times A^{l-1}(\hat{\boldsymbol{\theta}}_n)\hat{\boldsymbol{\theta}}_n$ and

$$A(\boldsymbol{\theta}) = \left(\begin{array}{c|c|c|c} \boldsymbol{\alpha} & \begin{array}{c} I_{p_1-1} \\ \hline \mathbf{0}_{p_1-1}^\top \end{array} & \multicolumn{2}{c}{\mathbf{0}_{p_1 \times p_2}} \\ \hline \boldsymbol{\beta} & \mathbf{0}_{p_2 \times (p_1-1)} & \mathbf{0}_{p_2} & \begin{array}{c} I_{p_2-1} \\ \hline \mathbf{0}_{p_2-1}^\top \end{array} \end{array}\right).$$

Here $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{p_1})^\top$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{p_2})^\top$, and $\mathbf{0}_m$, $\mathbf{0}_{m \times n}$ and I_m, respectively, denote the m-dimensional zero vector, the $m \times n$ zero matrix and the m-dimensional identity matrix. Define $\tilde{v}_{n+l}(\boldsymbol{\eta}_0) = -\mathbf{u}_n^\top(\boldsymbol{\eta}_0) A^{l-1}(\boldsymbol{\theta}_0)\boldsymbol{\theta}_0$. Then, it follows that $y_{n+h} = G_{d_0}(B) y_n + \sum_{s=0}^{h-1} \bar{c}_s(d_0) v_{n+h-s}(d_0) = G_{d_0}(B) y_n + \sum_{s=0}^{h-1} \bar{c}_s(d_0) \tilde{v}_{n+h-s}(\boldsymbol{\eta}_0) + \sum_{s=0}^{h-1} \underline{c}_s(\boldsymbol{\eta}_0)\varepsilon_{n+h-s}$. In this section, we establish an asymptotic expression for

$$\begin{aligned} &\mathrm{E}[y_{n+h} - \hat{y}_{n+h}(\hat{\boldsymbol{\eta}}_n)]^2 \\ &\quad = \sigma_h^2(\boldsymbol{\eta}_0) + \mathrm{E}\Biggl\{ G_{\hat{d}_n}(B) y_n + \sum_{s=0}^{h-1} \bar{c}_s(\hat{d}_n) \hat{v}_{n+h-s} \\ &\qquad\qquad - G_{d_0}(B) y_n - \sum_{s=0}^{h-1} \bar{c}_s(d_0) \tilde{v}_{n+h-s}(\boldsymbol{\eta}_0) \Biggr\}^2, \end{aligned} \tag{3.22}$$

where $\sigma_h^2(\boldsymbol{\eta}_0) = \sigma^2 \sum_{s=0}^{h-1} \underline{c}_s^2(\boldsymbol{\eta}_0)$. To state our result, first express $\Gamma(\boldsymbol{\eta}_0)$ as

$$\Gamma(\boldsymbol{\eta}_0) = \begin{pmatrix} \Gamma_{11}(\boldsymbol{\theta}_0) & \boldsymbol{\gamma}_{12}(\boldsymbol{\theta}_0) \\ \boldsymbol{\gamma}_{12}^\top(\boldsymbol{\theta}_0) & \pi^2\sigma^2/6 \end{pmatrix},$$

where $\Gamma_{11}(\boldsymbol{\theta}_0) = (\Gamma(\boldsymbol{\eta}_0)_{i,j})_{1\le i,j\le p_1+p_2}$ and $\boldsymbol{\gamma}_{12}^{\top}(\boldsymbol{\theta}_0) = (\Gamma(\boldsymbol{\eta}_0)_{\bar{p},i})_{1\le i\le p_1+p_2}$, noting that $\Gamma(\boldsymbol{\eta}_0)$ is independent of d_0. Then

$$\Gamma^{-1}(\boldsymbol{\eta}_0) = \begin{pmatrix} \tilde{\Gamma}_{11}(\boldsymbol{\theta}_0) & \tilde{\boldsymbol{\gamma}}_{12}(\boldsymbol{\theta}_0) \\ \tilde{\boldsymbol{\gamma}}_{12}^{\top}(\boldsymbol{\theta}_0) & \tilde{\gamma}_{22}(\boldsymbol{\theta}_0) \end{pmatrix},$$

where $\tilde{\Gamma}_{11}(\boldsymbol{\theta}_0) = (\Gamma_{11}(\boldsymbol{\theta}_0) - \boldsymbol{\gamma}_{12}(\boldsymbol{\theta}_0)\boldsymbol{\gamma}_{12}^{\top}(\boldsymbol{\theta}_0)\gamma_{22}^{-1}(\boldsymbol{\theta}_0))^{-1}$, $\tilde{\gamma}_{22}(\boldsymbol{\theta}_0) = (\pi^2\sigma^2/6 - \boldsymbol{\gamma}_{12}^{\top}(\boldsymbol{\theta}_0)\Gamma_{11}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{\gamma}_{12}(\boldsymbol{\theta}_0))^{-1}$ and $\tilde{\boldsymbol{\gamma}}_{12}(\boldsymbol{\theta}_0) = -\tilde{\gamma}_{22}(\boldsymbol{\theta}_0)\Gamma_{11}^{-1}(\boldsymbol{\theta}_0)\boldsymbol{\gamma}_{12}(\boldsymbol{\theta}_0)$. Define $\nabla^{(1)}\varepsilon_t(\boldsymbol{\theta}_0) = (\nabla\varepsilon_t(\boldsymbol{\eta}_0)_1, \dots, \nabla\varepsilon_t(\boldsymbol{\eta}_0)_{p_1+p_2})^{\top}$ [noting that $\nabla\varepsilon_t(\boldsymbol{\eta}_0)_i$, $1 \le i \le \bar{p}$, is independent of d_0], $\mathbf{w}_{t,h} = (\sum_{k=1}^{t-1}\varepsilon_{t-k}/(k+h-1), \dots, \sum_{k=1}^{t-1}\varepsilon_{t-k}/k)^{\top}$, $Q_h(\boldsymbol{\theta}_0) = \lim_{t\to\infty}\mathrm{E}(\nabla^{(1)}\varepsilon_t(\boldsymbol{\theta}_0)\mathbf{w}_{t,h}^{\top})$ and $R(h) = (\gamma_{i,j})_{h\times h}$, in which $\gamma_{i,i} = 6\pi^{-2}\sum_{l=h-i+1}^{\infty} l^{-2}$, $1 \le i \le h$, and $\gamma_{i,j} = \gamma_{j,i} = 6\pi^{-2}(j-i)^{-1}\sum_{l=1}^{j-i}(h-j+l)^{-1}$, $1 \le i < j \le h$. Now, an asymptotic expression for (3.22) is given as follows.

THEOREM 3.2. *Under the hypothesis of Theorem* 3.1,

$$\lim_{n\to\infty} n\{\mathrm{E}[y_{n+h} - \hat{y}_{n+h}(\hat{\boldsymbol{\eta}}_n)]^2 - \sigma_h^2(\boldsymbol{\eta}_0)\} = \{\underline{f}_h(p_1,p_2) + \underline{g}_h(\boldsymbol{\eta}_0) + 2J_h(\boldsymbol{\eta}_0)\}\sigma^2, \tag{3.23}$$

where $\underline{f}_h(p_1,p_2) = \mathrm{tr}\{\Gamma_{11}(\boldsymbol{\theta}_0)\underline{L}_h(\boldsymbol{\eta}_0)\tilde{\Gamma}_{11}(\boldsymbol{\theta}_0)\underline{L}_h^{\top}(\boldsymbol{\eta}_0)\}$, $\underline{g}_h(\boldsymbol{\eta}_0) = (\pi\sigma^2/6) \times \tilde{\gamma}_{22}(\boldsymbol{\theta}_0)\underline{\mathbf{c}}_h^{\top}(\boldsymbol{\eta}_0)R(h)\underline{\mathbf{c}}_h(\boldsymbol{\eta}_0)$, *and* $J_h(\boldsymbol{\eta}_0) = \tilde{\boldsymbol{\gamma}}_{12}^{\top}(\boldsymbol{\theta}_0)\underline{L}_h^{\top}(\boldsymbol{\eta}_0)Q_h(\boldsymbol{\theta}_0)\underline{\mathbf{c}}_h(\boldsymbol{\eta}_0)$, *with* $\underline{\mathbf{c}}_h(\boldsymbol{\eta}_0) = (\underline{c}_0(\boldsymbol{\eta}_0), \dots, \underline{c}_{h-1}(\boldsymbol{\eta}_0))^{\top}$, $\underline{L}_h(\boldsymbol{\eta}_0) = \sum_{s=0}^{h-1}\underline{c}_s(\boldsymbol{\eta}_0)\,\tilde{A}^{h-1-s}(\boldsymbol{\theta}_0)$ *and*

$$\tilde{A}(\boldsymbol{\theta}) = \left(\begin{array}{c|c} \boldsymbol{\alpha}\left|\begin{array}{c} I_{p_1-1} \\ \hline \mathbf{0}_{p_1-1}^{\top}\end{array}\right. & \mathbf{0}_{p_1\times p_2} \\ \hline \mathbf{0}_{p_2\times p_1} & \boldsymbol{\beta}\left|\begin{array}{c} I_{p_2-1} \\ \hline \mathbf{0}_{p_2-1}^{\top}\end{array}\right. \end{array}\right).$$

A few comments on Theorem 3.2 are in order. When $h = 1$, straightforward calculations imply $\underline{f}_h(p_1,p_2) + \underline{g}_h(\boldsymbol{\eta}_0) + 2J_h(\boldsymbol{\eta}_0) = \bar{p}$, which leads immediately to Theorem 3.1. When $p_1 = p_2 = 0$, $\underline{f}_h(p_1,p_2)$ and $2J_h(\boldsymbol{\eta}_0)$ vanish, and $\tilde{\gamma}_{22}(\boldsymbol{\eta}_0)$ and $\underline{\mathbf{c}}_h(\boldsymbol{\eta}_0)$ in $\underline{g}_h(\boldsymbol{\eta}_0)$ become $6/(\pi^2\sigma^2)$ and $\bar{\mathbf{c}}_h(d_0) = (\bar{c}_0(d_0), \dots, \bar{c}_{h-1}(d_0))^{\top}$, respectively. As a result, the right-hand side of (3.23) is simplified to

$$\bar{g}_h(d_0)\sigma^2 = \bar{\mathbf{c}}_h^{\top}(d_0)R(h)\bar{\mathbf{c}}_h(d_0)\sigma^2, \tag{3.24}$$

yielding the second-order MSPE of the h-step CSS predictor for a pure I(d) process. Alternatively, if $d_0 = 0$ is known, then $\underline{g}_h(\boldsymbol{\eta}_0)$ and $2J_h(\boldsymbol{\eta}_0)$ vanish, and the right-hand side of (3.23) becomes

$$\bar{f}_h(p_1,p_2) = \mathrm{tr}\{\Gamma_{11}(\boldsymbol{\theta}_0)\tilde{L}_h(\boldsymbol{\theta}_0)\Gamma_{11}^{-1}(\boldsymbol{\theta}_0)\tilde{L}_h^{\top}(\boldsymbol{\theta}_0)\}, \tag{3.25}$$

where $\tilde{L}_h(\boldsymbol{\theta}_0)$ is $\underline{L}_h(\boldsymbol{\eta}_0)$ with $d_0 = 0$. Note that (3.25) has been obtained by Yamamoto (1981) under the stationary ARMA(p_1, p_2) model through a somewhat heuristic argument that does not involve the moment bounds of the estimated parameters. When $p_2 = 0$, the right-hand side of (3.25) further reduces to $f_{1,h}(p_1)$ in (10) of Ing (2003), which is the second-order MSPE of the h-step plug-in predictor of a stationary AR(p_1) model. In view of the similarity between $\underline{f}_h(p_1, p_2)$ and $\bar{f}_h(p_1, p_2)$ and that between $\underline{g}_h(\boldsymbol{\eta}_0)$ and $\bar{g}_h(d_0)$, (3.23) displays not only an interesting structure of the multi-step prediction formula from the ARMA case to the I(d) case, and eventually to the ARFIMA case, but also reveals that the multi-step MSPE of an ARFIMA model is the sum of one ARMA term, $\underline{f}_h(p_1, p_2)$, one I($d$) term, $\underline{g}_h(\boldsymbol{\eta}_0)$ and the term $2\underline{J}_h(\boldsymbol{\eta}_0)$ that is related to the ARMA and I(d) joint effect. This expression is different from the ones obtained for the LS predictors in the integrated AR models, in which the AR and I(d) joint effect vanishes asymptotically; see Theorem 2.2 of Ing, Lin and Yu (2009) and Theorem 2 of Ing, Sin and Yu (2010) for details.

