Toward optimal multistep forecasts in unstable autoregressions

Ching-Kang Ing[a,*] Jin-Lung Lin[b] Shu-Hui Yu[c]

ABSTRACT This paper investigates multistep prediction errors for nonstationary autoregressive processes with both model order and true parameters unknown. We give asymptotic expressions for the multistep mean squared prediction errors and accumulated prediction errors of two important methods, plug-in and direct prediction. These expressions not only characterize how the prediction errors are influenced by the model orders, prediction methods, values of parameters, and unit roots, but also inspire us to construct some new predictor selection criteria that can ultimately choose the best combination of the model order and prediction method with probability 1. Finally, simulation analysis confirms the satisfactory finite sample performance of the newly proposed criteria.

Keywords: Accumulated prediction error, mean squared prediction error, plug-in method, direct prediction, model selection

---

[a]*Institute of Statistical Science, Academia Sinica and Department of Economics, National Taiwan University*

[b]*Institute of Economics, Academia Sinica and Department of Economics, National Chengchi University*

[c]*Institute of Statistical Science, Academia Sinica*

[*]Corresponding author. Tel.:886-2-27835611 ext 412; fax: 886-2-27831523
E-mail address: cking@stat.sinica.edu.tw

1. INTRODUCTION

Forecasting theory for stationary series with the true parameters known is well studied but not much is known about the case for nonstationary models with estimated parameters. To fill the gap, this paper investigates multistep prediction errors for autoregressive (AR) processes with unit root. The plug-in and direct predictors are the two most frequently used multistep prediction methods and comparing their relative performances has become a major issue in forecast theory. In the case of squared error losses, the plug-in predictor is obtained from repeatedly using the fitted (by least squares) AR model with an unknown future value replaced by their own forecasts, and the direct predictor is obtained by estimating the coefficient vector in the associated multistep prediction formula directly by linear least squares (see (1.2) and (1.3) below). Recently, many informative guidelines have been proposed to choose between these two methods in various time series models; see Findley (1983, 1984), Weiss (1991), Tiao and Tsay (1994), Lin and Tsay (1996), Stock and Watson (2002), Kang (2003), Ing (2003, 2004), Chevilloh and Hendry (2005), Schorfheide (2005), Massimilian, Stock, and Watson (2006), West (2006), White (2006), and Lin and Wei (2007), among many others. However, a theoretical resolution to the problem of how to select the optimal multistep predictor in nonstationary time series still seems to be lacking, at least when the estimation uncertainty is taken into account. In this paper, we have developed some predictor selection criteria to choose the model order and prediction method simultaneously and analyzed their theoretical properties rigorously. We consider this paper a major step toward optimal multistep forecasts in unstable regressions.

Assume that observations $x_1, \cdots, x_n$ are generated from a unit root AR model,

$$x_{t+1} = \sum_{i=1}^{p+1} a_i x_{t+1-i} + \varepsilon_{t+1}, \tag{1.1}$$

where $0 \le p < \infty$ is unknown, $a_{p+1} \neq 0$, ε_t's are white noises with zero means and common variance σ^2; and the characteristic polynomial

$$\begin{aligned} A(z) &= 1 - a_1 z - \cdots - a_p z^p - a_{p+1} z^{p+1} \\ &= (1-z)(1 - \alpha_1 z - \cdots - \alpha_p z^p), \end{aligned}$$

with $\alpha(z) = (1 - \alpha_1 z - \cdots - \alpha_p z^p) \neq 0$ for all $|z| \le 1$. x_t is called stationary or stable if all roots of A are outside the unit circle and unstable or nonstationary if some roots of A

are on the unit circle. For the sake of convenience, the initial conditions are set to $x_t = 0$ for all $t < 0$. To predict $x_{n+h}, h \geq 1$, based on $x_1, \cdots, x_n$ and a working model AR(k), one may use the plug-in predictor, $\hat{x}_{n+h}(k)$, or direct predictor, $\check{x}_{n+h}(k)$, where

$$\hat{x}_{n+h}(k) = \mathbf{x}_n'(k)\hat{\mathbf{a}}_n(h,k), \tag{1.2}$$

and

$$\check{x}_{n+h}(k) = \mathbf{x}_n'(k)\check{\mathbf{a}}_n(h,k), \tag{1.3}$$

with $\mathbf{x}_j(k) = (x_j, \ldots, x_{j-k+1})'$ being the regressor vector and $\hat{\mathbf{a}}_n(h,k)$ and $\check{\mathbf{a}}_n(h,k)$ being plug-in and direct estimators, respectively. Note that

$$\{\sum_{j=k}^{i-1} \mathbf{x}_j(k)\mathbf{x}_j(k)'\}\hat{\mathbf{a}}_i(1,k) = \sum_{j=k}^{i-1} \mathbf{x}_j(k)x_{j+1},$$
$$\{\sum_{j=k}^{i-h} \mathbf{x}_j(k)\mathbf{x}_j'(k)\}\check{\mathbf{a}}_i(h,k) = \sum_{j=k}^{i-h} \mathbf{x}_j(k)x_{j+h},$$

and $\hat{\mathbf{a}}_i(h,k) = \hat{A}_i^{h-1}(k)\hat{\mathbf{a}}_i(1,k)$, with $\hat{A}_i^0(k) = I_k$,

$$\hat{A}_i(k) = \left(\hat{\mathbf{a}}_i(1,k) \left| \begin{array}{c} I_{k-1} \\ \hline \mathbf{0}_{k-1}' \end{array} \right. \right),$$

and I_m and $\mathbf{0}_m$, respectively, denoting an identity matrix and a vector of zeros of dimension m. To assess the prediction performances of $\hat{x}_{n+h}(k)$ and $\check{x}_{n+h}(k)$, we consider their mean squared prediction errors (MSPEs),

$$\mathrm{MSPE}P_{n,h}(k) = E\left(x_{n+h} - \hat{x}_{n+h}(k)\right)^2,$$

and

$$\mathrm{MSPE}D_{n,h}(k) = E\left(x_{n+h} - \check{x}_{n+h}(k)\right)^2.$$

Theoretical investigations of $\mathrm{MSPE}P_{n,h}(k)$ (or $\mathrm{MSPE}D_{n,h}(k)$) in nonstationary AR models date back at least to Fuller and Hasza (1981). When $k \geq p+1$, an argument similar to that used in their Theorem 3.1 yields the following asymptotic expressions,

$$\mathrm{MSPE}P_{n,h}(k) = \sigma_h^2 + E\{R_{P,n}(k)\}, \tag{1.4}$$

and

$$\text{MSPE}D_{n,h}(k) = \sigma_h^2 + E\{R_{D,n}(k)\}, \tag{1.5}$$

where $R_{P,n}(k) = O_p(n^{-1})$, $R_{D,n}(k) = O_p(n^{-1})$, and $\sigma_h^2 = E(\eta_{t,h}^2)$, with $\eta_{t,h} = \sum_{j=0}^{h-1} b_j \varepsilon_{t+h-j}$, $b_j = \sum_{i=0}^{j} c_i$, $c_0 = 1$, and $c_j, j \geq 1$, satisfying $1 + \sum_{j=1}^{\infty} c_j z^j = 1/\alpha(z)$. The first term on the right-hand sides of (1.4) and (1.5), originating from the random disturbances, $\{\varepsilon_t\}$, is common for each multistep predictor, whereas the second terms on the right-hand sides of (1.4) and (1.5), arising from the estimation uncertainty, can vary with different k, different prediction methods, and different parameter values. However, since only rates of convergence of the second terms are reported, (1.4) and (1.5) fail to depict these features, which are indispensable in performing predictor comparisons. To remedy this difficulty, the constants associated with the terms of order n^{-1} in $E\{R_{P,n}(k)\}$ and $E\{R_{D,n}(k)\}$ need to be characterized. Recently, Ing (2001) made a first step towards this goal. In the special case where $p = 0$ in (1.1) (the random walk model) and $k = h = 1$, he showed that

$$n(\text{MSPE}P_{n,1}(1) - \sigma^2) = E\left\{\frac{x_n^2}{n} n^2(\hat{\mathbf{a}}_n(1,1) - 1)^2\right\} = 2\sigma^2. \tag{1.6}$$

The main obstacle in dealing with the above expectation, as argued by Ing, is the fact that the square of the normalized regressor, x_n^2/n, and the square of the normalized estimator, $n^2(\hat{\mathbf{a}}_n(1,1) - 1)^2$, are **not** asymptotically independent, a situation somewhat different from that encountered in the stationary case. While Ing was able to overcame this difficulty, his approach, focusing only on the random walk model and the case of one-step-ahead prediction, cannot be directly applied to more general nonstationary AR models or multistep prediction cases.

Another subtle problem, related to the direct method, can be illustrated using the following special case of (1.1),

$$(1 - B)(1 + 0.1B + 0.91B^2)x_{t+1} = (1 - 0.9B + 0.81B^2 - 0.91B^3)x_{t+1} = \varepsilon_{t+1}, \tag{1.7}$$

where B is the back shift operator. Simple algebra yields

$$x_{t+1} = 0.181x_{t-2} + 0.819x_{t-3} + \varepsilon_{t+1} + 0.9\varepsilon_t. \tag{1.8}$$

As observed in (1.8), the direct method only requires two regressors to make a three-step-ahead prediction, which indicates an interesting fact that the minimal correct order for the direct method, determined by the prediction lead time and unknown parameters, can be strictly less than that for the plug-in method. In general, model (1.1) can be rewritten as

$$x_{t+h} = (A^{h-1}(p+1)\mathbf{a}(p+1))'\mathbf{x}_t(p+1) + \eta_{t,h}, h \geq 1,$$

where $\mathbf{a}(k) = (a_1, \cdots, a_k)'$, with $a_j = 0$ for $j > p+1$,

$$A(k) = \left(\mathbf{a}(k) \left| \begin{array}{c} I_{k-1} \\ \hline \mathbf{0}'_{k-1} \end{array} \right. \right),$$

and $A^0(k) = I_k$. Let $\mathbf{a}(h, p+1) = (a_1(h, p+1), \cdots, a_{p+1}(h, p+1))' = A^{h-1}(p+1)\mathbf{a}(p+1)$. The above example leads us to define the minimal correct order for the h-step direct method, $p_h = \max\{j : 1 \leq j \leq p+1, a_j(h, p+1) \neq 0\}$. When $p_h < p_1$, by utilizing a more parsimonious prediction formula, the direct predictor can sometimes outperform the plug-in predictor in the correctly specified case, which creates additional complexity in predictor comparisons. For more details, see Examples 3 and 4 in Section 2.