Before leaving this section, we remark that the dependence structure of (1.1) has a substantial impact on the multi-step MSPE. To see this, consider the pure I(d) case. By (3.24) and a straightforward calculation, it follows that for any $d_0 \in R$, there exist $0 < C_{1,d_0} \leq C_{2,d_0} < \infty$ such that

$$C_{1,d_0} h^{-1+2d_0} \leq \bar{g}_h(d_0) \leq C_{2,d_0} h^{-1+2d_0}, \tag{3.26}$$

which shows that for $h > 1$, a larger d_0 (or a stronger dependence in the data) tends to result in a larger second-order MSPE. Finally, if the true model is a random walk model, $y_t = \alpha y_{t-1} + \varepsilon_t$, with $\alpha = 1$, and is modeled by an I(d) process, $(1 - B)^d y_t = \varepsilon_t$, in which $d = 1$ corresponds to the true model, then by Theorem 3.2 and (3.24), $\lim_{n\to\infty} n\{\mathrm{E}[y_{n+h} - \hat{y}_{n+h}(\hat{d}_n)]^2 - h\sigma^2\} = \sigma^2$ for $h = 1$, and the limit is smaller than $(4.87h + (1 + \log h)^2 - 2(1 + \log 2h))\sigma^2$ for $h \geq 2$. On the other hand, for the h-step LS predictor $\tilde{y}_{n+h}$ of the above AR(1) model, it follows from Theorem 2.2 of Ing, Lin and Yu (2009) that $\lim_{n\to\infty} n\{\mathrm{E}(y_{n+h} - \tilde{y}_{n+h})^2 - h\sigma^2\} = 2h^2\sigma^2$, which is larger than σ^2 when $h = 1$, and larger than $(4.87h + (1 + \log h)^2 - 2(1 + \log 2h))\sigma^2$ when $h \geq 2$. Hence $\hat{y}_{n+h}(\hat{d}_n)$ is always better than $\tilde{y}_{n+h}$ in terms of the MSPE. The convergence rates of their corresponding estimates, however, are completely reversed because the LS estimate converges much faster to 1 than $\hat{d}_n$ for a random walk model. This finding is reminiscent of the fact that when the true model simultaneously belongs to several different parametric families, the so-called optimal choice of parametric families may vary according to different objectives. For a random walk model, when estimation is the ultimate goal, then LS estimate may be preferable. On the other hand, for prediction purposes, CSS predictor is more desirable according to Theorem 3.2.

APPENDIX

PROOF OF LEMMA 2.1. We only prove (2.1) for $q \geq 1$ because for $0 < q < 1$, (2.1) is an immediate consequence of for the case of $q \geq 1$ and Jensen's inequality.

Since $\Pi \times D - B_\delta(\boldsymbol{\eta}_0)$ is compact, there exists a set of m points $\{\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_m\} \subset \Pi \times D - B_\delta(\boldsymbol{\eta}_0)$ and a small positive number $0 < \delta_1 < 1$, depending possibly on δ and Π, such that

$$\Pi \times D - B_\delta(\boldsymbol{\eta}_0) \subset \bigcup_{i=1}^{m} B_{\delta_1}(\boldsymbol{\eta}_i), \tag{A.1}$$

$$\|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| \geq \delta/2 \quad \text{and} \quad \boldsymbol{\theta} \text{ obeys (1.4)–(1.6)} \tag{A.2}$$

for each $\boldsymbol{\eta} = (\boldsymbol{\theta}^\top, d)^\top \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i)$ and $1 \leq i \leq m$, where $\bar{B}_{\delta_1}(\boldsymbol{\eta}_i)$ denotes the closure of $B_{\delta_1}(\boldsymbol{\eta}_i)$. In view of (A.1), it suffices for (2.1) to show that

$$\mathrm{E}\left[\left\{\inf_{\eta \in \bar{B}_{\delta_1}(\eta_i)} a_n^{-1}(d) \sum_{t=1}^{n} \big(\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t(\boldsymbol{\eta}_0)\big)^2\right\}^{-q}\right] = O\big((\log n)^{\theta}\big), \qquad i = 1, \dots, m, \tag{A.3}$$

hold for any $q \geq 1$ and $\theta > 0$. Let $D_i = \{d : (\boldsymbol{\theta}^\top, d)^\top \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i)\}$, $G_1 = \{i : 1 \leq i \leq m, \bar{D}_i = \sup D_i \leq d_0 - 1/2\}$, $G_2 = \{i : 1 \leq i \leq m, \underline{D}_i = \inf D_i \geq d_0 - 1/2\}$ and $G_3 = \{i : 1 \leq i \leq m, \underline{D}_i < d_0 - 1/2 < \bar{D}_i\}$. Then $\{1, \dots, m\} = \bigcup_{\ell=1}^{3} G_\ell$. In the following, we first prove (A.3) for the most challenging case, $i \in G_3$. The proofs of (A.3) for the cases of $i \in G_1$ or G_2 are similar, but simpler and are thus omitted.

By the convexity of x^{-q}, $x \geq 0$, it follows that for any fixed $0 < \iota < 1$,

$$\begin{aligned}
&\left\{a_n^{-1}(d) \sum_{i=1}^{n} \big(\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t(\boldsymbol{\eta}_0)\big)^2\right\}^{-q} \\
&\quad \leq \left\{n^{-1} \sum_{t=n\iota+1}^{n} g_t^2(\boldsymbol{\eta})\right\}^{-q} \\
&\quad \leq \{\ell q/(1-\iota)\}^q z_n^{-1} \sum_{j=0}^{z_n - 1} \left\{\sum_{r=0}^{\ell q - 1} g_{n\iota+1+rz_n+j}^2(\boldsymbol{\eta})\right\}^{-q},
\end{aligned} \tag{A.4}$$

where $\ell > \max\{(2/\alpha_0) + (\ell_1 + 1)\bar{p}/(\alpha_0 q), (\iota^{-1} - 1)/q\}$, with $\ell_1 > 2q$, $g_t(\boldsymbol{\eta})$ is defined after Lemma 2.1, and $z_n = (1-\iota)n/(\ell q)$. Here $n\iota$, ℓq and z_n are assumed to be positive integers. According to (A.4), if for any $q \geq 1$ and all large n,

$$\mathrm{E}\left[\left\{\inf_{\eta \in \bar{B}_{\delta_1}(\eta_i)} \sum_{r=0}^{\ell q - 1} g_{n\iota+1+z_n r+j}^2(\boldsymbol{\eta})\right\}^{-q}\right] \leq C(\log n)^{5/2}, \qquad j = 0, \dots, z_n - 1, \tag{A.5}$$

holds, then (A.3) follows with $\theta = 5/2$. Moreover, since q is arbitrary, this result is easily extended to any $\theta > 0$ using Jensen's inequality. Consequently, (A.3) is proved. In the rest of the proof, we only show that (A.5) holds for $j=0$ because the proof of (A.5) for $1 \le j \le z_n - 1$ is almost identical. For $j=0$, the left-hand side of (A.5) is bounded above by

$$
\begin{aligned}
&K + \int_K^\infty P\left(\inf_{\boldsymbol{\eta}\in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i)} \sum_{r=0}^{\ell q-1} g^2_{n\iota+1+rz_n}(\boldsymbol{\eta}) < \mu^{-q^{-1}}, R(\mu)\right) d\mu \\
&\quad + \int_K^\infty P\big(R^c(\mu)\big)\, d\mu \qquad\qquad \text{(A.6)}\\
&\quad := K + (\mathrm{I}) + (\mathrm{II}),
\end{aligned}
$$

where K, independent of n and not smaller than 1, will be specified later, with $c_\mu = 2\bar{p}^{1/2}\mu^{-(\ell_1+1)/(2q)}$,

$$
R(\mu) = \bigcap_{r=0}^{\ell q-1} \Big\{ \sup_{\substack{\|\boldsymbol{\eta}_a - \boldsymbol{\eta}_b\| \le c_\mu \\ \boldsymbol{\eta}_a, \boldsymbol{\eta}_b \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i)}} \big| g_{n\iota+1+rz_n}(\boldsymbol{\eta}_a) - g_{n\iota+1+rz_n}(\boldsymbol{\eta}_b)\big| < 2\mu^{-1/(2q)} \Big\},
$$

and $R^c(\mu)$ is the complement of $R(\mu)$.

We first show that

$$
(\mathrm{II}) \le C(\log n)^{5/2}. \qquad \text{(A.7)}
$$

Define $Q_1^{(i)} = \{(\boldsymbol{\theta}^\top, d)^\top : (\boldsymbol{\theta}^\top, d)^\top \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i), d \le d_0 - 1/2\}$ and $Q_2^{(i)} = \{(\boldsymbol{\theta}^\top, d)^\top : (\boldsymbol{\theta}^\top, d)^\top \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i), d \ge d_0 - 1/2\}$. It is clear that $g_t(\boldsymbol{\eta})$ is differentiable on $Q_{1,0}^{(i)}$ and $Q_{2,0}^{(i)}$, the interior of $Q_1^{(i)}$ and $Q_2^{(i)}$, and is continuous on $\bar{B}_{\delta_1}(\boldsymbol{\eta}_i)$. By the mean value theorem, we have for any $\boldsymbol{\eta}_a, \boldsymbol{\eta}_b \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i)$,

$$
\begin{aligned}
&\big|g_t(\boldsymbol{\eta}_a) - g_t(\boldsymbol{\eta}_b)\big| \qquad\qquad \text{(A.8)}\\
&\quad \le \|\boldsymbol{\eta}_a - \boldsymbol{\eta}_b\| \Big(\sup_{\boldsymbol{\eta}\in Q_{1,0}^{(i)}} \|\nabla g_t(\boldsymbol{\eta})\| + \sup_{\boldsymbol{\eta}\in Q_{2,0}^{(i)}} \|\nabla g_t(\boldsymbol{\eta})\|\Big),
\end{aligned}
$$

noting that $\partial g_t(\boldsymbol{\eta})/\partial \bar{p} = \partial g_t(\boldsymbol{\eta})/\partial d$ does not exist at any point in $\bar{B}_{\delta_1}(\boldsymbol{\eta}_i)$ with $d = d_0 - 1/2$. As will be seen later, (A.8) together with

$$
\max_{2\le t\le n} \mathrm{E}\Big(\sup_{\boldsymbol{\eta}\in Q_{v,0}^{(i)}} \|\nabla g_t(\boldsymbol{\eta})\|\Big) = O\big((\log n)^{5/2}\big), \qquad v = 1, 2, \qquad \text{(A.9)}
$$

constitutes a key step in the proof of (A.7).

To prove (A.9) for $v=1$, define $g_t^{(L)}(\boldsymbol{\eta}) = \sqrt{n}\{n(1-B)\}^{d-d_0} A^{-1}_{1,\boldsymbol{\theta}_0}(B) \times A_{2,\boldsymbol{\theta}_0}(B) A_{1,\boldsymbol{\theta}}(B) A^{-1}_{2,\boldsymbol{\theta}}(B)\varepsilon_t - n^{d-d_0+1/2}\varepsilon_t$. Then, $g_t^{(L)}(\boldsymbol{\eta}) = g_t(\boldsymbol{\eta})$ for $\boldsymbol{\eta} \in Q_1^{(i)}$.

In addition, $\nabla g_t^{(L)}(\boldsymbol{\eta}) = ((\nabla g_t^{(L)}(\boldsymbol{\eta}))_i)_{1\le i\le \bar{p}}$ satisfies

$$\begin{aligned}
&\frac{(\nabla g_t^{(L)}(\boldsymbol{\eta}))_k}{n^{1/2}} \\
&= \begin{cases}
-\{n(1-B)\}^{d-d_0}\dfrac{A_{2,\boldsymbol{\theta}_0}(B)}{A_{1,\boldsymbol{\theta}_0}(B)A_{2,\boldsymbol{\theta}}(B)}\varepsilon_{t-k}, \\
\qquad\qquad\qquad\qquad \text{if } 1\le k\le p_1, \\
\{n(1-B)\}^{d-d_0}\dfrac{A_{2,\boldsymbol{\theta}_0}(B)}{A_{1,\boldsymbol{\theta}_0}(B)}\dfrac{A_{1,\boldsymbol{\theta}}(B)}{A_{2,\boldsymbol{\theta}}^2(B)}\varepsilon_{t+p_1-k}, \\
\qquad\qquad\qquad\qquad \text{if } p_1+1\le k\le p_1+p_2, \\
\{n(1-B)\}^{d-d_0}(\log n+\log(1-B))\dfrac{A_{2,\boldsymbol{\theta}_0}(B)}{A_{1,\boldsymbol{\theta}_0}(B)}\dfrac{A_{1,\boldsymbol{\theta}}(B)}{A_{2,\boldsymbol{\theta}}(B)}\varepsilon_t \\
\qquad -(\log n)n^{d-d_0}\varepsilon_t, \qquad \text{if } k=\bar{p}.
\end{cases}
\end{aligned} \tag{A.10}$$

Denote $\boldsymbol{\eta}_i$ by $(\boldsymbol{\theta}_i^{\top}, d_i)^{\top}$. We first consider the case of $d_i \ge d_0 - 1/2$. Write $(\nabla g_t^{(L)}(\boldsymbol{\eta}))_k = \sum_{s=1}^{t-1} c_{s,k}^{(n)}(\boldsymbol{\eta})\varepsilon_{t-s}$, and define $\boldsymbol{\eta}_i^{(L)} = (\eta_{i,1}^{(L)}, \dots, \eta_{i,\bar{p}}^{(L)})^{\top} := (\boldsymbol{\theta}_i^{\top}, d_0 - 1/2)^{\top}$. Clearly, $c_{s,k}^{(n)}(\boldsymbol{\eta})$ has continuous partial derivatives, $D_{\mathbf{j}}c_{s,k}^{(n)}(\boldsymbol{\eta})$ on $Q_1^{(i)}$. Moreover, since for any $\boldsymbol{\eta} = (\eta_1, \dots, \eta_{\bar{p}})^{\top} \in Q_1^{(i)}$, the hypercube formed by $\boldsymbol{\eta}_i^{(L)}$ and $\boldsymbol{\eta}$ is included in $Q_1^{(i)}$, it follows from (3.10) of Lai (1994) and the Cauchy–Schwarz inequality that for any $\boldsymbol{\eta} \in Q_1^{(i)}$, $1 \le k \le \bar{p}$ and $t \ge 2$,

$$\begin{aligned}
&\{(\nabla g_t^{(L)}(\boldsymbol{\eta}))_k - (\nabla g_t^{(L)}(\boldsymbol{\eta}_i^{(L)}))_k\}^2 \\
&= \left\{\sum_{m=1}^{\bar{p}}\sum_{\mathbf{j}\in J(m,\bar{p})}\int_{\eta_{i,j_m}^{(L)}}^{\eta_{j_m}}\cdots\int_{\eta_{i,j_1}^{(L)}}^{\eta_{j_1}}\sum_{s=1}^{t-1}R_{s,\mathbf{j}}^{(k)}\varepsilon_{t-s}\,d\xi_{j_1}\cdots d\xi_{j_m}\right\}^2 \\
&\le 2^{\bar{p}}\sum_{m=1}^{\bar{p}}\sum_{\mathbf{j}\in J(m,\bar{p})}\left\{\int_{\eta_{i,j_m}^{(L)}}^{\eta_{j_m}}\cdots\int_{\eta_{i,j_1}^{(L)}}^{\eta_{j_1}}\sum_{s=1}^{t-1}R_{s,\mathbf{j}}^{(k)}\varepsilon_{t-s}\,d\xi_{j_1}\cdots d\xi_{j_m}\right\}^2 \qquad \text{(A.11)} \\
&\le 2^{\bar{p}}\sum_{m=1}^{\bar{p}}\sum_{\mathbf{j}\in J(m,\bar{p})}\operatorname{vol}(Q_1^{(i)}(m,\mathbf{j})) \\
&\qquad\qquad \times\int\cdots\int_{Q_1^{(i)}(m,\mathbf{j})}\left(\sum_{s=1}^{t-1}R_{s,\mathbf{j}}^{(k)}\varepsilon_{t-s}\right)^2 d\xi_{j_1}\cdots d\xi_{j_m},
\end{aligned}$$

where

$$R_{s,\mathbf{j}}^{(k)} = R_{s,\mathbf{j}}^{(k)}(\xi_{j_1},\dots,\xi_{j_m}) = D_{\mathbf{j}}c_{s,k}^{(n)}(\xi_1,\dots,\xi_{\bar{p}})|_{\xi_j=\eta_{i,j}^{(L)},\, j\notin\mathbf{j}}$$

and

$$
\begin{aligned}
&Q_1^{(i)}(m,\mathbf{j}) \\
&\quad = \{(\eta_{j_1},\dots,\eta_{j_m}) : (\eta_{i,1}^{(L)},\dots,\eta_{i,j_1-1}^{(L)},\eta_{j_1},\eta_{i,j_1+1}^{(L)},\dots,\eta_{i,j_2-1}^{(L)}, \\
&\qquad \eta_{j_2},\eta_{i,j_2+1}^{(L)},\dots,\eta_{i,j_m-1}^{(L)},\eta_{j_m},\eta_{i,j_m+1}^{(L)},\dots,\eta_{i,\bar{p}}^{(L)}) \in Q_1^{(i)}\}, \\
&\qquad \mathbf{j} \in J(m,\bar{p}),
\end{aligned}
$$

is an m-dimensional partial sphere. Now, by (A.2), (A.10), (A.11) and a change of the order of integration, we obtain for $1 \le k \le \bar{p}$,

$$
\begin{aligned}
&\mathrm{E}\Big[\sup_{\boldsymbol{\eta}\in Q_1^{(i)}} \{(\nabla g_t^{(L)}(\boldsymbol{\eta}))_k - (\nabla g_t^{(L)}(\boldsymbol{\eta}_i^{(L)}))_k\}^2\Big] \\
&\qquad \le C \sum_{m=1}^{\bar{p}} \sum_{\mathbf{j}\in J(m,\bar{p})} \mathrm{vol}^2(Q_1^{(i)}(m,\mathbf{j})) \\
&\qquad\qquad \times \left\{ \sum_{s=1}^{t-1} \max_{\mathbf{j}\in J(m,\bar{p}),1\le m\le \bar{p}} \sup_{\boldsymbol{\eta}\in Q_1^{(i)}} (D_{\mathbf{j}} c_{s,k}^{(n)}(\boldsymbol{\eta}))^2 \right\} \\
&\qquad = O((\log n)^5),
\end{aligned} \tag{A.12}
$$

where the last relation follows from for any $1 \le s \le t-1$ with $2 \le t \le n$, $\sup_{\boldsymbol{\eta}\in Q_1^{(i)}} |D_{\mathbf{j}} c_{s,k}^{(n)}(\boldsymbol{\eta})| \le C s^{-1/2}$ if $1 \le k < \bar{p}$ and $\bar{p} \notin \mathbf{j}$; $C(\log n)s^{-1/2}$ if $k = \bar{p}$ and $\bar{p} \notin \mathbf{j}$ or if $1 \le k < \bar{p}$ and $\bar{p} \in \mathbf{j}$; and $C(\log n)^2 s^{-1/2}$ if $k = \bar{p}$ and $\bar{p} \in \mathbf{j}$. Similarly, for all $1 \le t \le n-1$, $\mathrm{E}((\nabla g_t^{(L)}(\boldsymbol{\eta}_i^{(L)}))_k^2) \le C \log n$ if $1 \le k < \bar{p}$, and $C(\log n)^3$ if $k = \bar{p}$. Combining these, with (A.12), yields that for $d_i \ge d_0 - 1/2$,

$$
\begin{aligned}
\max_{2\le t\le n} \mathrm{E}\Big(\sup_{\boldsymbol{\eta}\in Q_{1,0}^{(i)}} \|\nabla g_t(\boldsymbol{\eta})\|^2\Big) &\le \max_{2\le t\le n} \mathrm{E}\Big(\sup_{\boldsymbol{\eta}\in Q_1^{(i)}} \|\nabla g_t^{(L)}(\boldsymbol{\eta})\|^2\Big) \\
&= O((\log n)^5).
\end{aligned} \tag{A.13}
$$

Replacing $\boldsymbol{\eta}_i^{(L)}$ by $\boldsymbol{\eta}_i$ and using an argument similar to that used in proving (A.13), it can be shown that (A.13) is still valid in the case of $d_i < d_0 - 1/2$. As a result, (A.9) holds for $v = 1$.

Define

$$
g_t^{(R)}(\boldsymbol{\eta}) = (1-B)^{d-d_0} A_{2,\boldsymbol{\theta}_0}(B) A_{1,\boldsymbol{\theta}_0}^{-1}(B) A_{1,\boldsymbol{\theta}}(B) A_{2,\boldsymbol{\theta}}^{-1}(B)\varepsilon_t - \varepsilon_t.
$$

Then, $g_t^{(R)}(\boldsymbol{\eta}) = g_t(\boldsymbol{\eta})$ on $\boldsymbol{\eta} \in Q_2^{(i)}$. Moreover, note that

$$(\nabla g_t^{(R)}(\boldsymbol{\eta}))_k = \begin{cases} -(1-B)^{d-d_0} \dfrac{A_{2,\boldsymbol{\theta}_0}(B)}{A_{1,\boldsymbol{\theta}_0}(B) A_{2,\boldsymbol{\theta}}(B)} \varepsilon_{t-k}, \\ \qquad \text{if } 1 \le k \le p_1, \\ (1-B)^{d-d_0} \dfrac{A_{2,\boldsymbol{\theta}_0}(B)}{A_{1,\boldsymbol{\theta}_0}(B)} \dfrac{A_{1,\boldsymbol{\theta}}(B)}{A_{2,\boldsymbol{\theta}}^2(B)} \varepsilon_{t+p_1-k}, \\ \qquad \text{if } p_1 + 1 \le k \le p_1 + p_2, \\ (1-B)^{d-d_0} (\log(1-B)) \dfrac{A_{2,\boldsymbol{\theta}_0}(B)}{A_{1,\boldsymbol{\theta}_0}(B)} \dfrac{A_{1,\boldsymbol{\theta}}(B)}{A_{2,\boldsymbol{\theta}}(B)} \varepsilon_t, \\ \qquad \text{if } k = \bar{p}. \end{cases} \tag{A.14}$$

Using (A.14) in place of (A.10), we can prove (A.9) with $v = 2$ in the same way as $v = 1$. The details are omitted to save space. Equipped with (A.8), (A.9) and Chebyshev's inequality, we obtain

$$\begin{aligned} &\int_K^\infty P\big(R^c(\mu)\big)\, d\mu \\ &\quad \le 2lq \int_K^\infty \max_{1 \le v \le 2, 2 \le t \le n} P\biggl(\sup_{\boldsymbol{\eta} \in Q_{v,0}^{(i)}} \|\nabla g_t(\boldsymbol{\eta})\| \ge \frac{\mu^{\ell_1/(2q)}}{2\bar{p}^{1/2}} \biggr) \, d\mu \\ &\quad \le C(\log n)^{5/2} \int_K^\infty \mu^{-\ell_1/(2q)}\, d\mu \\ &\quad \le C(\log n)^{5/2}, \end{aligned}$$

where the last inequality is ensured by $\ell_1 > 2q$. Hence (A.7) is proved.