In Section 2, we derive asymptotic expressions for $\mathrm{MSPE}P_{n,h}(k_1)$ and $\mathrm{MSPE}D_{n,h}(k_2)$ up to terms of order n^{-1}, where $k_1 \geq p_1 = p+1$ and $k_2 \geq p_h$. The constants associated with the terms of order n^{-1} in these expressions characterize how the prediction error is influenced by the orders, methods (plug-in or direct), values of parameters, and even the unit roots. In addition, they illuminate that to find the asymptotically optimal (from the MSPE point of view) multistep predictor among candidate plug-in and direct predictors, prediction orders and prediction methods must simultaneously be taken into account. The traditional order selection criteria can no longer serve the purpose. Section 3 is devoted to alleviating this difficulty. Our strategy is to find a statistic for each $\mathrm{MSPE}P_{n,h}(k)$ and $\mathrm{MSPE}P_{n,h}(k), k = 1, \cdots, K$, and show that the ordering of these statistics coincide with the ordering of their corresponding multistep MSPEs. Here, $K \geq p_1$ is a known integer. Inspired by Ing (2004), the statistics adopted in Section 3 are the multistep generalizations of accumulated prediction errors (APEs) based on sequential plug-in and direct predictors, namely,

$$\mathrm{APE}P_{n,h}(k) = \sum_{i=m_h}^{n-h} \left(x_{i+h} - \hat{x}_{i+h}(k) \right)^2, \tag{1.9}$$

and

$$\mathrm{APE}D_{n,h}(k)=\sum_{i=m_h}^{n-h}\left(x_{i+h}-\check{x}_{i+h}(k)\right)^2, \tag{1.10}$$

where m_h denotes the smallest positive number such that $\hat{\mathbf{a}}_i(h,K)$ and $\check{\mathbf{a}}_i(h,K)$ are well defined for all $i\geq m_h$. Note that $\mathrm{APE}P_{n,1}$ was first proposed by Rissanen (1986). A complete asymptotic analysis of $\mathrm{APE}P_{n,1}$ was given by Wei (1987, 1992) under a model more general than (1.1). However, due to some "nice" algebraic (or probability) structures in $\mathrm{APE}P_{n,1}$ are missing in its multistep counterparts (see Remarks 1 and 2 in Section 3), asymptotic properties of (1.9) and (1.10) in nonstationary AR processes still remain unknown. We propose a resolution of this problem, which shows that every $\mathrm{APE}P_{n,h}(k_1)$ and $\mathrm{APE}D_{n,h}(k_2)$, with $k_1\geq p_1$ and $k_2\geq p_h$, can be asymptotically decomposed into two terms, one of which, due to estimation uncertainty, is of order $\log n$, and the other, due to the random disturbances, is of order n and common for each predictor. More importantly, the constant associated with the term of $\log n$ in APE$P_{n,h}$ (APE$D_{n,h}$) is exactly the same as the one associated with the term of n^{-1} in its corresponding MSPE$P_{n,h}$ (MSPE$D_{n,h}$). This special feature enables us to show that Ing's (2004) asymptotically efficient predictor selection procedure (based on APE$P_{n,h}$ and APE$D_{n,h}$) in stationary AR processes is also asymptotically efficient in the presence of unit roots, and hence leads to a unified approach. Note that a predictor selection procedure is said to be asymptotically efficient if with probability 1, it can choose the order/method combination with the minimal MSPE for all sufficiently large n; see Section 3 for the exact definition.

Despite its theoretical advantage, Ing's procedure suffers from unsatisfactory finite-sample performances, as explained at the beginning of Section 4. To fix this flaw, a new predictor selection method is proposed in Section 4. This new method not only shares the same asymptotic advantage as Ing's procedure, it also has satisfactory finite-sample performances, which are illustrated at the end of Section 4 through a simulation experiment. Concluding remarks are given in Section 5. Appendices A, B, and C contain the proofs of the theorems in Sections 2, 3, and 4, respectively.

2. MSPEs of plug-in and direct predictors in the presence of unit roots

Throughout this section, it is assumed that in model (1.1) the ε_t's are independent random variables with zero means and variances $\sigma^2 > 0$. Moreover, there are small positive numbers α_1 and δ_1 and a large positive number M_1 such that for $0 \leq s - \nu \leq \delta_1$

$$\sup_{1 \leq m \leq t < \infty, \|\mathbf{v}_m\|=1} |F_{t,m,\mathbf{v}_m}(s) - F_{t,m,\mathbf{v}_m}(\nu)| \leq M_1(s-\nu)^{\alpha_1}, \tag{2.1}$$

where $\mathbf{v}_m = (v_1, \cdots, v_m)' \in R^m$, $\|\mathbf{v}_m\|^2 = \sum_{j=1}^m v_j^2$, and $F_{t,m,\mathbf{v}_m}(\cdot)$ denotes the distribution of $\sum_{l=1}^m v_l \varepsilon_{t+1-l}$.

In the case where ε_t's are i.i.d., the following lemma provides sufficient conditions under which (2.1) is fulfilled.

Lemma 2.1 *(Ing and Sin (2007, Lemma 4)). Let ε_t's be i.i.d. random variables satisfying $E(\varepsilon_1) = 0, E(\varepsilon_1^2) > 0$, and $E(|\varepsilon_1|^\alpha) < \infty$ for some $\alpha > 2$. Assume also that for some positive constant $M_2 < \infty$,*

$$\int_{-\infty}^{\infty} |\varphi(t)| dt \leq M_2, \tag{2.2}$$

where $\varphi(t) = E(e^{it\varepsilon_1})$ is the characteristic function of ε_1. Then, for all $-\infty < t < \infty$, $m \geq 1$, $\mathbf{r}_m \in R^m$, and $\|\mathbf{r}_m\| = 1$, there is a finite positive constant M_3 such that

$$\sup_{-\infty < x < \infty} f_{t,m,\mathbf{r}_m}(x) < M_3,$$

where $f_{t,m,\mathbf{r}_m}(\cdot)$ is the density function of $\sum_{j=1}^m r_j \varepsilon_{t+1-j}$, with $(r_1, \cdots, r_m)' = \mathbf{r}_m$. As a result, (2.1) follows.

Since (2.2) is satisfied by most absolutely continuous distributions, (2.1) is flexible enough to accommodate a wide range time series applications. Note that (2.1) is given to ensure that the inverses of the normalized Fisher's information matrices, $\hat{R}_n^{-1}(k)$ and $\bar{R}_{n,h}^{-1}(k)$, have finite positive moments in the senses of (A.1) and (A.19) (in Appendix A), where

$$\hat{R}_n(k) = \frac{1}{n} D_n(k) \sum_{j=k}^{n-1} \mathbf{x}_j(k) \mathbf{x}_j'(k) D_n(k)',$$

with

$$D_n(k) = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -1 \\ \frac{1}{\sqrt{n}} & \frac{-\alpha_1}{\sqrt{n}} & \dots & \dots & \frac{-\alpha_{k-1}}{\sqrt{n}} \end{pmatrix};$$

and

$$\bar{R}_{n,h}(k) = \frac{1}{n}\bar{D}_n(k)\sum_{j=k}^{n-h} \mathbf{x}_j(k)\mathbf{x}'_j(k)\bar{D}_n(k)',$$

with

$$\bar{D}_n(k) = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -1 \\ \frac{1}{\sqrt{n}} & 0 & \dots & \dots & 0 \end{pmatrix}.$$

These moment results will be used to deal with the asymptotic properties of MSPE$P_{h,n}$ and MSPE$D_{h,n}$; see the proofs of Theorems 2.2 and 2.3 for details. Theorems 2.2 below provides an asymptotic expression for $\mathrm{MSPE}P_{n,h}(k)$ with $k \geq p_1$. Before stating the result, we need to define $S_M^0(k) = I_k$ and

$$S_M(k) = \left(\begin{array}{c|c} \alpha(k) & \begin{array}{c} I_{k-1} \\ \hline \mathbf{0}'_{k-1} \end{array} \end{array} \right),$$

where $\alpha(k) = (\alpha_1, \cdots, \alpha_k)'$ with $\alpha_j = 0$ for $j > p$.

Theorem 2.2. Assume that $\{x_t\}$ satisfies model (1.1). Also assume that $\{\varepsilon_t\}$ satisfies (2.1) and

$$E\left(|\,\varepsilon_1\,|^{\theta_h}\right) < \infty,$$

where $\theta_h = \max\{8,\, 2(h+2)\} + \delta$ for some $\delta > 0$. Then, for $k \geq p_1$ and $h \geq 1$,

$$n\left(\mathrm{MSPE}P_{n,h}(k) - \sigma_h^2\right) = 2\sigma^2(\sum_{j=0}^{h-1} b_j)^2 + f_{1,h}(k-1) + o(1), \qquad (2.3)$$

where $f_{1,h}(0) = 0$ and for $k \geq 2$,

$$f_{1,h}(k-1) = \text{tr}\left(\Gamma(k-1)M_h(k-1)\Gamma^{-1}(k-1)M_h'(k-1)\right)\sigma^2,$$

with $M_h(k-1) = \sum_{j=0}^{h-1} b_j S_M^{h-1-j}(k-1)$, $\Gamma(k-1) = \lim_{j\to\infty} E(\mathbf{s}_j(k-1)\mathbf{s}_j'(k-1))$, $\mathbf{s}_j(k-1) = (s_j, \cdots, s_{j-k+2})'$, and $s_j = x_j - x_{j-1}$.

An asymptotic expression for $\text{MSPE}D_{n,h}(k)$, with $k \geq p_h$, is given as follows.

Theorem 2.3. Let the assumptions of Theorem 2.1 hold, with θ_h replaced by $8+\delta$ for some $\delta > 0$. Then, for $k \geq p_h$ and $h \geq 1$,

$$n\left(\text{MSPE}D_{n,h}(k) - \sigma_h^2\right) = 2\sigma^2(\sum_{j=0}^{h-1} b_j)^2 + f_{2,h}(k-1) + o(1), \tag{2.4}$$

where $f_{2,h}(0) = 0$, for $k \geq 2$,

$$f_{2,h}(k-1) = \text{tr}\left\{\Gamma^{-1}(k-1)\lim_{t\to\infty}\text{cov}\left(\sum_{j=0}^{h-1} b_j\mathbf{s}_{t+j}(k-1)\right)\right\}\sigma^2,$$

and for random vector $\mathbf{y}$, $\text{cov}(\mathbf{y}) = E\{(\mathbf{y}-E(\mathbf{y}))(\mathbf{y}-E(\mathbf{y}))'\}$.

Theorems 2.2 and 2.3 show that each $n(\text{MSPE}P_{n,h}(k_1) - \sigma_h^2)$ and $n(\text{MSPE}D_{n,h}(k_2) - \sigma_h^2)$, with $k_1 \geq p_1$ and $k_2 \geq p_h$, can be asymptotically decomposed as a sum of two terms. The first term, $2\sigma^2(\sum_{j=0}^{h-1} b_j)^2$, arising from predicting the nonstationary component in model (1.1), is common for each predictor, whereas the second term, $f_{1,h}(k-1)$ (or $f_{2,h}(k-1)$), arising from predicting the stationary component in model (1.1), can vary with different orders and methods. The following examples help provide a better understanding of Theorems 2.2 and 2.3.

Example 1. When $p_1 = 1$ (the random walk model) and $k = 1$, (2.3) and (2.4) imply that $\hat{x}_{n+h}(1)$ and $\check{x}_{n+h}(1)$ are asymptotically equivalent in the sense that

$$\lim_{n\to\infty}\frac{\text{MSPE}P_{n,h}(1) - \sigma_h^2}{\text{MSPE}D_{n,h}(1) - \sigma_h^2} = 1.$$

Example 2. When $k \geq \max\{2, p_1\}$ and $h = 2$, by (2.3) and (2.4), it is straightforward to show that

$$f_{1,2}(k-1) = \{(k-2) + \alpha_{k-1}^2 + 2\alpha_1 b_1 + b_1^2(k-1)\}\sigma^2, \tag{2.5}$$

and

$$f_{2,2}(k-1) = \{(k-1)(1+b_1^2) + 2\alpha_1 b_1\}\sigma^2, \tag{2.6}$$

which yields

$$f_{2,2}(k-1) - f_{1,2}(k-1) = (1-\alpha_{k-1}^2)\sigma^2 > 0. \tag{2.7}$$

Moreover, by an argument similar to that used to prove (17) of Ing (2003), it can be shown that for $k \geq \max\{2, p_1\}$ and $h \geq 2$,

$$f_{2,h}(k-1) - f_{1,h}(k-1) \geq f_{2,2}(k-1) - f_{1,2}(k-1) > 0, \tag{2.8}$$

and hence $\hat{x}_{n+h}(k)$ is asymptotically more efficient than $\check{x}_{n+h}(k)$ in this case.

As shown in Section 1, it is possible that $p_h < p_1$. In this case, it would be more interesting to compare $n(\text{MSPE}P_{n,h}(p_1) - \sigma_h^2)$ and $n(\text{MSPE}D_{n,h}(p_h) - \sigma_h^2)$ rather than those MSPEs of the same order. The following example shows that the advantage of the plug-in predictor illustrated in Example 2 vanishes in this kind of comparison.

Example 3. Assume

$$(1-B)(1+a_1B+\cdots+a_pB^p)x_t = e_t,$$

where $p \geq 2$, $1+a_1z+\cdots+a_pz^p \neq 0$ for $|z| \leq 1$, and $a_p \neq 0$. If $a_1 = 1$, then it is not difficult to see that $p_2 = p_1 - 1 = p$ and $f_{2,2}(p) - f_{2,2}(p-1) = \sigma^2$. In addition, (2.7) implies $f_{2,2}(p) - f_{1,2}(p) = (1-a_p^2)\sigma^2$. As a result,

$$n\{\text{MSPE}P_{n,2}(p_1) - \sigma^2\} - n\{\text{MSPE}D_{n,2}(p_2) - \sigma^2\} \to a_p^2\sigma^2 > 0,$$

as $n \to \infty$. Hence $\check{x}_{n+2}(p_2)$ is asymptotically more efficient than $\hat{x}_{n+2}(p_1)$ in this case.

When $h = 2$ and $p_1 \geq 2$, Examples 2 and 3 together suggest a simple rule that $\hat{x}_{n+2}(p_1)$ is asymptotically more efficient than $\check{x}_{n+2}(p_2)$ if $p_1 = p_2$; and the conclusion is reversed if $p_1 > p_2$. This rule, however, fails to hold for $h \geq 3$, as detailed in the following example.

Example 4. Consider the following AR(4) model

$$\begin{aligned}
& (1-B)(1+a_1B)(1+a_2B^2)x_t \\
= \; & \{1-(1-a_1)B-(a_1-a_2)B^2-a_2(1-a_1)B^3-a_1a_2B^4\}x_t = e_t,
\end{aligned}$$

where $0 < a_1 < 1$ and $a_2 = a_1^2 - a_1 + 1$. It is straightforward to show that $p_3 = 3 = p_1 - 1$. By numerical calculations, we obtain the values of $f_{2,3}(2) - f_{1,3}(3)$, with $a_1 = 0.1, 0.2, \cdots, 0.9$; see Table I. According to Table I, $\check{x}_{n+3}(p_3)$ is asymptotically more efficient than $\hat{x}_{n+3}(p_1)$ in cases of $a_1 = 0.1, 0.2, 0.9$, and less efficient than $\hat{x}_{n+3}(p_1)$ in all other cases.

Consequently, when $h \geq 3$, the rankings of $\hat{x}_{n+h}(p_1)$ and $\check{x}_{n+h}(p_h)$ are determined not only by whether $p_h < p_1$, but also by the values of the unknown parameters. Simply determining p_1 or p_h through certain consistent model selection techniques cannot guarantee optimal multistep prediction (from the MSPE point of view) in situations where plug-in and direct predictors are simultaneously taken into account. To fundamentally solve this problem, some ideas that go beyond the conventional model (order) selection are required. Our proposals toward this problem are given in the next two sections.

3. Multistep accumulated prediction errors

Let $\hat{x}_{n+h}(k), k = 1, \cdots, K$, and $\check{x}_{n+h}(k), k = 1, \cdots, K$, be candidate plug-in and direct predictors, where $h \geq 1$ and $K \geq p_1$. For convenience, we use $(k, 1)$ to denote $\hat{x}_{n+h}(k)$ and $(k, 2)$ to denote $\check{x}_{n+h}(k)$. Instead of identifying p_1 or p_h, this section attempts to choose the order/method combination having the minimal MSPE, at least when n is sufficiently large. To this end, the loss function of $(k, 1)$ and $(k, 2)$ are defined to be

$$L_{1,h}(k) = \begin{cases} \lim_{n\to\infty} n\left(\text{MSPE}P_{n,h}(k) - \sigma_h^2\right) & \text{if} \quad p_1 \leq k \leq K \\ \infty & \text{if} \quad k < p_1, \end{cases} \tag{3.1}$$

and

$$L_{2,h}(k) = \begin{cases} \lim_{n\to\infty} n\left(\text{MSPE}D_{n,h}(k) - \sigma_h^2\right) & \text{if} \quad p_h \leq k \leq K \\ \infty & \text{if} \quad k < p_h, \end{cases} \tag{3.2}$$

respectively. Note that the existence of the above limits are ensured by Theorems 2.2 and 2.3; and to have the prediction loss due to underspecification be much larger than the one

TABLE I

THE VALUES OF Diff $= f_{2,3}(2) - f_{1,3}(3)$

a_1	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Diff	-0.378	-0.013	0.197	0.310	0.354	0.336	0.247	0.051	-0.321

due to overspecification, the loss function values of $(k, 1)$ with $k < p_1$ and $(k, 2)$ with $k < p_h$ are set to ∞. A predictor selection criterion, $(\tilde{k}_n, \tilde{j}_n)$, with $1 \leq \tilde{k}_n \leq K$ and $1 \leq \tilde{j}_n \leq 2$, is said to be asymptotically efficient if

$$P\left((\tilde{k}_n, \tilde{j}_n) \in C_{h,K},\ \mathit{eventually}\ \right) = 1, \tag{3.3}$$

where

$$C_{h,K} = \left\{(k, j) : 1 \leq k \leq K, 1 \leq j \leq 2,\ \text{and}\ L_{j,h}(k) = \min_{1\leq k_0 \leq K, 1 \leq j_0 \leq 2} L_{j_0,h}(k_0)\right\}.$$

Therefore, with probability 1 $(\tilde{k}_n, \tilde{j}_n)$ can choose the predictor having the minimal loss function value for all sufficiently large n.

The goal of this section is to show that (3.3) is fulfilled by $(\hat{k}_n, \hat{j}_n)$. Here, $(\hat{k}_n, \hat{j}_n)$, first proposed by Ing (2004), is obtained through the following procedure.

Step 1. Define $\hat{k}_{D,n}^{(1)} = \arg \min_{1\leq k \leq K} \mathrm{APE}D_{n,1}(k)$.

Step 2. Define

$$\hat{k}_{D,n}^{(h)} = \arg \min_{1\leq k \leq K} \mathrm{APE}D_{n,h}(k),$$

and define

$$\hat{k}_{n}^{(1,h)} = \arg \min_{\hat{k}_{D,n}^{(1)} \leq k \leq K} \mathrm{APE}P_{n,h}(k).$$

Step 3. If $\mathrm{APE}D_{n,h}(\hat{k}_{D,n}^{(h)}) > \mathrm{APE}P_{n,h}(\hat{k}_{n}^{(1,h)})$, then $(\hat{k}_n, \hat{j}_n) = (\hat{k}_{n}^{(1,h)}, 1)$; otherwise $(\hat{k}_n, \hat{j}_n) = (\hat{k}_{D,n}^{(h)}, 2)$.

In the sequel, the above procedure will be referred to as Procedure I. We begin by investigating the asymptotic properties of $\mathrm{APE}P_{n,h}(k)$ and $\mathrm{APE}D_{n,h}(k)$ in the correctly specified case. Note that for $k \geq p_1$,

$$\mathrm{APE}P_{n,h}(k) = \sum_{i=m_h}^{n-h} \left\{ \eta_{i,h} - \mathbf{x}_i'(k)\hat{L}_{i,h}(k)\left(\hat{\mathbf{a}}_i(1, k) - \mathbf{a}(k)\right)\right\}^2, \tag{3.4}$$

where $\hat{L}_{i,h}(k) = \sum_{j=0}^{h-1} b_j \hat{A}_i^{h-1-j}(k)$; and for $k \geq p_h$,

$$\text{APE}D_{n,h}(k) = \sum_{i=m_h}^{n-h} \{\, \eta_{i,h} - \mathbf{x}_i'(k)\, (\, \check{\mathbf{a}}_i(h,k) - \mathbf{a}_D(h,k)\,)\,\}^2 \,, \tag{3.5}$$

where $\mathbf{a}_D(h,k) = (a_1(h,p+1), \cdots, a_k(h,p+1))'$, with $a_j(h,p+1), 1 \leq j \leq p+1$, defined in Section 1 and $a_j(h,p+1) = 0$ if $j > p+1$.

Theorem 3.1. *Assume that $\{x_t\}$ satisfies model (1.1) and $\{\varepsilon_t\}$ is a sequence of independent random noises with zero means and common variance $\sigma^2 > 0$. Moreover, assume $\sup_t E(|\varepsilon_t|^\alpha) < \infty$, for some $\alpha > 2$. Then, for $k \geq p_1$ and $h \geq 1$,*

$$\begin{aligned} \text{APE}P_{n,h}(k) - \sum_{i=m_h}^{n-h} \eta_{i,h}^2 &= \{2\sigma^2 (\sum_{j=0}^{h-1} b_j)^2 + f_{1,h}(k-1)\} \log n + o\,(\log n) \quad \text{a.s.} \\ &= L_{1,h}(k) \log n + o\,(\log n) \quad \text{a.s.} \end{aligned} \tag{3.6}$$

Remark 1. As shown in (B.26),

$$\begin{aligned} \text{APE}P_{n,h}(k) - \sum_{i=m_h}^{n-h} (\eta_{i,h})^2 &= \sum_{i=m_h}^{n-h} \left\{ \mathbf{x}_i'(k) \hat{L}_{i,h}(k)\, (\hat{\mathbf{a}}_i(1,k) - \mathbf{a}(k)) \right\}^2 (1 + o(1)) \\ &+ O(1) \quad \text{a.s.} \end{aligned}$$

Therefore, the main task of proving (3.6) is to explore the almost sure properties of the first term on the right-hand side of the above equality. Through a recursive expression for $Q_n(1,k)$, where with $V_i^{-1}(k) = \sum_{i=k}^{i} \mathbf{x}_i(k)\mathbf{x}_i'(k)$,

$$\begin{aligned} Q_n(1,k) &= \sum_{i=k}^{n-1} \{\mathbf{x}_i'(k)\, (\, \hat{\mathbf{a}}_n(1,k) - \mathbf{a}(k)\,)\,\}^2 \\ &= \left(\sum_{i=k}^{n-1} \mathbf{x}_i'(k) \varepsilon_{i+1} \right) V_{n-1}(k) \left(\sum_{i=k}^{n-1} \mathbf{x}_i(k) \varepsilon_{i+1} \right) \end{aligned}$$

is the (second-order) residual sum of squares for one-step predictions, Lai and Wei (1982) established a connection between $Q_n(1,k)$ and its sequential counterpart,

$$\sum_{i=m_1}^{n-1} \{\mathbf{x}_i'(k)\, (\, \hat{\mathbf{a}}_i(1,k) - \mathbf{a}(k)\,)\,\}^2 = \sum_{i=m_h}^{n-1} \left\{ \mathbf{x}_i'(k) V_{i-1}(k) \left(\sum_{j=k}^{i-1} \mathbf{x}_j(k) \varepsilon_{j+1} \right) \right\}^2 . \tag{3.7}$$