In view of (A.6) and (A.7), (A.5) follows by showing that for all large n,

$$\text{(I)} \le C. \tag{A.15}$$

To establish (A.15), motivated by page 1543 of Chan and Ing (2011), we are led to consider the $\bar{p}$-dimensional hypercube, $H^{(i)}$, circumscribing $\bar{B}_{\delta_1}(\boldsymbol{\eta}_i)$. Choose $\mu \ge K$. We then divide $H^{(i)}$ into sub-hypercubes, $H_j^{(i)}(\mu)$ (indexed by j), of equal size, each of which has an edge of length $2\delta_1(\lfloor \mu_1^{(\ell_1+1)/(2q)} \rfloor + 1)^{-1}$ and a circumscribed hypersphere of radius $\sqrt{\bar{p}}\delta_1(\lfloor \mu_1^{(\ell_1+1)/(2q)} \rfloor + 1)^{-1}$, where $\lfloor a \rfloor$ denotes the largest integer $\le a$. Let $G_j^{(i)}(\mu) = \bar{B}_{\delta_1}(\boldsymbol{\eta}_i) \cap H_j^{(i)}(\mu)$ and $\{G_{v_j}^{(i)}(\mu), j = 1, \dots, m^*\}$ be the collection of nonempty $G_j^{(i)}(\mu)$'s. Then it follows that

$$\bar{B}_{\delta_1}(\boldsymbol{\eta}_i) = \bigcup_{j=1}^{m^*} G_{v_j}^{(i)}(\mu) \quad \text{and} \quad m^* \le C^* \mu^{(\ell_1+1)\bar{p}/(2q)}, \tag{A.16}$$

where $C^* > 0$ is independent of μ. In addition, we have

$$\begin{aligned}
A(\mu) &:= \left\{ \inf_{\boldsymbol{\eta} \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i)} \sum_{r=0}^{\ell q-1} g^2_{n\iota+1+rz_n}(\boldsymbol{\eta}) < \mu^{-1/q}, R(\mu) \right\} \\
&\subset \bigcup_{j=1}^{m^*} \bigcap_{r=0}^{\ell q-1} \Big\{ \inf_{\boldsymbol{\eta} \in G^{(i)}_{v_j}(\mu)} |g_{n\iota+1+rz_n}(\boldsymbol{\eta})| < \mu^{-1/(2q)}, \\
&\qquad \sup_{\substack{\|\boldsymbol{\eta}_a - \boldsymbol{\eta}_b\| \le c_\mu \\ \boldsymbol{\eta}_a, \boldsymbol{\eta}_b \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i)}} |g_{n\iota+1+rz_n}(\boldsymbol{\eta}_a) - g_{n\iota+1+rz_n}(\boldsymbol{\eta}_b)| < 2\mu^{-1/(2q)} \Big\} \\
&:= \bigcup_{j=1}^{m^*} \bigcap_{r=0}^{\ell q-1} D_{j,r}(\mu).
\end{aligned} \tag{A.17}$$

Let $\boldsymbol{\eta}^{(j)}_a \in G^{(i)}_{v_j}(\mu)$, $j = 1, \dots, m^*$, be arbitrarily chosen. Then for any $\boldsymbol{\eta} \in G^{(i)}_{v_j}(\mu)$, we have

$$\begin{aligned}
|g_{n\iota+1+rz_n}(\boldsymbol{\eta}^{(j)}_a)| &\le |g_{n\iota+1+rz_n}(\boldsymbol{\eta}^{(j)}_a) - g_{n\iota+1+rz_n}(\boldsymbol{\eta})| + |g_{n\iota+1+rz_n}(\boldsymbol{\eta})| \\
&\le \sup_{\substack{\|\boldsymbol{\eta}_a - \boldsymbol{\eta}_b\| \le c_\mu \\ \boldsymbol{\eta}_a, \boldsymbol{\eta}_b \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i)}} |g_{n\iota+1+rz_n}(\boldsymbol{\eta}_a) - g_{n\iota+1+rz_n}(\boldsymbol{\eta}_b)| + |g_{n\iota+1+rz_n}(\boldsymbol{\eta})|,
\end{aligned}$$

where the last inequality is ensured by $\|\boldsymbol{\eta}^{(j)}_a - \boldsymbol{\eta}\| \le 2\delta_1\sqrt{\bar{p}}(\lfloor \mu^{(\ell_1+1)/(2q)} \rfloor + 1)^{-1} \le c_\mu$, and hence, $D_{j,r}(\mu) \subset S_{j,r}(\mu) := \{|g_{n\iota+1+rz_n}(\boldsymbol{\eta}^{(j)}_a)| < 3\mu^{-1/(2q)}\}$. Combining this with (A.17) yields

$$(\mathrm{I}) \le \int_K^\infty \sum_{j=1}^{m^*} P\left(\bigcap_{r=0}^{\ell q-1} S_{j,r}(\mu) \right) d\mu. \tag{A.18}$$

It is shown at the end of the proof that for some $s_0 > 0$ and all large t,

$$\inf_{\boldsymbol{\eta} \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i)} \mathrm{E}^{1/2}\{(\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t)^2 a_t^{-1}(d)t\} > s_0. \tag{A.19}$$

By letting $K = \max\{1, [(6\sigma/(s_0\delta_0)) \max\{(\ell q/(1-\iota))^{d_0-1/2-L}, 1\}]^{2q}\}$, (A.15) follows. To see this, denote by $\mathcal{F}_t$ the σ-algebra generated by $\{\varepsilon_s, 1 \le s \le t\}$, and recall that $\varepsilon_t(\boldsymbol{\eta}) = \sum_{s=0}^{t-1} b_s(\boldsymbol{\eta})\varepsilon_{t-s}$, with $b_0(\boldsymbol{\eta}) = 1$. We obtain, after some algebraic manipulations,

$$P\left(\bigcap_{r=0}^{\ell q-1} S_{j,r}(\mu) \right) = \mathrm{E}\left[\prod_{r=0}^{\ell q-2} I_{S_{j,r}(\mu)} P(S_{j,\ell q-1}(\mu) | \mathcal{F}_{n+1-2z_n}) \right] \tag{A.20}$$

and

$$P\big(S_{j,\ell q-1}(\mu)|\mathcal{F}_{n+1-2z_n}\big)$$

(A.21)
$$= P\biggl(M_1\big(\boldsymbol{\eta}_a^{(j)},\mu\big) < \frac{\sum_{s=1}^{z_n-1} b_s(\boldsymbol{\eta}_a^{(j)})\varepsilon_{n+1-z_n-s}}{\operatorname{var}^{1/2}((\varepsilon_{z_n}(\boldsymbol{\eta}_a^{(j)})-\varepsilon_{z_n})/\sigma)} < M_2\big(\boldsymbol{\eta}_a^{(j)},\mu\big)\bigg| \mathcal{F}_{n+1-2z_n}\biggr),$$

where

$$M_i(\boldsymbol{\eta},\mu) = \frac{(-1)^i 3\mu^{-1/(2q)} a_n^{1/2}(d) n^{-1/2}}{\operatorname{var}^{1/2}((\varepsilon_{z_n}(\boldsymbol{\eta})-\varepsilon_{z_n})/\sigma)} - \frac{\sum_{s=z_n}^{n-z_n} b_s(\boldsymbol{\eta})\varepsilon_{n+1-z_n-s}}{\operatorname{var}^{1/2}((\varepsilon_{z_n}(\boldsymbol{\eta})-\varepsilon_{z_n})/\sigma)}, \qquad i=1,2.$$

Since (A.19) yields that for $\mu \geq K$ and n sufficiently large,

$$\begin{aligned} &M_2\big(\boldsymbol{\eta}_a^{(j)},\mu\big) - M_1\big(\boldsymbol{\eta}_a^{(j)},\mu\big) \\ &\quad = 6\sigma\mu^{-1/(2q)} \mathrm{E}^{-1/2}\big\{\big(\varepsilon_{z_n}\big(\boldsymbol{\eta}_a^{(j)}\big)-\varepsilon_{z_n}\big)^2 a_{z_n}^{-1}(d) z_n\big\} \frac{a_n^{1/2}(d) n^{-1/2}}{a_{z_n}^{1/2}(d) z_n^{-1/2}} \\ &\quad \leq 6\sigma\mu^{-1/(2q)} s_0^{-1} \max\biggl\{\biggl(\frac{\ell q}{1-\iota}\biggr)^{d_0-1/2-L}, 1\biggr\} \leq \delta_0 \end{aligned}$$

and since $\sum_{s=1}^{z_n-1} b_s^2(\boldsymbol{\eta}_a^{(j)}) = \operatorname{var}((\varepsilon_{z_n}(\boldsymbol{\eta}_a^{(j)})-\varepsilon_{z_n})/\sigma)$, it follows from (A1), (A.20) and (A.21) that

$$\begin{aligned} &P\biggl(\bigcap_{t=0}^{\ell q-1} S_{j,r}(\mu)\biggr) \\ &\quad \leq M_1\biggl(6\sigma\mu^{-1/(2q)} s_0^{-1} \max\biggl\{\biggl(\frac{\ell q}{1-\iota}\biggr)^{d_0-1/2-L}, 1\biggr\}\biggr)^{\alpha_0} \mathrm{E}\biggl(\prod_{r=0}^{\ell q-2} I_{S_{j,r}(\mu)}\biggr). \end{aligned}$$

Moreover, since $\ell > q^{-1}(\iota^{-1}-1)$, we can repeat the same argument ℓq times to get

(A.22)
$$\begin{aligned} &P\biggl(\bigcap_{r=0}^{\ell q-1} S_{j,r}(\mu)\biggr) \\ &\quad \leq M_1^{\ell q}\biggl(6\sigma\mu^{-1/(2q)} s_0^{-1} \max\biggl\{\biggl(\frac{\ell q}{1-\iota}\biggr)^{d_0-1/2-L}, 1\biggr\}\biggr)^{\alpha_0 \ell q}. \end{aligned}$$

By noting that the bound on the right-hand side of (A.22) is independent of j and $\ell > (2/\alpha_0) + \{(\ell_1 + 1)\bar{p}/(\alpha_0 q)\}$, it is concluded from (A.16), (A.18) and (A.22) that for all large n, (I) is bounded above by

$$C^* M_1^{\ell q} \bigl(6\sigma s_0^{-1} \max\{(\ell q/(1-\iota))^{d_0 - 1/2 - L}, 1\}\bigr)^{\alpha_0 \ell q} \times \int_K^\infty \mu^{-(\alpha_0 \ell/2 - (\ell_1+1)\bar{p}/(2q))}\, d\mu \le C.$$

To complete the proof of the lemma, it only remains to show (A.19). Define

$$U_{t,\boldsymbol{\theta}} = \frac{A_{2,\boldsymbol{\theta}_0}(B)}{A_{1,\boldsymbol{\theta}_0}(B)} \frac{A_{1,\boldsymbol{\theta}}(B)}{A_{2,\boldsymbol{\theta}}(B)} \varepsilon_t = \sum_{s=0}^{t-1} d_s(\boldsymbol{\theta}) \varepsilon_{t-s}, \tag{A.23}$$

noting that $d_0(\boldsymbol{\theta}) = 1$. By (1.4)–(1.6), there exist $0 < c_1, c_2 < \infty$ such that

$$\sup_{\boldsymbol{\theta} \in \Pi} |d_s(\boldsymbol{\theta})| \le c_1 \exp(-c_2 s). \tag{A.24}$$