Based on this connection and some strong laws for martingales, Wei (1987, 1992) subsequently obtained an asymptotic expression for the left-hand side of (3.6) in the case of $h = 1$. However, it is extremely difficult to obtain an analyzable recursive formula for the multistep analog of $Q_n(1,k)$, $Q_n(h,k) = \sum_{i=k}^{n-h} \{\mathbf{x}_i'(k)\hat{L}_{n,h}(k)(\hat{\mathbf{a}}_n(1,k) - \mathbf{a}(k))\}^2, h \geq 2$, due to the appearance of $\hat{L}_{n,h}(k)$. Hence, Wei's approach is not easy to be extended to the case of multistep predictions. By observing

$$Q_n(h,k) = (\sum_{i=k}^{n-1} \mathbf{x}_i'(k)\varepsilon_{i+1})S_n'(k)V_{n-h}(k)S_n(k)(\sum_{i=k}^{n-1} \mathbf{x}_i'(k)\varepsilon_{i+1}),$$

where

$$S_n(k) = (\sum_{i=k}^{n-h} \mathbf{x}_i(k)\mathbf{x}_i'(k))\hat{L}_{n,h}(k)(\sum_{i=k}^{n-1} \mathbf{x}_i(k)\mathbf{x}_i'(k))^{-1},$$

Ing (2004), under stationary AR processes, adopted

$$Q_n^*(h,k) = (\sum_{i=k}^{n-1} \mathbf{x}_i'(k)\varepsilon_{i+1})S'(k)V_{n-h}(k)S(k)(\sum_{i=k}^{n-1} \mathbf{x}_i'(k)\varepsilon_{i+1})$$

to replace $Q_n(h,k)$, where $S(k)$ is the almost sure limit of $S_n(k)$ that is a nonrandom matrix. He then obtained a recursive formula for $Q_n^*(h,k)$ and established a connection between $Q_n^*(h,k)$ and $\sum_{i=k}^{n-h}\{\mathbf{x}_i'(k)L_{i,h}(k)(\hat{\mathbf{a}}_i(1,k) - \mathbf{a}(k))\}^2$, which further yields an asymptotic expression for the latter. Unfortunately, when model (1.1) is assumed, $S_n(k)$, with $k \geq 2$, no longer has an almost sure and nonrandom limit, which makes it hard to apply Ing's (2004) approach to the nonstationary case. To obtain (3.6), extra effort is made to overcome the above difficulties; see Appendix B for details.

Theorem 3.2. *Let the assumptions of Theorem 3.1 hold. Then, for $k \geq p_h$ and $h \geq 1$,*

$$\begin{aligned} \text{APED}_{n,h}(k) - \sum_{i=m_h}^{n-h} \eta_{i,h}^2 &= \{2\sigma^2(\sum_{j=0}^{h-1} b_j)^2 + f_{2,h}(k-1)\} \log n + o(\log n) \quad \text{a.s.} \\ &= L_{2,h}(k) \log n + o(\log n) \quad \text{a.s.} \end{aligned} \tag{3.8}$$

Remark 2. As indicated in (B.43),

$$\begin{aligned} \text{APED}_{n,h}(k) - \sum_{i=m_h}^{n-h} \eta_{i,h}^2 &= (1+o(1)) \sum_{i=m_h}^{n-h} \left\{ \mathbf{x}_i'(k)V_{i-h}(k) \left(\sum_{j=k}^{i-h} \mathbf{x}_j(k)\eta_{j,h} \right) \right\}^2 \\ &+ O(1) \quad \text{a.s.} \end{aligned}$$

While

$$\sum_{i=m_h}^{n-h}\left\{\mathbf{x}_i'(k)V_{i-h}(k)\left(\sum_{j=k}^{i-h}\mathbf{x}_j(k)\eta_{j,h}\right)\right\}^2$$

looks very similar to (3.7), Wei's approach for the one-step APE still cannot be applied to it because $\sum_{j=k}^{i-h}\mathbf{x}_j(k)\eta_{j,h}, h \geq 2$, is not a martingale transformation. While Ing (2004) resolved this difficulty in stationary AR model, his method, which is highly reliant on the stationary assumption, is not applicable to the unit root processes.

Remark 3. Theorems 2.2, 2.3, 3.1, and 3.2 together disclose a fascinating fact that the constants associated with the terms of order n^{-1} in MSPE$P_{n,h}(k_1)$ and MSPE$D_{n,h}(k_2)$, with $k_1 \geq p_1$ and $k_2 \geq p_h$, are exactly the same as the constants associated with the terms of order $\log n$ in their corresponding multistep APEs. While MSPE$P_{n,h}(k_1)$ and MSPE$D_{n,h}(k_2)$ are unobservable, this special property allows us to preserve their asymptotic rankings through the values of the associated multistep APEs, which can be easily obtained from the data. This is also the driving motivation for constructing $(\hat{k}_n, \hat{j}_n)$ in model (1.1).

Before showing the asymptotic efficiency of $(\hat{k}_n, \hat{j}_n)$, we need to investigate the asymptotic properties of APE$D_{n,k}(k)$ in misspecified cases.

Theorem 3.3. *Let the assumptions of Theorem 3.1 hold. Then, for* $1 \leq k < p_h$ *and* $h \geq 1$,

$$\liminf_{n\to\infty}\frac{1}{n}\left(\mathrm{APE}D_{n,h}(k) - \sum_{j=m_h}^{n-h}\eta_{j,h}^2\right) > 0 \text{ a.s.} \tag{3.9}$$

We are now in a position to state the main result of this section.

Theorem 3.4. *Let the assumptions of Theorem 3.1 hold. Then, for* $K \geq p_1$, $(\hat{k}_n, \hat{j}_n)$ *is asymptotically efficient in the sense of (3.3).*

Remark 4. Since Ing (2004) showed that $(\hat{k}_n, \hat{j}_n)$ is also asymptotically efficient in stationary AR models, Theorem 3.4, together with Ing's result, provides a unified approach for choosing the (asymptotically) optimal multistep predictor for AR processes with or without unit roots. While it is possible to select multistep predictors after unit root tests are performed (which means that the selection procedure will be carried out based on the differenced data if the unit-root hypothesis is rejected), all unit root tests suffer from low power when the process is near unity. One can hardly expect a reliable selection/prediction result once the process is erroneously differenced.

4. New Criteria

In view of (3.4) and (3.5), for $k_1 \geq p_1$ and $k_2 \geq p_h$,

$$\mathrm{APE}P_{n,h}(k_1) - \mathrm{APE}D_{n,h}(k_2) = \sum_{i=m_h}^{n-h} \left\{ \mathbf{x}_i'(k_1)\hat{L}_{i,h}(k_1)\,(\,\hat{\mathbf{a}}_i(1,k_1) - \mathbf{a}(k_1)\,)\,\right\}^2$$
$$- \sum_{i=m_h}^{n-h} \{\mathbf{x}_i'(k_2)\,(\,\check{\mathbf{a}}_i(h,k_2) - \mathbf{a}_D(h,k_2)\,)\,\}^2 - 2\sum_{i=m_h}^{n-h} \mathbf{x}_i'(k_1)\hat{L}_{i,h}(k_1)\,(\,\hat{\mathbf{a}}_i(1,k_1) - \mathbf{a}(k_1)\,)\,\eta_{i,h}$$
$$+ \quad 2\sum_{i=m_h}^{n-h} \mathbf{x}_i'(k_2)\,(\,\check{\mathbf{a}}_i(h,k_2) - \mathbf{a}_D(h,k_2)\,)\,\eta_{i,h} \equiv (I) - (II) - (III) + (IV). \tag{4.1}$$

Although the cross-product terms, (III) and (IV), in (4.1) are almost surely of order $o(\log n)$ and asymptotically negligible compared to (I) and (II) (see Appendix B), we have found that the finite sample effects of (III) and (IV) can differ remarkably. This ”nonuniformity” feature causes ”rank-distortion” when we perform cross-method comparisons. Simulation results show that the rankings of APE$P_{n,h}(k_1)$ and APE$D_{n,h}(k_2)$ are often inconsistent with the rankings $L_{1,h}(k_1)$ and $L_{2,h}(k_2)$ even when $n > 500$, which leads to unsatisfactory selection results in finite samples. (Recall that the rankings of $L_{1,h}(k_1)$ and $L_{2,h}(k_2)$ almost surely coincide with the ”limiting” rankings of APE$P_{n,h}(k_1)$ and APE$D_{n,h}(k_2)$.)

To overcome the above difficulty, we consider using $\mathrm{PMIC}_{n,h}(k)$ and $\mathrm{DMIC}_{n,h}(k)$ to replace $\mathrm{APE}P_{n,h}(k)$ and $\mathrm{APE}D_{n,h}(k)$ in Procedure I, where

$$\mathrm{PMIC}_{n,h}(k) = \hat{\sigma}^2_{P,n}(h,k) +$$
$$tr\left\{ (\sum_{j=k}^{n-h} \mathbf{x}_j(k)\mathbf{x}_j'(k))\ddot{L}_{h,n}(k)(\sum_{j=k}^{n-h} \mathbf{x}_j(k)\mathbf{x}_j'(k))^{-1}\ddot{L}_{h,n}'(k) \right\} \tilde{\sigma}^2_n C_n, \tag{4.2}$$

and

$$\mathrm{DMIC}_{n,h}(k) = \hat{\sigma}^2_D(h,k) +$$

$$tr\left\{(\sum_{j=k}^{n-h}\mathbf{x}_j(k)\mathbf{x}_j'(k))^{-1}(\sum_{j=k}^{n-2h+1}\mathbf{z}_j(k)\mathbf{z}_j'(k))\right\}\tilde{\sigma}_n^2C_n, \tag{4.3}$$

where $\lim_{n\to\infty}C_n = 0$ and $\liminf_{n\to\infty}C_nn/\log n > 0$. Note that $\hat{\sigma}^2_{P,n}(h,k) = (n-h-K)^{-1}\sum_{j=K}^{n-h}\{x_{j+h}-\hat{\mathbf{a}}_n(h,k)\mathbf{x}_j(k)\}^2$ and $\hat{\sigma}^2_{D,n}(h,k) = (n-h-K)^{-1}\sum_{j=K}^{n-h}\{x_{j+h}-\check{\mathbf{a}}_n(h,k)\mathbf{x}_j(k)\}^2$ are the h-step residual mean squared errors obtained from the k-regressor plug-in and direct methods, respectively, $\tilde{\sigma}_n^2 = \hat{\sigma}^2_{P,n}(1,K) = \hat{\sigma}^2_{D,n}(1,K)$ is the one-step residual mean squared error obtained from the largest candidate model, $\mathbf{z}_j(k) = \sum_{i=0}^{h-1}\hat{b}_{i,n}\mathbf{x}_{j+i}(k)$, and $\ddot{L}_{h,n}(k) = \sum_{j=0}^{h-1}\hat{b}_{j,n}\hat{A}_n^{h-1-j}(k)$, where $\hat{b}_{0,n} = 1$, and for $j\geq 1$, $\hat{b}_{j,n} = \sum_{l=1}^{j}\hat{b}_{j-l,n}\hat{a}_{l,n}(1,K)$, with $(\hat{a}_{1,n}(1,K),\cdots,\hat{a}_{k,n}(1,K))' = \hat{\mathbf{a}}_n(1,K)$ and $\hat{a}_{l,n}(1,K) = 0$ if $l > K$.