Let $\Sigma_t(\boldsymbol{\theta}) = \mathrm{E}[(U_{1,\boldsymbol{\theta}}, \dots, U_{t,\boldsymbol{\theta}})^\top (U_{1,\boldsymbol{\theta}}, \dots, U_{t,\boldsymbol{\theta}})]$. It can be shown that there exists $C_0 > 0$ such that for all $t \ge 1$,

$$\inf_{\boldsymbol{\theta} \in \Pi} \lambda_{\min}(\Sigma_t(\boldsymbol{\theta})) \ge C_0 \sigma^2. \tag{A.25}$$

Express $\varepsilon_t(\boldsymbol{\eta})$ as $(1-B)^{d-d_0} U_{t,\boldsymbol{\theta}} = \sum_{s=0}^{t-1} v_s(d) U_{t-s,\boldsymbol{\theta}}$. Then, there exists $G = G(L, U) > 0$, depending only on L and U, such that for any $s \ge 0$ and $d \in D$,

$$|v_s(d)| \le G(s+1)^{d_0 - d - 1}. \tag{A.26}$$

Let $0 < \eta^* < 1/2$ be given. Straightforward calculations yield that there exists $C_{\eta^*} > 0$ such that for any $s \ge 0$ and $L \le d \le d_0 - \eta^*$,

$$|v_s(d)| \ge C_{\eta^*}(s+1)^{d_0 - d - 1}. \tag{A.27}$$

Let $\iota_1 > 0$ be small enough such that

$$\begin{gathered} 0 < \iota_1 < \tfrac{1}{2} - \eta^*, \qquad d_0 - \tfrac{1}{2} - \iota_1 > \underline{D}_i, \\ d_0 - \tfrac{1}{2} + \iota_1 < \bar{D}_i \quad \text{and} \quad C_0 C_{\eta^*}^2 (\log \iota_1^{-1}) \iota_1^{2\iota_1} > 2. \end{gathered} \tag{A.28}$$

Define

$$\begin{aligned}
A_1 &= \{(\boldsymbol{\theta}^\top, d)^\top : (\boldsymbol{\theta}^\top, d)^\top \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i), \underline{D}_i \le d \le d_0 - \tfrac{1}{2} - \iota_1\}, \\
A_2 &= \{(\boldsymbol{\theta}^\top, d)^\top : (\boldsymbol{\theta}^\top, d)^\top \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i), d_0 - \tfrac{1}{2} - \iota_1 < d < d_0 - \tfrac{1}{2}\}, \\
A_3 &= \{(\boldsymbol{\theta}^\top, d)^\top : (\boldsymbol{\theta}^\top, d)^\top \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i), d = d_0 - \tfrac{1}{2}\}, \\
A_4 &= \{(\boldsymbol{\theta}^\top, d)^\top : (\boldsymbol{\theta}^\top, d)^\top \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i), d_0 - \tfrac{1}{2} < d < d_0 - \tfrac{1}{2} + \iota_1\}, \\
A_5 &= \{(\boldsymbol{\theta}^\top, d)^\top : (\boldsymbol{\theta}^\top, d)^\top \in \bar{B}_{\delta_1}(\boldsymbol{\eta}_i), d_0 - \tfrac{1}{2} + \iota_1 \le d \le \bar{D}_i\}.
\end{aligned}$$

Then, (A.19) is ensured by showing that there exist $\zeta_i > 0$, $i = 1, \dots, 5$, such that for all large t

$$\inf_{\boldsymbol{\eta} \in A_i} \mathrm{E}\{(\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t)^2 a_t^{-1}(d)t\} > \zeta_i, \qquad i = 1, \dots, 5. \tag{A.29}$$

To show (A.29) with $i = 1$, we deduce from (A.25), (A.27), (A.28) and a straightforward calculation that for all $\boldsymbol{\eta} \in A_1$ and $t \geq 2$,

$$\begin{aligned} \mathrm{E}(\varepsilon_t^2(\boldsymbol{\eta}) a_t^{-1}(d)t) &\geq \frac{\sigma^2 C_0 C_{\eta^*}^2}{t^{2(d_0-d)-1}} \sum_{s=0}^{t-1} (s+1)^{2(d_0-d)-2} \\ &\geq \frac{\sigma^2 C_0 C_{\eta^*}^2 \{1 - (1/2)^{2\iota_1}\}}{2(d_0 - \underline{D}_i) - 1}. \end{aligned} \tag{A.30}$$

In addition, it is clear that $\sup_{\boldsymbol{\eta} \in A_1} \mathrm{E}(\varepsilon_t^2 a_t^{-1}(d)t) \leq \sigma^2 / t^{2\iota_1} \to 0$, as $t \to \infty$, which, together with (A.30), yields (A.29) with $i = 1$. By (A.28), Taylor's theorem and an argument similar to that of (A.30), we have for all $\eta \in A_2$ and sufficiently large t,

$$\begin{aligned} \mathrm{E}(\varepsilon_t^2(\boldsymbol{\eta}) a_t^{-1}(d)t) &\geq \frac{\sigma^2 C_0 C_{\eta^*}^2}{t^{2(d_0-d)-1}} \int_{\iota_1 t}^{t} x^{2(d_0-d)-2}\, dx \\ &\geq \frac{\sigma^2 C_0 C_{\eta^*}^2}{2(d_0 - d) - 1} (1 - \iota_1^{2(d_0-d)-1}) \\ &\geq \sigma^2 C_0 C_{\eta^*}^2 (\log \iota_1^{-1}) \iota_1^{2\iota_1} > 2\sigma^2. \end{aligned} \tag{A.31}$$

Moreover, $\sup_{\boldsymbol{\eta} \in A_2} \mathrm{E}(\varepsilon_t^2 a_t^{-1}(d)t) \leq \sigma^2$. This and (A.31) together imply (A.29) for $i = 2$. Equation (A.29) for $i = 3$ follows directly from (A.25) and (A.27). The details are skipped. To show that (A.29) holds for $i = 4$, we get from Taylor's theorem and (A.28) that for all $\boldsymbol{\eta} \in A_4$ and sufficiently large t,

$$\begin{aligned} \mathrm{E}(\varepsilon_t^2(\boldsymbol{\eta}) a_t^{-1}(d)t) = \mathrm{E}(\varepsilon_t^2(\boldsymbol{\eta})) &\geq \sigma^2 C_0 C_{\eta^*}^2 \int_1^{\iota_1^{-1}} x^{2(d_0-d)-2}\, dx \\ &\geq \sigma^2 C_0 C_{\eta^*}^2 (\log \iota_1^{-1}) \iota_1^{2\iota_1} > 2\sigma^2. \end{aligned}$$

Hence, the desired conclusion follows. For $\boldsymbol{\eta} \in A_5$, define $W_t(\boldsymbol{\eta}) = \mathrm{E}(\varepsilon_t(\boldsymbol{\eta}) - \varepsilon_t)^2$. Then, it follows from (A.24), (A.26) and the Weierstrass convergence theorem that $W_t(\boldsymbol{\eta})$ converges uniformly on A_5 to some function $W_\infty(\boldsymbol{\eta})$. Moreover, since $W_t(\boldsymbol{\eta})$ is continuous on A_5, the uniform convergence of $W_t(\boldsymbol{\eta})$ ensures that $W_\infty(\boldsymbol{\eta})$ is also continuous. In view of (1.4)–(1.6) and (A.2), we have $(1 - z)^{d-d_0} A_{2,\boldsymbol{\theta}_0}(z) A_{1,\boldsymbol{\theta}_0}^{-1}(z) A_{1,\boldsymbol{\theta}}(z) A_{2,\boldsymbol{\theta}}^{-1}(z) \neq 1$. Hence, for each $\boldsymbol{\eta} \in A_5$, $W_t(\boldsymbol{\eta}) > 0$ for all sufficiently large t. This, the continuity of $W_\infty(\boldsymbol{\eta})$ and the compactness of A_5 yield that there exists $\tilde{\theta} > 0$ such that $\inf_{\boldsymbol{\eta} \in A_5} W_\infty(\boldsymbol{\eta}) > \tilde{\theta}$. By making use of this finding and the uniform convergence of $W_t(\boldsymbol{\eta})$ to $W_\infty(\boldsymbol{\eta})$, we obtain (A.29) with $i=5$. This complete the proof of Lemma 2.1. □

PROOF OF LEMMA 2.2. Following the proof of Lemma 2.1, write $\varepsilon_t(\boldsymbol{\eta})-\varepsilon_t=\sum_{s=1}^{t-1}b_s(\boldsymbol{\eta})\varepsilon_{t-s}$. Note first that $b_s(\boldsymbol{\eta})$ has continuous partial derivatives, $D_{\mathbf{j}}b_s(\boldsymbol{\eta})$, on $\Pi\times D$. By an argument similar to that used to derive bounds for $\sup_{\boldsymbol{\eta}\in Q_1^{(i)}}|D_{\mathbf{j}}c_{s,k}^{(n)}(\boldsymbol{\eta})|$ in Lemma 2.1, we have for all $s\geq 1$,

$$\sup_{\boldsymbol{\eta}\in B_{j,v}}\left|D_{\mathbf{j}}b_s(\boldsymbol{\eta})\right|\leq\begin{cases}Cs^{-1/2}, & \text{if } j=0 \text{ and } \bar{p}\notin\mathbf{j},\\ Cs^{-1/2}\log(s+1), & \text{if } j=0 \text{ and } \bar{p}\in\mathbf{j},\\ Cs^{vj-1/2}, & \text{if } j\geq 1 \text{ and } \bar{p}\notin\mathbf{j},\\ Cs^{vj-1/2}\log(s+1), & \text{if } j\geq 1 \text{ and } \bar{p}\in\mathbf{j}.\end{cases}\tag{A.32}$$

Moreover, it follows from (A.24) and (A.26) that

$$\sup_{\boldsymbol{\eta}\in B_{j,v}}\sum_{s=0}^{n-1}b_s^2(\boldsymbol{\eta})=\begin{cases}O(\log n), & \text{if } j=0,\\ O(n^{2vj}), & \text{if } j\geq 1.\end{cases}\tag{A.33}$$

In view of (A.32), (A.33), the compactness of $B_{j,v}$, $j\geq 0$ and (B.5) of Chan and Ing (2011), we get

$$\begin{aligned}\mathrm{E}&\left\{\sup_{\boldsymbol{\eta}\in B_{j,v}}\left|\sum_{t=1}^{n}(\varepsilon_t(\boldsymbol{\eta})-\varepsilon_t(\boldsymbol{\eta}_0))\varepsilon_t\right|^{q_1}\right\}\\ &\leq Cn^{q_1/2}\left[\left\{\sup_{\boldsymbol{\eta}\in B_{j,v}}\sum_{s=1}^{n-1}b_s^2(\boldsymbol{\eta})\right\}^{q_1/2}\right.\\ &\qquad\left.+\left\{\sum_{s=1}^{n-1}\max_{\mathbf{j}\in J(m,\bar{p}),1\leq m\leq\bar{p}}\sup_{\boldsymbol{\eta}\in B_{j,v}}\left(D_{\mathbf{j}}b_s(\boldsymbol{\eta})\right)^2\right\}^{q_1/2}\right]\\ &=\begin{cases}O\left(n(\log n)^3\right)^{q_1/2}, & \text{if } j=0,\\ O\left(n^{1+2vj}(\log n)^2\right)^{q_1/2}, & \text{if } j\geq 1.\end{cases}\end{aligned}\tag{A.34}$$