First observe that

$$\begin{aligned}
& \mathrm{PMIC}_{n,h}(k_1) - \mathrm{DMIC}_{n,h}(k_2)\\
= & \ \hat{\sigma}^2_{P,n}(h,k_1) - \hat{\sigma}^2_D(h,k_2)\\
+ & \ tr\left\{(\sum_{j=k_1}^{n-h}\mathbf{x}_j(k_1)\mathbf{x}_j'(k_1))\ddot{L}_{h,n}(k_1)(\sum_{j=k_1}^{n-h}\mathbf{x}_j(k_1)\mathbf{x}_j'(k_1))^{-1}\ddot{L}'_{h,n}(k_1)\right\}\tilde{\sigma}_n^2C_n\\
- & \ tr\left\{(\sum_{j=k_2}^{n-h}\mathbf{x}_j(k_2)\mathbf{x}_j'(k_2))^{-1}(\sum_{j=k_2}^{n-2h+1}\mathbf{z}_j(k_2)\mathbf{z}_j'(k_2))\right\}\tilde{\sigma}_n^2C_n. \qquad (4.4)
\end{aligned}$$

It is shown in Appendix C that when $k_1\geq p_1$ and $k_2\geq p_h$,

$$\begin{aligned}
& tr\left\{(\sum_{j=k_1}^{n-h}\mathbf{x}_j(k_1)\mathbf{x}_j'(k_1))\ddot{L}_{h,n}(k_1)(\sum_{j=k_1}^{n-h}\mathbf{x}_j(k_1)\mathbf{x}_j'(k_1))^{-1}\ddot{L}'_{h,n}(k_1)\right\}\tilde{\sigma}_n^2\\
- & \ tr\left\{(\sum_{j=k_2}^{n-h}\mathbf{x}_j(k_2)\mathbf{x}_j'(k_2))^{-1}(\sum_{j=k_2}^{n-2h+1}\mathbf{z}_j(k_2)\mathbf{z}_j'(k_2))\right\}\tilde{\sigma}_n^2\\
= & \ L_{1,h}(k_1) - L_{2,h}(k_2) + o(1)\ a.s. \qquad (4.5)
\end{aligned}$$

Therefore, the trace terms in (4.2) and (4.3) play roles in keeping the rankings of their corresponding loss functions. On the other hand, for $k_1\geq p_1$ and $k_2\geq p_h$, the weight associated with the trace terms, C_n, asymptotically dominates $\hat{\sigma}^2_{P,n}(h,k_1) - \hat{\sigma}^2_D(h,k_2)$ (see (C.2)), which helps to protect the trace term effects in (4.4) from being distorted by $\hat{\sigma}^2_{P,n}(h,k_1) - \hat{\sigma}^2_D(h,k_2)$. In fact, our simulations reveal that this domination usually occurs quite early (particular when C_n is relatively large), and hence considerably alleviate the dilemma encountered by Procedure I in finite samples. (Note that $\hat{\sigma}^2_{P,n}(h,k)$ and $\hat{\sigma}^2_{D,n}(h,k)$ cannot be dropped from (4.2) and (4.3) because they are necessary for preventing underspecification; see, e.g., (C.1).) The following is the new predictor selection procedure (which is referred to as Procedure II) and its asymptotic property.

Step 1. Define $\hat{O}_n^{(1)} = \arg\min_{1\le k\le K} \mathrm{DMIC}_{n,1}(k)$.

Step 2. Define

$$\hat{O}_n^{(h)} = \arg\min_{1\le k\le K} \mathrm{DMIC}_{n,h}(k),$$

and define

$$\hat{O}_n^{(1,h)} = \arg\min_{\hat{O}_n^{(1)}\le k\le K} \mathrm{PMIC}_{n,h}(k).$$

Step 3. If $\mathrm{DMIC}_{n,h}(\hat{O}_n^{(h)}) > \mathrm{PMIC}_{n,h}(\hat{O}_n^{(1,h)})$, then $(\hat{O}_n, \hat{M}_n) = (\hat{O}_n^{(1,h)}, 1)$; otherwise $(\hat{O}_n, \hat{M}_n) = (\hat{O}_n^{(h)}, 2)$.

Theorem 4.1 *Let the assumptions of Theorem 3.1 hold. Then, for $K \ge p_1$, $(\hat{O}_n, \hat{M}_n)$ is asymptotically efficient in the sense of (3.3).*

Remark 5. Although (4.5) holds, it is worth mentioning the trace terms in (4.5) are not consistent estimators of their corresponding loss functions $L_{1,h}(k_1)$ and $L_{2,h}(k_2)$; see (C.5) and (C.6) in Appendix C.

Remark 6. Following an argument similar to that used in the proof of Theorem 4.1, it is not difficult to show that $(\hat{O}_n, \hat{M}_n)$ is also asymptotically efficient in stationary AR models.

To illustrate the asymptotic results obtained in Theorem 4.1, we conduct a simulation study. The data generating processes (DGPs) are given by

DGP I. $x_t = -0.8x_{t-2} + \varepsilon_t$,

DGP II $x_t = 0.3x_{t-1} - 0.8x_{t-2} + \varepsilon_t$,

DGP III $x_t = 0.2x_{t-2} + 0.8x_{t-3} + \varepsilon_t$,

DGP IV $x_t = 0.3x_{t-1} - 0.1x_{t-2} + 0.8x_{t-3} + \varepsilon_t$,

DGP V $x_t = 0.9x_{t-1} - 0.81x_{t-2} + \varepsilon_t$,

DGP VI $x_t = 0.6x_{t-1} - 0.36x_{t-2} + \varepsilon_t$,

DGP VII $x_t = 0.9x_{t-1} - 0.81x_{t-2} + 0.91x_{t-3} + \varepsilon_t$,

DGP VIII $x_t = 0.9x_{t-1} - 0.56x_{t-2} + 0.66x_{t-3} + \varepsilon_t$,

where ε_t's are independent and identically $\mathcal{N}(0, 25)$ distributed. We aim to select two-step ($h = 2$) predictors for DGPs I-IV and three-step ($h = 3$) predictors for DGPs V-VIII using Procedure II with $C_n = \log n/n, 2\log n/n$, and $3\log n/n$, which will be referred to as Procedures A, B, and C, respectively. The candidate predictors are set to $(i, j), i = 1, \cdots, 10$ and $j = 1, 2$. According to Section 2 and Section 2 of Ing (2004), the asymptotically optimal multistep predictors (or the order/method combinations with the minimal loss function values) for DGPs I-VIII are listed in Table II. We generated 100 replications for each of these DGPs and carried out predictor selection for each replication. The frequency of these combinations selected by Procedures A, B, and C are shown in Table III for n = 150, 300, 500, 1000, and 2000.

The simulation results are summarized as follows.
(1) Two-step predictions. Procedures A, B, and C can efficiently select the best order/method combination (listed in Table II) regardless of whether the DGP is stationary or nonstationary. (Note that DGPs I and II are stationary, but DGPs III and IV are not.) In particular, the proportion of the best combination selected by Procedure C often exceeds 95 percent even when $n = 150$. Note that while the differences between the parameter values of DGPs I and II (or III and IV) are not sizable, different order/method combinations are required

TABLE II

OREDER/METHOD COMBINATION WITH THE MINIMAL LOSS FUNCTION VALUE

	h=2				h=3			
DGP	I	II	III	IV	V	VI	VII	VIII
Combination	(1,2)	(2,1)	(2,2)	(3,1)	(1,2)	(2,1)	(2,2)	(3,1)

to attain the minimal loss function value (defined in (3.1) and (3.2)). Table III shows that these procedures are sensitive to small parameter changes and can efficiently switch to the ”right track”. However, we also notice that the finite sample performances of Procedure A seem to be slightly worse than those of Procedures B and C.

(2) Three-step predictions. Note that DGPs V and VI are stationary AR(2) models with AR coefficients satisfying $0 < a_1 < 1$ and $a_1^2 + a_2 = 0$. Ing (2004, Section 2) recently showed that $(1, 2)$ is asymptotically more efficient than $(2, 1)$ in DGP V, whereas $(2, 1)$ is asymptotically more efficient than $(1, 2)$ in DGP VI. Procedures A, B, and C perform quite well in this subtle case. More specifically, for $(a_1, a_2) = (0.9, -0.81)$, they can correctly choose (1, 2) over 90 percent of the time for all sample sizes (except for Procedure A in the sample size of 150). On the other hand, when $(a_1, a_2) = (0.6, -0.36)$, Procedures B and C successfully select another combination, (2, 1), with rather high frequency for $n \geq 300$. While Procedure A performs slightly worse than the other two procedures, it can still choose (2,1) with over 90 percent frequency as $n \geq 1000$. DGPs VII and VIII are unit root processes. In DGP VII, the direct method only requires two regressors to perform three-step predictions, and according to Section 2, (2, 2) can attain the minimal loss function value. On the other hand, (3, 1) is the best combination for DGP VIII. Table III shows that the performances of Procedures A, B, and C in DGPs VII and VIII are similar to theirs in DGPs V and VI.

Finally, we note that the choice of C_n in Procedure II does influence its finite-sample results. While we do not intend to suggest the best C_n in finite-sample cases, the C_n's used in this paper may serve as good ”initial values” for pursuing better performances based on Procedure II.

5. Conclusion

This paper provides asymptotic expressions for the MSPEs (up to terms of order $1/n$) and the multistep generalizations of APE (up to terms of order $\log n$) for the plug-in and direct predictors in correctly specified AR processes with unit roots. These asymptotic results, which are new to the literature, enable us to show that Ing's (2004) predictor selection procedure is asymptotically efficient regardless of whether there exist unit roots or not. The finite-sample performances of Ing's procedure, however, are not as good as expected. Inspired by the asymptotic analysis, we have constructed a new predictor selection criterion which shares the same asymptotic properties as Ing's criterion, but also has satisfactory finite sample performances as illustrated by various simulation experiments. As a final remark, we believe Procedures I and II are also asymptotically efficient in AR processes with several unit roots (e.g., Chan and Wei (1988)). Further investigation to verify this conjecture would be of interest.