By Hölder's inequality, the left-hand side of (2.3) is bounded above by

$$\begin{aligned}&\left\{\mathrm{E}\left(\inf_{\boldsymbol{\eta}\in B_{j,v}}a_n^{-1}(d)\sum_{t=1}^{n}\left(\varepsilon_t(\boldsymbol{\eta})-\varepsilon_t(\boldsymbol{\eta}_0)\right)^2\right)^{-qq_1/(q_1-q)}\right\}^{(q_1-q)/q_1}\\ &\quad\times\left\{\mathrm{E}\left(\sup_{\boldsymbol{\eta}\in B_{j,v}}\left|\sum_{t=1}^{n}(\varepsilon_t(\boldsymbol{\eta})-\varepsilon_t(\boldsymbol{\eta}_0))\varepsilon_t\right|^{q_1}\right)\right\}^{q/q_1}\\ &\quad\times\left(n^{1+2(j-1)v}I_{\{j\geq 1\}}+nI_{\{j=0\}}\right)^{-q},\end{aligned}$$

which together with (A.34) and Lemma 2.1, gives the desired conclusion. □

PROOF OF THEOREM 3.2. By a calculation similar but more complicated than that in the proof of Theorem 3.1, we obtain

$$\begin{aligned} & y_{n+h} - \hat{y}_{n+h}(\hat{\boldsymbol{\eta}}_n) - \sum_{s=0}^{h-1} \underline{c}_s(\boldsymbol{\eta}_0)\varepsilon_{n+h-s} \\ & \quad = (\nabla^{(1)}\varepsilon_{n+1}(\boldsymbol{\theta}_0))^\top \underline{L}_h(\boldsymbol{\eta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \underline{\mathbf{c}}_h^\top(\boldsymbol{\eta}_0)\mathbf{w}_{n+1,h}(\hat{d}_n - d_0) + r_n, \end{aligned} \tag{A.35}$$

where r_n satisfies $n\mathrm{E}(r_n^2) = o(1)$, and $((\nabla^{(1)}\varepsilon_{n+1}(\boldsymbol{\theta}_0))^\top \underline{L}_h(\boldsymbol{\eta}_0), \underline{\mathbf{c}}_h^\top(\boldsymbol{\eta}_0)\mathbf{w}_{n+1,h})^\top$ and $n^{1/2}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)$ are asymptotically independent. The desired conclusion (3.23) follows by a direct application of (A.35), (3.12) and Theorem 2.1. □

Acknowledgments. We would like to thank an Associate Editor and two anonymous referees for their insightful and constructive comments, which greatly improve the presentation of this paper.

SUPPLEMENTARY MATERIAL

Supplement to "Moment bounds and mean squared prediction errors of long-memory time series" (DOI: 10.1214/13-AOS1110SUPP; .pdf). The supplementary material contains a Monte Carlo experiment of finite sample performance of the CSS predictor and the proof of (2.9).

REFERENCES

BERAN, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57** 659–672. MR1354073

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

BLOOMFIELD, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika* **60** 217–226. MR0323048

CHAN, N. H., HUANG, S. F. and ING, C. K. (2013). Supplement to "Moment bounds and mean squared prediction errors of long-memory time series." DOI:10.1214/13-AOS1110SUPP.

CHAN, N. H. and ING, C.-K. (2011). Uniform moment bounds of Fisher's information with applications to time series. *Ann. Statist.* **39** 1526–1550. MR2850211

CHAN, N. H. and WEI, C. Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Ann. Statist.* **16** 367–401. MR0924877

DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17** 1749–1766. MR1026311

DOUKHAN, P., OPPENHEIM, G. and TAQQU, M. S., eds. (2003). *Theory and Applications of Long-Range Dependence*. Birkhäuser, Boston, MA. MR1956041

FINDLEY, D. F. and WEI, C.-Z. (2002). AIC, overfitting principles, and the boundedness of moments of inverse matrices for vector autoregressions and related models. *J. Multivariate Anal.* **83** 415–450. MR1945962

FOX, R. and TAQQU, M. S. (1986). Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* **14** 517–532. MR0840512

FULLER, W. A. and HASZA, D. P. (1981). Properties of predictors for autoregressive time series. *J. Amer. Statist. Assoc.* **76** 155–161. MR0608187

GIRAITIS, L. and SURGAILIS, D. (1990). A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields* **86** 87–104. MR1061950

HUALDE, J. and ROBINSON, P. M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Ann. Statist.* **39** 3152–3181. MR3012404

ING, C.-K. (2003). Multistep prediction in autoregressive processes. *Econometric Theory* **19** 254–279. MR1966030

ING, C.-K., LIN, J.-L. and YU, S.-H. (2009). Toward optimal multistep forecasts in non-stationary autoregressions. *Bernoulli* **15** 402–437. MR2543868

ING, C.-K. and SIN, C.-Y. (2006). On prediction errors in regression models with nonstationary regressors. In *Time Series and Related Topics* (H.-C. Ho, C.-K. Ing and T. L. Lai, eds.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **52** 60–71. IMS, Beachwood, OH. MR2427839

ING, C.-K., SIN, C.-Y. and YU, S.-H. (2010). Prediction errors in nonstationary autoregressions of infinite order. *Econometric Theory* **26** 774–803. MR2646479

ING, C.-K. and WEI, C.-Z. (2003). On same-realization prediction in an infinite-order autoregressive process. *J. Multivariate Anal.* **85** 130–155. MR1978181

KATAYAMA, N. (2008). Asymptotic prediction of mean squared error for long-memory processes with estimated parameters. *J. Forecast.* **27** 690–720. MR2654619

KUNITOMO, N. and YAMAMOTO, T. (1985). Properties of predictors in misspecified autoregressive time series models. *J. Amer. Statist. Assoc.* **80** 941–950. MR0819598

LAI, T. L. (1994). Asymptotic properties of nonlinear least squares estimates in stochastic regression models. *Ann. Statist.* **22** 1917–1930. MR1329175

LING, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *Ann. Statist.* **35** 1213–1237. MR2341704

ROBINSON, P. M. (2006). Conditional-sum-of-squares estimation of models for stationary time series with long memory. In *Time Series and Related Topics* (H.-C. Ho, C.-K. Ing and T. L. Lai, eds.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **52** 130–137. IMS, Beachwood, OH. MR2427843

ROBINSON, P. M. and HIDALGO, F. J. (1997). Time series regression with long-range dependence. *Ann. Statist.* **25** 77–104. MR1429918

SCHORFHEIDE, F. (2005). VAR forecasting under misspecification. *J. Econometrics* **128** 99–136. MR2022928

WEI, C. Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist.* **15** 1667–1682. MR0913581

YAMAMOTO, T. (1981). Predictions of multivariate autoregressive-moving average models. *Biometrika* **68** 485–492. MR0626411

N. H. CHAN
DEPARTMENT OF STATISTICS
ROOM 118, LADY SHAW BUILDING
CHINESE UNIVERSITY OF HONG KONG
SHATIN, NEW TERRITORIES
HONG KONG
E-MAIL: nhchan@sta.cuhk.edu.hk

S.-F. HUANG
DEPARTMENT OF APPLIED MATHEMATICS
NATIONAL UNIVERSITY OF KAOHSIUNG
KAOHSIUNG, 811
TAIWAN, ROC
E-MAIL: huangsf@nuk.edu.tw

C.-K. ING
INSTITUTE OF STATISTICAL SCIENCE
ACADEMIA SINICA
TAIPEI, 115
TAIWAN, ROC
E-MAIL: cking@stat.sinica.edu.tw

Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

SUPPLEMENT TO "MOMENT BOUNDS AND MEAN SQUARED PREDICTION ERRORS OF FRACTIONAL TIME SERIES MODELS"

By Ngai Hang Chan, Shih-Feng Huang and Ching-Kang Ing

Chinese University of Hong Kong, National University of Kaohsiung, and Academia Sinica and National Taiwan University

This supplement contains a Monte Carlo experiment of finite sample performance of the CSS predictor and the proof of (2.9).

NUMERICAL EXAMPLES

To illustrate the finite sample properties of the one-step and multi-step prediction results obtained in Section 3, we conduct a Monte Carlo simulation to assess the performance of the empirical estimates of $n\{\mathrm{E}(y_{n+h} - \hat{y}_{n+h}(\hat{\boldsymbol{\eta}}_n))^2 - \sigma_h^2(\boldsymbol{\eta}_0)\}$ under $\mathrm{I}(d_0)$ and ARFIMA$(1, d_0, 1)$ models with ε_t's being i.i.d. standard normal random variables, $-0.6 \le d_0 \le 2.0$ and $(\alpha_{0,1}, \beta_{0,1})$ = $(0.5, 0.2)$ and (0.3,0.8). These estimates, denoted by $g_{h,n}$ and $m_{h,n}$ for $\mathrm{I}(d_0)$ and ARFIMA$(1, d_0, 1)$ models, respectively, are obtained based on 5,000 replications for n =100, 500, 2,000 and $h = 1, \ldots, 10$. The closeness of $g_{h,n}$ and $m_{h,n}$ to the corresponding limiting values $\bar{g}_h(d_0)$ and $m_h(\boldsymbol{\eta_0}) = \{\underline{f}_h(1,1) + \underline{g}_h(\boldsymbol{\eta_0}) + 2J_h(\boldsymbol{\eta_0})\}$, defined in (3.24) and (3.23), is measured by the ratios $R^{(1)}_{h,n} = g_{h,n}/\bar{g}_h(d_0)$ and $R^{(2)}_{h,n} = m_{h,n}/m_h(\boldsymbol{\eta_0})$, which is summarized in Tables 1–3.

Table 1, listing $\bar{g}_h(d_0)$ and $R^{(1)}_{h,n}$ with n =100 and 500, shows that except for a few cases where $n = 100$, $d_0 = -0.6$ and $h = 2, 3$ and 4, all values of $R^{(1)}_{h,n}$ are between 0.88 and 1.15. This result suggests that $g_{h,n}$ can be well predicted by $\bar{g}_h(d_0)$ even for moderate sample sizes. Note also that in the exceptional cases mentioned above, $R^{(1)}_{h,n}$'s fall within the interval $[1.2, 1.4]$. Tables 2 and 3, listing $m_h(\boldsymbol{\eta_0})$ and $R^{(2)}_{h,n}$ with n =500 and 2,000, show that the behaviours of $R^{(2)}_{h,n}$ vary not only with n, h and d_0, but also with $\alpha_{0,1}$ and $\beta_{0,1}$, and hence are somewhat different from those of $R^{(1)}_{h,n}$. In particular, it is shown in Table 2 ($(\alpha_{0,1}, \beta_{0,1}) = (0.5, 0.2)$) and Table 3 ($(\alpha_{0,1}, \beta_{0,1}) = (0.3, 0.8)$) that when $m_h(\boldsymbol{\eta_0}) > 0.06$, $R^{(2)}_{h,2000}$'s fall between 0.9 and 1.1 ; and when $m_h(\boldsymbol{\eta_0}) \le 0.06$, the values are between 1.12 and 2.11 except for a small number of exceptional cases in Table 3 in which $1.04 \leq R^{(2)}_{h,2000} \leq 1.1$. This feature seems to reflect the desirable property that as long as the asymptotic MSPEs of the h-step CSS predictors are large enough (e.g., > 0.06), they can be readily estimated by the corresponding finite sample counterparts. It is also worth noting that the cases where $m_h(\boldsymbol{\eta_0}) \leq 0.06$ in Table 2, only including the pairs $(d_0, h) = (-0.6, 8), (-0.6, 9), (-0.6, 10), (-0.5, 9)$ and (-0.5, 10), are substantially rarer than those in Table 3 which contain 21 such pairs in the upper right corner.

The behaviours of $R^{(2)}_{h,500}$ are not explicable by a simple rule related to the value of $m_h(\boldsymbol{\eta_0})$, unlike those of $R^{(2)}_{h,2000}$. When $(\alpha_{0,1}, \beta_{0,1}) = (0.5, 0.2)$ and $d_0 \geq -0.4$, all $R^{(2)}_{h,500}$'s lie between 0.9 and 1.1, suggesting the similarities between $m_h(\boldsymbol{\eta_0})$ and $m_{h,500}$ in these cases. Alternatively, when $(\alpha_{0,1}, \beta_{0,1}) = (0.5, 0.2)$ and $d_0 \leq -0.5$, $|R^{(2)}_{h,500} - 1|$ is usually slightly larger than $|R^{(2)}_{h,2000} - 1|$ and varies between 0.83 and 1.5. On the other hand, in the case of $(\alpha_{0,1}, \beta_{0,1}) = (0.3, 0.8)$, $R^{(2)}_{h,500}$ is significantly smaller than 1 if $d_0 \geq 0$, and oscillates between 0.75 and 1.46 if $d_0 \leq -0.4$.

To conclude, our numerical findings are consistent with the asymptotic results established in Theorems 3.1 and 3.2 even when the prediction lead time h is large. For I(d_0) models, $n = 500$ is large enough for our asymptotic results to become effective. However, to attain a similar precision for ARFIMA(1, d_0, 1) models, $n = 2,000$ seems indispensable.

PROOF OF (2.9)

In view of Theorem 2.1 of Chan and Ing (2011), it suffices for (2.9) to show that conditions (C1)-(C4) of the same paper hold when $\mathbf{f}_t(\cdot)$ and $\boldsymbol{\Theta}$ therein equal $\nabla\varepsilon_t(\boldsymbol{\eta})$ and $B_{\tau_1}(\boldsymbol{\eta}_0)$, respectively. It is clear that $\nabla\varepsilon_t(\boldsymbol{\eta})$ is continuous on $B_{\tau_1}(\boldsymbol{\eta}_0)$, and hence (C1) of Chan and Ing (2011) follows. To verify (C2) of Chan and Ing (2011), define $\boldsymbol{\Lambda} = \{\mathbf{v} : \mathbf{v} \in R^{\bar{p}}, ||\mathbf{v}|| = 1\}$. We will first prove that for any $\boldsymbol{\eta} \in \bar{B}_{\tau_1}(\boldsymbol{\eta}_0)$ and $\mathbf{v} = (\mu_1, \ldots, \mu_{p_1}, s_1, \ldots, s_{p_2}, \mu_0)^\mathsf{T} \in \boldsymbol{\Lambda}$, there exists some $c_{\boldsymbol{\eta},\mathbf{v}} > 0$ such that

$$\lim_{t\to\infty} \mathrm{E}(\mathbf{v}^\mathsf{T}\nabla\varepsilon_t(\boldsymbol{\eta}))^2 := H(\mathbf{v}, \boldsymbol{\eta}) > c_{\boldsymbol{\eta},\mathbf{v}}. \tag{0.1}$$

If $\mu_0 = 0$, then (0.1) follows from (1.4)-(1.6) and an argument similar to that used in proving (3.20) of Chan and Ing (2011). On the other hand, if $\mu_0 \neq 0$, straightforward calculations yield

$$\mathbf{v}^\mathsf{T}\nabla\varepsilon_t(\boldsymbol{\eta}) = F_{\boldsymbol{\eta},\mathbf{v}}(B)W_{\boldsymbol{\eta}}(B)\varepsilon_t, \tag{0.2}$$

where $F_{\boldsymbol{\eta},\mathbf{v}}(z) = -\sum_{i=1}^{p_1} \mu_i z^i + \mu_0 A_{1,\boldsymbol{\theta}}(z)\log(1-z) + A_{1,\boldsymbol{\theta}}(z)A_{2,\boldsymbol{\theta}}^{-1}(z)\sum_{j=1}^{p_2} s_j z^j$ and $W_{\boldsymbol{\eta}}(z) = A_{2,\boldsymbol{\theta}_0}(z)A_{2,\boldsymbol{\theta}}^{-1}(z)A_{1,\boldsymbol{\theta}_0}^{-1}(z)(1-z)^{d-d_0}$. By virtue of $A_{1,\boldsymbol{\theta}}(z) \neq 0$ for $|z| \leq 1$, it can be shown that $A_{1,\boldsymbol{\theta}}(z)\log(1-z) = \sum_{j=1}^{\infty} l_j z^j$ with $|l_j| \geq \xi/j$ for some $\xi > 0$ and all large j. This together with $A_{2,\boldsymbol{\theta}}(z) \neq 0$ for $|z| \leq 1$ yields $F_{\boldsymbol{\eta},\mathbf{v}}(z) \neq 0$. In addition, it is clear that $W_{\boldsymbol{\eta}}(z) \neq 0$. Combining $F_{\boldsymbol{\eta},\mathbf{v}}(z)W_{\boldsymbol{\eta}}(z) \neq 0$ with (2.22) and (0.2) gives (0.1). Now, by (0.1), (2.22), the continuity of $\mathrm{E}(\mathbf{v}^{\top}\nabla\varepsilon_t(\boldsymbol{\eta}))^2$ on $\boldsymbol{\Lambda} \times \bar{B}_{\tau_1}(\boldsymbol{\eta}_0)$ (which is compact) and the Weierstrass convergence theorem, we obtain

$$\mathrm{E}(\mathbf{v}^{\top}\nabla\varepsilon_t(\boldsymbol{\eta}))^2 \text{ converges to } H(\mathbf{v},\boldsymbol{\eta}) \text{ uniformly on } \boldsymbol{\Lambda} \times \bar{B}_{\tau_1}(\boldsymbol{\eta}_0), \tag{0.3}$$

and

$$\inf_{\mathbf{v}\in\boldsymbol{\Lambda},\boldsymbol{\eta}\in\bar{B}_{\tau_1}(\boldsymbol{\eta}_0)} H(\mathbf{v},\boldsymbol{\eta}) > 0. \tag{0.4}$$

Consequently, (C2) of Chan and Ing (2011) follows from (0.3), (0.4), (A1) and an argument similar to that used in proving (3.25) of Chan and Ing (2011).

To show (C3) of Chan and Ing (2011), note first that by the mean value theorem for vector-valued functions, we have for any $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2 \in B_{\tau_1}(\boldsymbol{\eta}_0)$,

$$\begin{aligned} &||\nabla\varepsilon_t(\boldsymbol{\eta}_2) - \nabla\varepsilon_t(\boldsymbol{\eta}_1)||^2 \\ &\leq ||\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1||^2 || \int_0^1 \nabla^2\varepsilon_t(\boldsymbol{\eta}_1 + v(\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1))\, dv||^2 \leq ||\boldsymbol{\eta}_2 - \boldsymbol{\eta}_1||^2 \tilde{B}_t^2, \end{aligned} \tag{0.5}$$

where $\tilde{B}_t^2 = \sum_{1\leq i,j\leq r} \sup_{\boldsymbol{\theta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} (\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j})^2$. Our aim is to show that

$$\sup_{t\geq 2} \mathrm{E}(\tilde{B}_t^2) < \infty. \tag{0.6}$$

In view of the first inequality in (2.20), it suffices for (0.6) to show that for any $1 \leq i,j \leq r$ and $t \geq 2$, there exists $0 < \bar{U} < \infty$, independent of i, j and t, such that

$$\mathrm{E}\{\sup_{\boldsymbol{\theta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} [\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j} - \nabla^2\varepsilon_t(\boldsymbol{\eta}_0)_{i,j}]^2\} < \bar{U}. \tag{0.7}$$

Using an argument similar to that used in proving (A.12), we have for any $1 \leq i,j \leq r$ and $t \geq 2$,

$$\begin{aligned} &\mathrm{E}\Big\{\sup_{\boldsymbol{\theta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} [\nabla^2\varepsilon_t(\boldsymbol{\eta})_{i,j} - \nabla^2\varepsilon_t(\boldsymbol{\eta}_0)_{i,j}]^2\Big\} \\ \leq &C\Big\{\sum_{s=1}^{\infty} \max_{\mathbf{j}\in\mathbf{J}(m,\bar{p}),1\leq m\leq\bar{p}} \sup_{\boldsymbol{\eta}\in B_{\tau_1}(\boldsymbol{\eta}_0)} (\mathbf{D}_{\mathbf{j}} c_{s,ij}(\boldsymbol{\eta}))^2\Big\}. \end{aligned} \tag{0.8}$$

Combining (0.8), with (2.19), yields (0.7), and hence (C3) of Chan and Ing (2011) is proved. Finally, (C4) of Chan and Ing (2011) follows immediately from (2.22), (0.5) and (0.6).

References.

[1] Chan, N. H., Huang, S. F. and Ing, C.-K. (2011). Moment bounds and mean squared prediction errors of long-memory Time Series.

Ngai Hang, Chan
Department of Statistics
Room 118, Lady Shaw Building
Chinese University of Hong Kong
Shatin, New Territories
Hong Kong
E-mail: nhchan@sta.cuhk.edu.hk

Shih-Feng Huang
Department of Applied Mathematics
National University of Kaohsiung
Kaohsiung, 811
Taiwan, ROC
E-mail: huangsf@nuk.edu.tw

Ching-Kang Ing
Institute of Statistical Science
Academia Sinica
Taipei, 115
Taiwan, ROC
E-mail: cking@stat.sinica.edu.tw

Table 1
Values of $R^{(1)}_{h,n}$ and $\bar{g}_h(d_0)$, with $-0.6 \leq d_0 \leq 2.0$, $n = 100, 500$, and $h = 1, \ldots, 10$.

	h									
	1	2	3	4	5	6	7	8	9	10
$d_0 = -0.6$										
$\bar{g}_h(d_0)$	1.0000	0.0226	0.0090	0.0048	0.0029	0.0019	0.0014	0.0010	0.0008	0.0006
$R^{(1)}_{h,100}$	1.09	1.39	1.27	1.20	1.15	1.12	1.09	1.07	1.05	1.03
$R^{(1)}_{h,500}$	1.06	1.11	1.14	1.15	1.14	1.13	1.12	1.11	1.10	1.09
$d_0 = -0.5$										
$\bar{g}_h(d_0)$	1.0000	0.0341	0.0118	0.0062	0.0038	0.0026	0.0019	0.0015	0.0012	0.0009
$R^{(1)}_{h,100}$	1.04	0.97	0.97	0.97	0.97	0.98	0.97	0.97	0.97	0.97
$R^{(1)}_{h,500}$	0.99	1.02	1.03	1.01	1.00	0.99	0.98	0.98	0.97	0.96
$d_0 = -0.4$										
$\bar{g}_h(d_0)$	1.0000	0.0657	0.0230	0.0121	0.0076	0.0053	0.0039	0.0030	0.0024	0.0020
$R^{(1)}_{h,100}$	1.04	0.92	0.96	0.96	0.96	0.95	0.94	0.93	0.91	0.90
$R^{(1)}_{h,500}$	1.03	0.98	0.99	0.99	0.99	1.00	0.99	0.99	0.99	0.99
$d_0 = 0.0$										
$\bar{g}_h(d_0)$	1.00	0.39	0.24	0.17	0.13	0.11	0.09	0.08	0.07	0.06
$R^{(1)}_{h,100}$	1.05	0.95	0.93	0.92	0.91	0.90	0.90	0.89	0.89	0.88
$R^{(1)}_{h,500}$	1.04	1.02	1.01	1.01	1.02	1.02	1.02	1.02	1.02	1.02
$d_0 = 0.4$										
$\bar{g}_h(d_0)$	1.00	1.04	1.02	0.99	0.96	0.94	0.92	0.90	0.89	0.87
$R^{(1)}_{h,100}$	1.09	1.04	1.01	1.00	0.99	0.98	0.97	0.97	0.96	0.96
$R^{(1)}_{h,500}$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$d_0 = 0.5$										
$\bar{g}_h(d_0)$	1.00	1.25	1.35	1.41	1.45	1.47	1.49	1.50	1.52	1.52
$R^{(1)}_{h,100}$	1.05	1.03	1.02	1.02	1.01	1.01	1.01	1.00	1.00	1.00
$R^{(1)}_{h,500}$	0.97	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.98	0.98
$d_0 = 0.6$										
$\bar{g}_h(d_0)$	1.00	1.48	1.76	1.96	2.12	2.24	2.35	2.44	2.52	2.60
$R^{(1)}_{h,100}$	1.09	1.05	1.03	1.03	1.03	1.03	1.02	1.02	1.02	1.02
$R^{(1)}_{h,500}$	1.06	1.04	1.03	1.02	1.01	1.01	1.00	1.00	1.00	1.00
$d_0 = 1.0$										
$\bar{g}_h(d_0)$	1.00	2.61	4.37	6.20	8.06	9.95	11.86	13.78	15.70	17.64
$R^{(1)}_{h,100}$	1.12	1.07	1.05	1.03	1.02	1.01	1.01	1.00	1.00	0.99
$R^{(1)}_{h,500}$	1.03	1.03	1.02	1.02	1.02	1.01	1.01	1.01	1.01	1.01
$d_0 = 1.5$										
$\bar{g}_h(d_0)$	1.00	4.47	10.68	19.75	31.73	46.66	64.56	85.45	109.35	136.25
$R^{(1)}_{h,100}$	1.14	1.11	1.09	1.08	1.08	1.07	1.07	1.06	1.06	1.06
$R^{(1)}_{h,500}$	1.03	1.02	1.01	1.01	1.00	1.00	1.00	1.00	0.99	0.99
$d_0 = 2.0$										
$\bar{g}_h(d_0)$	1.00	6.82	22.06	51.50	100.06	172.71	274.47	410.36	585.47	804.87
$R^{(1)}_{h,100}$	1.08	1.06	1.05	1.04	1.03	1.03	1.02	1.02	1.02	1.01
$R^{(1)}_{h,500}$	1.00	1.00	0.99	0.99	0.99	0.99	0.99	0.98	0.98	0.98