APPENDIX A

TABLE III

FREQUENCY OF CHOOSING PREDICTORS WITH MINIMAL LOSS FUNCTION VALUES IN 100 REPLICATIONS

		$h = 2$				$h = 3$		
		Procedure				Procedure		
n	Model(Unit Root)	A	B	C	Model(Unit Root)	A	B	C
150	I(No)	86	94	96	V(No)	83	92	95
300		88	94	99		90	96	99
500		95	99	100		91	98	98
1000		93	96	99		93	97	98
2000		93	98	99		97	100	100
150	II(No)	84	92	94	VI(No)	67	75	73
300		84	89	95		84	90	96
500		88	95	98		86	97	100
1000		91	95	99		90	98	100
2000		94	99	99		93	98	99
150	III(Yes)	85	94	97	VII(Yes)	85	93	95
300		85	92	97		87	98	98
500		91	95	99		91	96	99
1000		97	100	100		91	97	99
2000		94	99	100		94	97	99
150	IV(Yes)	80	87	91	VIII(Yes)	62	77	75
300		87	96	97		83	91	91
500		90	96	100		86	94	98
1000		90	94	97		86	97	98
2000		90	96	100		93	99	100

Throughout this section, we only consider the case $k \geq 2$ (recall that k is the order of the AR model under consideration), since the results for the case $k = 1$ can be verified similarly. We start with some useful lemmas.

Lemma A.1. *Assume that $\{x_t\}$ satisfies model (1.1) with $\{\varepsilon_t\}$ obeying (2.1). Then, for any $q > 0$ and $k \geq p_1$,*

$$E\|\hat{R}_n^{-1}(k)\|^q = O(1), \tag{A.1}$$

where $\hat{R}_n(k)$ is defined after (2.2) and for a matrix A, $\|A\|^2 = \sup_{\|\mathbf{z}\|=1} \mathbf{z}'A'A\mathbf{z}$ with $\|\mathbf{z}\|$ denoting Euclidean norm for vector $\mathbf{z}$.

PROOF. (A.1) can be verified by an argument similar to that used in the proof of Lemma 1 in Ing and Sin (2006). The details are omitted. □

Lemma A.2. *Assume that $\{x_t\}$ satisfies model (1.1) with $\{\varepsilon_t\}$ obeying (2.1) and for some $q_1 \geq 2$, $\sup_{-\infty<t<\infty} E|\varepsilon_t|^{2q_1} < \infty$. Then, for any $0 < q < q_1$ and $k \geq p + 1$,*

$$E\|\hat{R}_n^{-1}(k) - \hat{R}_n^{*^{-1}}(k)\|^q = O(n^{-q/2}), \tag{A.2}$$

where

$$\hat{R}_n^*(k) = \begin{pmatrix} \hat{\Gamma}_n(k-1) & \mathbf{0}'_{k-1} \\ \mathbf{0}'_{k-1} & \frac{1}{n^2}\sum_{j=k}^{n-1} N_j^2 \end{pmatrix},$$

$\hat{\Gamma}_n(k-1) = (1/n)\sum_{j=k}^{n-1} \mathbf{s}_j(k-1)\mathbf{s}'_j(k-1)$, *and* $N_j = x_j - \sum_{l=0}^{k-1} \alpha_j x_{j-l}$.

PROOF. First note that Lemma 1 ensures for any $q > 0$,

$$E\|\hat{R}_n^{*^{-1}}(k)\|^q = O(1). \tag{A.3}$$

We also have

$$\begin{aligned} &\|\hat{R}_n^{-1}(k) - \hat{R}_n^{*^{-1}}(k)\|^q \leq \|\hat{R}_n^{-1}(k)\|^q \|\hat{R}_n^{*^{-1}}(k)\|^q \|\hat{R}_n(k) - \hat{R}_n^*(k)\|^q \\ \leq\ & C_1 \|\hat{R}_n^{-1}(k)\|^q \|\hat{R}_n^{*^{-1}}(k)\|^q \|n^{-3/2}\sum_{j=k}^{n-1} \mathbf{s}_j(k-1)N_j\|^q, \end{aligned} \tag{A.4}$$

where C_1 is some positive constant. By analogy with Lemma 2 in Ing and Sin (2006),

$$E\|n^{-3/2}\sum_{j=k}^{n-1} \mathbf{s}_j(k-1)N_j\|^{q_1} = O(n^{-q_1/2}). \tag{A.5}$$

Consequently, (A.2) follows from (A.1), (A.3)-(A.5), and Hölder's inequality. □

Lemma A.3. *Assume that* $\{x_t\}$ *satisfies model (1.1) with* $\sup_{-\infty<t<\infty} E|\varepsilon_t|^q < \infty$, *where* $q \geq 2$. *Then, for* $k \geq p+1$,

$$E\|n^{-1/2} D_n \sum_{j=k}^{n-1} \mathbf{x}_j(k)\varepsilon_{j+1}\|^q \quad = \quad O(1). \tag{A.6}$$

PROOF. (A.6) can be shown by an argument similar to that used in the proof in Ing and Wei (2003, Lemma 4). We skip the details. □

Lemma A.4. *Assume that* $\{x_t\}$ *satisfies model (1.1) with* $\sup_{-\infty<t<\infty} E|\varepsilon_t|^r < \infty$, *for some* $r > 4$. *Then, for* $k \geq p_1$,

$$\lim_{n\to\infty} E(F_{n,k}) = 0, \tag{A.7}$$

where

$$F_{n,k} = \frac{\mathbf{s}_n(k-1)M_h(k-1)\hat{\Gamma}_n^{-1}(k-1)\{\sum_{j=k}^{n-1}\mathbf{s}_j(k-1)\varepsilon_{j+1}\}N_n \sum_{j=k}^{n-1} N_j\varepsilon_{j+1}}{\sum_{j=k}^{n-1} N_j^2}. \tag{A.8}$$

PROOF. Omitted.

PROOF OF THEOREM 2.2. Some algebraic manipulations give

$$x_{n+h} - \hat{x}_{n+h}(k) = \eta_{n,h} - \mathbf{x}_n'(k)\hat{L}_{n,h}(k)(\hat{\mathbf{a}}_n(1,k) - \mathbf{a}(k)), \tag{A.9}$$

where $\hat{L}_{n,h}(k)$ is defined after (3.4). We also have

$$\begin{aligned} & nE\left\{\mathbf{x}_n'(k)(\hat{L}_{n,h}(k) - L_h(k))(\hat{\mathbf{a}}_n(1,k) - \mathbf{a}(k))\right\}^2 \\ = \quad & E\left\{\mathbf{x}_n'(k)(\hat{L}_{n,h}(k) - L_h(k))D_n'(k)\hat{R}_n^{-1}(k)\frac{1}{\sqrt{n}}D_n(k)\sum_{j=k}^{n-1}\mathbf{x}_j(k)\varepsilon_{j+1}\right\}^2 \\ \equiv \quad & E\{G_n^2(k)\}, \end{aligned} \tag{A.10}$$

where $L_h(k) = \sum_{j=0}^{h-1} b_j A^{h-1-j}(k)$, with $A(k)$ defined in Section 1. Let

$$\hat{D}_n(k) = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & -1 \\ \frac{1}{\sqrt{n}} & \frac{-\hat{\alpha}(n,1)}{\sqrt{n}} & \dots & \dots & \frac{-\hat{\alpha}(n,k-1)}{\sqrt{n}} \end{pmatrix},$$

where $(\hat{\alpha}(n,1), \cdots, \hat{\alpha}(n,k-1))' = \hat{\alpha}_n(k-1) = \Psi(k)\hat{\mathbf{a}}_n(1,k)$, and $\Psi(k)$ is a $(k-1)\times k$ matrix, with the (i,j)th component equal to 0 if $j \le i$ and equal to -1 if $j > i$. By observing $A(k)D_n'(k) = D_n'(k)\bar{A}(k)$ and $\hat{A}_n(k)\hat{D}_n'(k) = \hat{D}_n'(k)A_n^*(k)$, where

$$\bar{A}(k) = \begin{pmatrix} S_M(k-1) & \mathbf{0}_{k-1} \\ \mathbf{0}_{k-1}' & 1 \end{pmatrix},$$

$$A_n^*(k) = \begin{pmatrix} \hat{S}_{M,n}(k-1) & \mathbf{0}_{k-1} \\ \mathbf{0}_{k-1}' & 1 \end{pmatrix},$$

and

$$\hat{S}_{M,n}(k-1) = \left(\hat{\alpha}_n(k-1) \;\middle|\; \begin{array}{c} I_{k-2} \\ \hline \mathbf{0}_{k-2}' \end{array} \right),$$

we have

$$L_h(k)D_n'(k) = D_n'(k)\bar{L}_h(k), \tag{A.11}$$

and

$$\hat{L}_{n,h}(k)\hat{D}_n'(k) = \hat{D}_n'(k)L_{n,h}^*(k), \tag{A.12}$$

where

$$\begin{aligned} \bar{L}_h(k) &= \begin{pmatrix} M_h(k-1) & \mathbf{0}_{k-1} \\ \mathbf{0}_{k-1}' & \sum_{j=0}^{h-1} b_j \end{pmatrix}, \\ L_{n,h}^*(k) &= \begin{pmatrix} \hat{M}_{n,h}(k-1) & \mathbf{0}_{k-1} \\ \mathbf{0}_{k-1}' & \sum_{j=0}^{h-1} b_j \end{pmatrix}, \end{aligned}$$

and $\hat{M}_{n,h}(k-1) = \sum_{j=0}^{h-1} b_j \hat{S}_{M,n}^{h-1-j}(k-1)$, with $\hat{S}_{M,n}^0(k-1) = I_{k-1}$. (A.11) and (A.12) yield

$$\begin{aligned}(\hat{L}_{n,h}(k) - L_h(k))D_n'(k) &= \hat{L}_{n,h}(k)(D_n'(k) - \hat{D}_n'(k)) \\ + \quad & (\hat{D}_n'(k) - D_n'(k))L_{n,h}^*(k) + D_n'(k)(L_{n,h}^*(k) - \bar{L}_h(k)),\end{aligned}$$

and hence

$$|G_n(k)| \le G_n^*(k), \tag{A.13}$$

where $G_n^*(k) = (I) + (II)$, with

$$\begin{aligned}(I) &= \|n^{-1/2}\mathbf{x}_n(k)\|\|\hat{\alpha}_n(k-1) - \alpha(k-1)\|(\|\hat{L}_{n,h}(k)\| + \|L_{n,h}^*(k)\|)G_{1,n}^*(k), \\ (II) &= (\|\mathbf{s}_n(k-1)\| + |n^{-1/2}N_n|)\|L_{n,h}^*(k) - \bar{L}_h(k)\|G_{1,n}^*(k),\end{aligned}$$

and $G_{1,n}^*(k) = \|\hat{R}_n^{-1}(k)\|\|n^{-1/2}D_n(k)\sum_{j=k}^{n-1}\mathbf{x}_j(k)\varepsilon_{j+1}\|$. By (A.10), (A.13), Lemmas 1-3, and Hölder's inequality, it can be shown that

$$nE\left\{\mathbf{x}_n'(k)(\hat{L}_{n,h}(k) - L_h(k))\left(\hat{\mathbf{a}}_n(1,k) - \mathbf{a}(k)\right)\right\}^2 = E(G_n^*(k))^2 = O(n^{-1}). \tag{A.14}$$

Similarly, we have

$$E\left\{\mathbf{x}_n'(k)L_h(k)D_n'(k)(\hat{R}_n^{-1}(k) - \hat{R}_n^{*^{-1}}(k))\frac{1}{\sqrt{n}}D_n(k)\sum_{j=k}^{n-1}\mathbf{x}_j(k)\varepsilon_{j+1}\right\}^2 = O(n^{-1}). \tag{A.15}$$

By (A.11) and some algebraic manipulations,

$$\begin{aligned}& E\left\{\mathbf{x}_n'(k)L_h(k)D_n'(k)\hat{R}_n^{*^{-1}}(k)\frac{1}{\sqrt{n}}D_n(k)\sum_{j=k}^{n-1}\mathbf{x}_j(k)\varepsilon_{j+1}\right\}^2 \\ =\ & E_{1,n}(k) + E_{2,n}(k) + E_{3,n}(k),\end{aligned} \tag{A.16}$$

where

$$\begin{aligned}E_{1,n}(k) &= E\left\{\mathbf{s}_n'(k-1)M_h(k-1)\hat{\Gamma}_n^{-1}(k-1)n^{-1/2}\sum_{j=k}^{n-1}\mathbf{s}_j(k-1)\varepsilon_{j+1}\right\}^2, \\ E_{2,n}(k) &= (\sum_{j=0}^{h-1} b_j)^2 E\left\{n\frac{N_n^2(\sum_{j=k}^{n-1}N_j\varepsilon_{j+1})^2}{(\sum_{j=k}^{n-1}N_j^2)^2}\right\}, \\ E_{3,n}(k) &= 2(\sum_{j=0}^{h-1} b_j)E(F_{n,k}),\end{aligned}$$

By an analogy with Ing (2003, Theorem 1),

$$\lim_{n\to\infty} E_{1,n}(k) = f_{1,h}(k-1). \tag{A.17}$$

In view of Ing (2001), it is straightforward to show that

$$\lim_{n\to\infty} E_{2,n}(k) = 2\sigma^2(\sum_{j=0}^{h-1} b_j)^2. \tag{A.18}$$

Consequently, the desired result follows from (A.9), (A.14)-(A.18), and Lemma A.4. □

PROOF OF THEOREM 2.3. By analogies with Lemmas A.1-A.4, for $k \geq p_h$,

$$E\|\bar{R}_{n,h}^{-1}(k)\|^q = O(1), \tag{A.19}$$

$$E\|\bar{R}_{n,h}^{-1}(k) - \bar{R}_{n,h}^{*-1}(k)\|^4 = O(n^{-2}), \tag{A.20}$$

$$E\|n^{-1/2}\bar{D}_n \sum_{j=k}^{n-1} \mathbf{x}_j(k)\eta_{j+h}\|^8 = O(1), \tag{A.21}$$

and

$$\lim_{n\to\infty} E(\bar{F}_{n,k}) = 0, \tag{A.22}$$

where $q > 0$,

$$\bar{R}_{n,h}^{*}(k) \quad = \quad \begin{pmatrix} \frac{1}{n}\sum_{j=k}^{n-h} \mathbf{s}_j(k-1)\mathbf{s}_j'(k-1) & \mathbf{0} \\ \mathbf{0}' & \frac{1}{n^2}\sum_{j=k}^{n-h} x_j^2 \end{pmatrix},$$

and

$$\bar{F}_{n,k} = \frac{\mathbf{s}_n(k-1)\{\frac{1}{n}\sum_{j=k}^{n-h} \mathbf{s}_j(k-1)\mathbf{s}_j'(k-1)\}^{-1}\{\sum_{j=k}^{n-h} \mathbf{s}_j(k-1)\eta_{j,h}\}x_n \sum_{j=k}^{n-h} x_j\eta_{j,h}}{\sum_{j=k}^{n-h} x_j^2}. \tag{A.23}$$

In addition, according to (1.9) and (3.5) of Ing and Sin (2007), it can be shown that

$$\lim_{n\to\infty} E\left\{n\frac{x_n^2(\sum_{j=k}^{n-h} x_j\eta_{j,h})^2}{\sum_{j=k}^{n-h} x_j^2}\right\} = 2\sigma^2(\sum_{j=0}^{h-1} b_j)^2. \tag{A.24}$$

As a result, Theorem 2.2 follows from (A.19)-(A.22), (A.24), and arguments similar to those used in the proofs of Theorem 2 in Ing (2003) and Theorem 2.1 above. □

APPENDIX B

Lemma B.1 below provides (almost sure) asymptotic bounds for $\|\hat{\Gamma}_n(k-1)-\Gamma(k-1)\|$, $\|\hat{R}_n(k)-\hat{R}_n^*(k)\|$, and $\|\hat{R}_n^{-1}(k)\|$ under a minimal moment condition, $\sup_{-\infty<t<\infty} E|\varepsilon_t|^\alpha$ for some $\alpha > 2$. As will be seen later, these bounds play subtle roles in our asymptotic analysis.

Lemma B.1. *Assume that the assumptions of Theorem 3.1 holds. Then,*

(i) for $k \geq 2$, $k \geq p_1$, and some $\iota > 0$,

$$\|\hat{\Gamma}_n(k-1)-\Gamma(k-1)\| = o(n^{-\iota}) \ a.s; \tag{B.1}$$

(ii) for $k \geq p_1$ and some $\eta > 0$,

$$\|\hat{R}_n(k)-\hat{R}_n^*(k)\| = o(n^{-\eta}) \ a.s; \tag{B.2}$$

(iii) for $k \geq p_1$,

$$\|\hat{R}_n^{-1}(k)\| = O(\log\log n) \ a.s. \tag{B.3}$$

PROOF. First note that

$$\|\hat{\Gamma}_n(k-1)-\Gamma(k-1)\| \leq \sum_{l=0}^{k-2}\sum_{m=0}^{k-2} |n^{-1}\sum_{j=k}^{n-1} s_{j-l}s_{j-m}-\gamma_{l,m}|,$$

where $\gamma_{l,m}$ is the (l,m)-th component of $\Gamma(k-1)$. Therefore, (B.1) is ensured by showing that for any $1 \leq l \leq k-1$ and $1 \leq m \leq k-1$,

$$|\frac{1}{n}\sum_{j=k}^{n-1} s_{j-l}s_{j-m}-\gamma_{l,m}| = o(n^{-\iota}) \ a.s. \tag{B.4}$$

In the following, we only prove the case of $l = m = 0$ since the proofs of other cases can be similarly obtained. For $l = m = 0$, the left-hand side of (B.4) can be rewritten as

$$\left| \frac{1}{n} \sum_{j=k}^{n-1} (s_j^2 - \gamma_{0,0}^{(j)}) + \frac{1}{n} \sum_{j=k}^{n-1} (\gamma_{0,0}^{(j)} - \gamma_{0,0}) + \frac{k\gamma_{0,0}}{n} \right|, \tag{B.5}$$

where $\gamma_{0,0}^{(j)} = \sigma^2 \sum_{r=0}^{j-1} c_r^2$ with c_j's defined in Section 1. By observing $\gamma_{0,0} = \sigma^2 \sum_{r=0}^{\infty} c_r^2$ and $|c_r| \leq C_1 e^{-\beta_1 r}$ for all r and some $C_1, \beta_1 > 0$, we have $(1/n) \sum_{j=k}^{n-1} (\gamma_{0,0}^{(j)} - \gamma_{0,0}) = O(1/n)$ and $k\gamma_{0,0}/n = O(1/n)$. In addition, straightforward calculations yield that

$$s_j^2 - \gamma_{0,0}^{(j)} = \sum_{l=1}^{j} c_{j-l}^2 (\varepsilon_l^2 - \sigma^2) + 2 \sum_{l_2=2}^{j} \sum_{l_1=1}^{l_2-1} c_{j-l_1} c_{j-l_2} \varepsilon_{l_1} \varepsilon_{l_2}. \tag{B.6}$$

In view of (B.6), one obtains, through changing the order of summations, that

$$\begin{aligned} \sum_{j=n_1}^{n_2} \frac{s_j^2 - \gamma_{0,0}^{(j)}}{j^\theta} &= \sum_{l=1}^{n_1} (\sum_{j=n_1}^{n_2} \frac{c_{j-l}^2}{j^\theta}) \eta_l + \sum_{l=n_1+1}^{n_2} (\sum_{j=l}^{n_2} \frac{c_{j-l}^2}{j^\theta}) \eta_l + 2 \sum_{l_2=2}^{n_1} \{ \sum_{l_1=1}^{l_2-1} (\sum_{j=n_1}^{n_2} \frac{c_{j-l_1} c_{j-l_2}}{j^\theta}) \varepsilon_{l_1} \} \varepsilon_{l_2} \\ &+ 2 \sum_{l_2=n_1+1}^{n_2} \{ \sum_{l_1=1}^{l_2-1} (\sum_{j=l_2}^{n_2} \frac{c_{j-l_1} c_{j-l_2}}{j^\theta}) \varepsilon_{l_1} \} \varepsilon_{l_2} \equiv (I) + (II) + (III) + (IV), \end{aligned}$$

where $\eta_t = \varepsilon_t^2 - \sigma^2$, $\theta < 1$ and $\theta\alpha/2 > 1$. If we can show that for any $1 \leq n_1 \leq n_2 < \infty$,

$$E|(G)|^{\alpha/2} \leq C(\sum_{j=n_1}^{n_2} \frac{1}{j^{\xi_1}})^{\xi_2}, \tag{B.7}$$

where $G = I, II, III$ and IV; and $C > 0, \xi_1 > 1, \xi_2 > 1$ are some positive constant independent of n_1 and n_2 (but they can vary with G), then by Móricz (1976) (see, also, Ing and Wei (2006)), for all sufficiently large n_1,

$$E \max_{n_1 \leq l \leq n_2} |\sum_{j=n_1}^{l} \frac{s_j^2 - \gamma_{0,0}^{(j)}}{j^\theta}|^{\alpha/2} \leq C^*(\sum_{j=n_1}^{n_2} \frac{1}{j^{\xi_1^*}})^{\xi_2^*}, \tag{B.8}$$

where $C^* > 0$, $\xi_1^* > 1$ and $\xi_2^* > 1$ are some positive constant independent of n_1 and n_2. (B.8) and Kronecker's lemma yield

$$\frac{1}{n^\theta} \sum_{j=1}^{n} (s_j^2 - \gamma_{0,0}^{(j)}) = o(1) \ a.s. \tag{B.9}$$

As a result, (B.1) holds with $\iota = 1 - \theta$.