Table 2

Values of $R^{(2)}_{h,n}$ *and* $m_h(\boldsymbol{\eta}_0)$, *with* $(\alpha_{0,1},\beta_{0,1})=(0.5,0.2)$, $-0.6\le d_0\le 2.0$, $n=500,2000$, *and* $h=1,\dots,10$.

	h									
	1	2	3	4	5	6	7	8	9	10
$d_0=-0.6$										
$m_h(\boldsymbol{\eta}_0)$	3.00	1.31	0.48	0.29	0.21	0.14	0.08	0.05	0.03	0.02
$R^{(2)}_{h,500}$	1.00	1.17	1.15	1.02	0.89	0.87	0.93	1.10	1.25	1.50
$R^{(2)}_{h,2000}$	1.04	1.09	1.09	1.08	1.03	1.02	1.06	1.13	1.20	1.25
$d_0=-0.5$										
$m_h(\boldsymbol{\eta}_0)$	3.00	1.48	0.67	0.41	0.30	0.21	0.14	0.09	0.06	0.04
$R^{(2)}_{h,500}$	1.03	1.08	1.03	0.97	0.86	0.83	0.85	0.95	1.02	1.14
$R^{(2)}_{h,2000}$	1.03	1.05	1.04	1.05	1.03	1.03	1.05	1.09	1.12	1.13
$d_0=-0.4$										
$m_h(\boldsymbol{\eta}_0)$	3.00	1.71	0.92	0.59	0.45	0.34	0.25	0.17	0.12	0.09
$R^{(2)}_{h,500}$	1.00	1.07	1.00	1.00	0.94	0.90	0.91	0.94	0.97	1.03
$R^{(2)}_{h,2000}$	1.04	1.08	1.07	1.07	1.05	1.03	1.04	1.06	1.08	1.10
$d_0=0.0$										
$m_h(\boldsymbol{\eta}_0)$	3.00	3.25	3.04	2.69	2.43	2.23	2.04	1.85	1.67	1.52
$R^{(2)}_{h,500}$	1.00	0.98	0.94	0.97	0.97	0.96	0.95	0.94	0.94	0.94
$R^{(2)}_{h,2000}$	1.04	1.07	1.06	1.06	1.06	1.05	1.04	1.04	1.04	1.04
$d_0=0.4$										
$m_h(\boldsymbol{\eta}_0)$	3.00	5.74	7.95	9.41	10.51	11.42	12.17	12.77	13.23	13.58
$R^{(2)}_{h,500}$	0.99	0.95	0.92	0.92	0.92	0.92	0.91	0.91	0.90	0.90
$R^{(2)}_{h,2000}$	1.04	1.05	1.05	1.06	1.07	1.07	1.07	1.07	1.07	1.07
$d_0=0.5$										
$m_h(\boldsymbol{\eta}_0)$	3.00	6.52	9.81	12.40	14.56	16.46	18.17	19.66	20.96	22.07
$R^{(2)}_{h,500}$	1.03	0.97	0.93	0.94	0.94	0.93	0.92	0.92	0.91	0.90
$R^{(2)}_{h,2000}$	1.03	1.02	1.02	1.02	1.03	1.03	1.03	1.03	1.03	1.02
$d_0=0.6$										
$m_h(\boldsymbol{\eta}_0)$	3.00	7.35	11.99	16.13	19.88	23.39	26.69	29.76	32.60	35.21
$R^{(2)}_{h,500}$	1.00	0.95	0.94	0.95	0.96	0.97	0.97	0.96	0.96	0.96
$R^{(2)}_{h,2000}$	1.04	1.05	1.05	1.05	1.05	1.05	1.05	1.04	1.04	1.04
$d_0=1.0$										
$m_h(\boldsymbol{\eta}_0)$	3.00	11.28	24.55	41.51	61.37	83.82	108.62	135.51	164.21	194.44
$R^{(2)}_{h,500}$	1.00	0.95	0.95	0.95	0.96	0.96	0.97	0.97	0.96	0.96
$R^{(2)}_{h,2000}$	1.04	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.04	1.04
$d_0=1.5$										
$m_h(\boldsymbol{\eta}_0)$	3.00	17.55	51.96	112.42	203.66	329.83	494.85	702.20	954.93	1255.57
$R^{(2)}_{h,500}$	1.00	0.96	0.95	0.95	0.96	0.96	0.96	0.96	0.96	0.96
$R^{(2)}_{h,2000}$	1.04	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05	1.05
$d_0=2.0$										
$m_h(\boldsymbol{\eta}_0)$	3.00	25.32	97.85	262.90	571.13	1080.50	1855.85	2968.30	4494.44	6515.39
$R^{(2)}_{h,500}$	1.05	1.02	1.01	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$R^{(2)}_{h,2000}$	1.03	1.02	1.02	1.03	1.03	1.03	1.04	1.04	1.04	1.04

TABLE 3
Values of $R^{(2)}_{h,n}$ and $m_h(\boldsymbol{\eta_0})$, with $(\alpha_{0,1},\beta_{0,1})=(0.3,0.8)$, $-0.6\le d_0\le 2.0$, $n=500,2000$, and $h=1,\dots,10$.

	h									
	1	2	3	4	5	6	7	8	9	10
$d_0=-0.6$										
$m_h(\boldsymbol{\eta_0})$	3.0000	1.1229	0.1056	0.0350	0.0192	0.0094	0.0046	0.0023	0.0012	0.0006
$R^{(2)}_{h,500}$	0.87	1.04	1.12	1.24	1.01	0.95	0.96	1.00	1.13	1.46
$R^{(2)}_{h,2000}$	1.04	1.05	1.10	1.22	1.11	1.10	1.17	1.31	1.57	2.11
$d_0=-0.5$										
$m_h(\boldsymbol{\eta_0})$	3.0000	0.9076	0.1039	0.0299	0.0190	0.0113	0.0067	0.0041	0.0026	0.0016
$R^{(2)}_{h,500}$	0.86	1.23	1.01	1.31	1.02	0.88	0.83	0.83	0.87	0.97
$R^{(2)}_{h,2000}$	0.99	1.03	1.04	1.18	1.07	1.04	1.06	1.12	1.23	1.41
$d_0=-0.4$										
$m_h(\boldsymbol{\eta_0})$	3.0000	0.7523	0.1078	0.0271	0.0192	0.0140	0.0100	0.0072	0.0053	0.0039
$R^{(2)}_{h,500}$	0.87	1.05	1.02	1.39	1.08	0.89	0.81	0.76	0.75	0.77
$R^{(2)}_{h,2000}$	1.04	1.07	1.05	1.31	1.19	1.12	1.11	1.14	1.20	1.29
$d_0=0.0$										
$m_h(\boldsymbol{\eta_0})$	3.00	0.73	0.29	0.13	0.09	0.09	0.08	0.08	0.08	0.07
$R^{(2)}_{h,500}$	0.87	0.92	0.82	0.82	0.75	0.67	0.62	0.59	0.57	0.56
$R^{(2)}_{h,2000}$	1.04	1.09	1.04	1.10	1.09	1.07	1.04	1.03	1.02	1.02
$d_0=0.4$										
$m_h(\boldsymbol{\eta_0})$	3.00	1.67	1.24	0.98	0.85	0.81	0.81	0.82	0.84	0.86
$R^{(2)}_{h,500}$	0.86	0.82	0.77	0.72	0.68	0.65	0.62	0.60	0.59	0.57
$R^{(2)}_{h,2000}$	0.99	1.00	1.01	1.03	1.04	1.04	1.03	1.02	1.01	1.00
$d_0=0.5$										
$m_h(\boldsymbol{\eta_0})$	3.00	2.05	1.71	1.47	1.35	1.32	1.34	1.38	1.43	1.49
$R^{(2)}_{h,500}$	0.86	0.81	0.77	0.73	0.69	0.67	0.64	0.62	0.60	0.59
$R^{(2)}_{h,2000}$	0.99	0.99	1.00	1.02	1.03	1.03	1.02	1.01	1.00	0.99
$d_0=0.6$										
$m_h(\boldsymbol{\eta_0})$	3.00	2.50	2.30	2.05	2.08	2.10	2.17	2.27	2.40	2.53
$R^{(2)}_{h,500}$	0.87	0.82	0.79	0.76	0.73	0.70	0.68	0.66	0.64	0.62
$R^{(2)}_{h,2000}$	1.04	1.06	1.06	1.06	1.06	1.05	1.04	1.03	1.02	1.01
$d_0=1.0$										
$m_h(\boldsymbol{\eta_0})$	3.00	4.88	6.46	7.88	9.26	10.69	12.24	13.93	15.74	17.70
$R^{(2)}_{h,500}$	0.87	0.83	0.80	0.78	0.76	0.73	0.71	0.70	0.68	0.66
$R^{(2)}_{h,2000}$	1.04	1.06	1.06	1.06	1.06	1.05	1.05	1.04	1.03	1.03
$d_0=1.5$										
$m_h(\boldsymbol{\eta_0})$	3.00	9.20	17.80	28.64	41.67	57.00	74.85	95.45	119.05	145.88
$R^{(2)}_{h,500}$	0.87	0.83	0.81	0.79	0.78	0.76	0.74	0.73	0.71	0.70
$R^{(2)}_{h,2000}$	1.04	1.05	1.06	1.06	1.06	1.05	1.05	1.05	1.04	1.04
$d_0=2.0$										
$m_h(\boldsymbol{\eta_0})$	3.00	15.03	40.09	82.09	144.81	232.22	348.61	498.73	687.77	921.35
$R^{(2)}_{h,500}$	0.90	0.87	0.85	0.84	0.83	0.81	0.80	0.79	0.78	0.76
$R^{(2)}_{h,2000}$	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.98	0.97