Without loss of generality, assume $2 < \alpha < 4$. Then,

$$\begin{aligned}
E|(I)|^{\alpha/2} &\leq C_2 E\{\sum_{l=1}^{n_1}(\sum_{j=n_1}^{n_2} \frac{c_{j-l}^2}{j^\theta})^2 \eta_l^2\}^{\alpha/4} \\
&\leq C_2 \sum_{j_1=n_1}^{n_2} \sum_{j_2=n_1}^{n_2} \frac{1}{j_1^{\theta\alpha/4} j_2^{\theta\alpha/4}} \sum_{l=1}^{n_1} |c_{j_1-l}c_{j_2-l}|^{\alpha/2} E|\eta_l|^{\alpha/2} \\
&\leq C_3 \left(\sum_{j=n_1}^{n_2} \frac{1}{j^{\theta\alpha/2}} + \sum_{j_1=n_1}^{n_2-1} \frac{1}{j_1^{\theta\alpha/4}} \sum_{j_2=j_1+1}^{n_2} \frac{1}{j_2^{\theta\alpha/4}} (j_2 - j_1)^{-s} \right) \\
&\leq C_4 \left(\sum_{j=n_1}^{n_2} \frac{1}{j^{\theta\alpha/2}} \right) \leq C_4 \left(\sum_{j=n_1}^{n_2} \frac{1}{j^{\xi_1}} \right)^{\xi_2}, \qquad \text{(B.10)}
\end{aligned}$$

where $C_i > 0, i = 2, \cdots, 4$, and $s > 1$ are some positive constant independent of n_1 and n_2, $1 < \xi_1 < \theta\alpha/2$, $\xi_2 = \theta\alpha/2\xi_1$, the first inequality follows from Burkholder's inequality, the second one follows from the fact that $\alpha/4 < 1$ and changing the order of summations, the third one is ensured by $\sup_t E|\varepsilon_t|^\alpha < \infty$ and $c_j \leq C_1 e^{-\beta_1 j}$, which implies for all $n_1 \leq j_1 \neq j_2 \leq n_2$, $\sum_{l=1}^{n_1} |c_{j_1-l}c_{j_2-l}|^{\alpha/2} \leq C_5 |j_1 - j_2|^{-s}$, for some $C_5 > 0$. As a result, (B.7) holds with $G = I$. The proof of (B.7) for the case of $G = II$ is similar. The details are omitted. To show (B.7) for the case of $G = III$, note that

$$\begin{aligned}
& E|\sum_{l_2=2}^{n_1}\{\sum_{l_1=1}^{l_2-1}(\sum_{j=n_1}^{n_2} \frac{c_{j-l_1}c_{j-l_2}}{j^\theta})\varepsilon_{l_1}\}\varepsilon_{l_2}|^{\alpha/2} \\
\leq\ & \{E(\sum_{l_2=2}^{n_1}\{\sum_{l_1=1}^{l_2-1}(\sum_{j=n_1}^{n_2} \frac{c_{j-l_1}c_{j-l_2}}{j^\theta})\varepsilon_{l_1}\}\varepsilon_{l_2})^2\}^{\alpha/4} \\
=\ & |\sigma|^\alpha \left(\sum_{l_2=2}^{n_1} \sum_{l_1=1}^{l_2-1} (\sum_{j=n_1}^{n_2} \frac{c_{j-l_1}c_{j-l_2}}{j^\theta})^2 \right)^{\alpha/4}. \qquad \text{(B.11)}
\end{aligned}$$

By arguments similar to those used to verify the second to fifth inequalities in (B.10), the desired result follows. Similarly, it can be shown that (B.7) holds for the case of $G = IV$.

To show (B.2), first observe that

$$\|\hat{R}_n(k) - \hat{R}_n^*(k)\| \leq \sqrt{2} \sum_{l=0}^{k-2} |\frac{1}{n^{3/2}} \sum_{j=k}^{n-1} s_{j-l} N_j|.$$

Therefore, it suffices to show that for $l = 0, \cdots, k-2$ and some $\eta > 0$,

$$\frac{1}{n^{3/2}} \sum_{j=k}^{n-1} s_{j-l} N_j = o(n^{-\eta}) \ a.s. \qquad \text{(B.12)}$$

We only verify (B.12) for the case $l = 0$ since the proof of the case $l > 0$ can be similarly obtained. Let $\max\{1, (1/2) + (2/\alpha)\} < \theta_1 < 3/2$. Some algebraic manipulations yield

$$\begin{aligned}
\sum_{j=n_1}^{n_2} \frac{s_j N_j}{j^{\theta_1}} &= \sigma^2 \sum_{j=n_1}^{n_2} \frac{1}{j^{\theta_1}} \sum_{m=1}^{j} c_{j-m} + \sum_{m=1}^{n_1} \sum_{j=n_1}^{n_2} \frac{c_{j-m}}{j^{\theta_1}} \eta_m + \sum_{m=n_1+1}^{n_2} \sum_{j=m}^{n_2} \frac{c_{j-m}}{j^{\theta_1}} \eta_m \\
&+ \sum_{m=2}^{n_1} \sum_{l=1}^{m-1} \sum_{j=n_1}^{n_2} \frac{c_{j-l}}{j^{\theta_1}} \varepsilon_l \varepsilon_m + \sum_{m=n_1+1}^{n_2} \sum_{l=1}^{m-1} \sum_{j=m}^{n_2} \frac{c_{j-l}}{j^{\theta_1}} \varepsilon_l \varepsilon_m \\
&+ \sum_{l=2}^{n_1} \sum_{m=1}^{l-1} \sum_{j=n_1}^{n_2} \frac{c_{j-l}}{j^{\theta_1}} \varepsilon_m \varepsilon_l + \sum_{l=n_1+1}^{n_2} \sum_{m=1}^{l-1} \sum_{j=l}^{n_2} \frac{c_{j-l}}{j^{\theta_1}} \varepsilon_m \varepsilon_l \\
&= (\ddot{I}) + (\ddot{II}) + (I\ddot{I}I) + (I\ddot{V}) + (\ddot{V}) + (\ddot{V}I) + (V\ddot{I}I). \qquad \text{(B.13)}
\end{aligned}$$

It is clear that

$$|(\ddot{I})|^{\alpha/2} \leq C_6 (\sum_{j=n_1}^{n_2} \frac{1}{j^{\xi_1}})^{\xi_2}, \qquad \text{(B.14)}$$

where $\xi_1 = \theta_1$ and $\xi_2 = \alpha/2$. By an argument similar to that used in (B.10),

$$E|(W)|^{\alpha/2} \leq C_7 (\sum_{j=n_1}^{n_2} \frac{1}{j^{\xi_1}})^{\xi_2}, \qquad \text{(B.15)}$$

where $W = \ddot{II}, I\ddot{I}I$, $1 < \xi_1 < \theta_1\alpha/2$, and $\xi_2 = \theta_1\alpha/2\xi_1$. An argument similar to that used in (B.11) yields

$$E|(W)|^{\alpha/2} \leq C_8 (\sum_{j=n_1}^{n_2} \frac{1}{j^{\xi_1}})^{\xi_2}, \qquad \text{(B.16)}$$

where $W = I\ddot{V}, \ddot{V}, \ddot{V}I, V\ddot{I}I$ and $1 < \xi_1 < (2\theta_1 - 1)\alpha/4$ and $\xi_2 = (2\theta_1 - 1)\alpha/4\xi_1$. Consequently, (B.12) (with $\eta = (3/2) - \theta_1$) follows from (B.13)-(B.16), Móricz (1976), and Kronecker's lemma.

To show (B.3), observe that $\|\hat{R}_n^{-1}(k)\| \leq \|\hat{R}_n^{-1}(k)\| \|\hat{R}_n(k) - \hat{R}_n^*(k)\| \|\hat{R}_n^{*^{-1}}(k)\| + \|\hat{R}_n^{*^{-1}}(k)\|$. By Lai and Wei (1982, (3.23); 1983, (3.2)),

$$\|\hat{R}_n^{*^{-1}}(k)\| = O(\log\log n) \ a.s.$$

This and (B.2) yield (B.3). □

Remark 7. Since

$$\|\hat{\Gamma}_n^{-1}(k-1) - \Gamma^{-1}(k-1)\| \le \|\hat{\Gamma}_n^{-1}(k-1)\| \|\hat{\Gamma}_n(k-1) - \Gamma(k-1)\| \|\Gamma^{-1}(k-1)\|,$$

by Lai and Wei (1983, (3.2)) and (B.1), for some $\iota > 0$,

$$\|\hat{\Gamma}_n^{-1}(k-1) - \Gamma^{-1}(k-1)\| = o(n^{-\iota}) \ a.s.$$

Lemma B.2. *Assume that the assumptions of Theorem 3.1 hold. Then, for $k \ge 2$ and $k \ge p_1$,*

$$\sum_{i=m_h}^{n-h} F_{i,k} = o(n) \ a.s., \tag{B.17}$$

where $F_{i,k}$ is defined in Lemma A.4.

PROOF. We only prove (3.9) for $k = 2$ since the proof of the case $k > 2$ is similar. As noted in the proof of Lemma B.1, there are positive numbers C_1, β_1 such that $|c_j| \le C_1 e^{-\beta_1 j}$ for all $j \ge 0$. Define $s_i^* = \sum_{j=0}^{g_i} c_j \varepsilon_{i-j}$, where $g_i = \lfloor (1/\beta_1) \log i \rfloor + 1$ and $\lfloor a \rfloor$ denotes the largest integer $\le a$. Also define

$$\begin{aligned}
L_{1,i} &= \frac{1}{\sqrt{i}} \sum_{j=2}^{i-1} s_j \varepsilon_{j+1}, L_{1,i}^* = \frac{1}{\sqrt{i}} \sum_{j=2}^{i-Mg_i-1} s_j \varepsilon_{j+1}, \\
L_{2,i} &= \frac{1}{\sqrt{i}} N_i, L_{2,i}^* = \frac{1}{\sqrt{i}} \sum_{j=1}^{i-Mg_i} \varepsilon_j, \\
L_{3,i} &= \frac{1}{i} \sum_{j=2}^{i-1} N_j \varepsilon_{j+1}, L_{3,i}^* = \frac{1}{i} \sum_{j=2}^{i-Mg_i-1} N_j \varepsilon_{j+1}, \\
L_{4,i} &= \frac{i^2}{\sum_{j=2}^{i-1} N_j^2}, L_{4,i}^* = \frac{i^2}{\sum_{j=2}^{i-Mg_i} N_j^2},
\end{aligned}$$

where $M > 1$.

By (B.4), the Cauchy-Schwarz inequality, Burkholder's inequality, Minkowski's inequality, and the Borel-Cantelli lemma, it can be shown that, all with probability 1,

$$\sum_{i=k}^{n-1} |s_i| = O(n), s_n - s_n^* = o(n^{-1/2}), \ and \ L_{2,n} - L_{2,n}^* = o(n^{-\iota}), \tag{B.18}$$

where $\iota > 0$ is some positive number. By the first and second equalities in (B.18),

$$\sum_{i=k}^{n-1} |s_i^*| = O(n). \tag{B.19}$$

In addition, according to Wei (1987, Lemma 1), (3.3) of Lai and Wei (1983), the law of the iterated logarithm, and (3.23) of Lai and Wei (1982), we have, all with probability 1,

$$\begin{aligned} & L_{1,n} = o((\log n)^{\theta_2}), L_{1,n}^* = o((\log n)^{\theta_2}), L_{2,n} = O((\log\log n)^{1/2}), L_{2,n}^* = O((\log\log n)^{1/2}), \\ & L_{3,n} = o((\log n)^{\theta_2}(\log\log n)^{1/2}), L_{3,n}^* = o((\log n)^{\theta_2}(\log\log n)^{1/2}), L_{4,n} = O(\log\log n), \\ & L_{4,n}^* = O(\log\log n),\ and\ L_{4,n} - L_{4,n}^* = O(\log n(\log\log n)^3 n^{-1}), \end{aligned} \tag{B.20}$$

where $\theta_2 > 1/\alpha$. By an analogy with Ing and Wei (2003, Lemma 4) and the Borel-Cantelli lemma, we also have, both with probability 1,

$$L_{1,n} - L_{1,n}^* = o(n^{-\iota_1})\ and\ L_{3,n} - L_{3,n}^* = o(n^{-\iota_2}), \tag{B.21}$$

where ι_1 and ι_2 are some positive numbers. (B.18)-(B.21) and Kronecker's lemma imply

$$\begin{aligned} & \frac{1}{n}\sum_{i=m_h}^{n-h} \frac{\mathbf{s}_i(1)M_h(1)\hat{\Gamma}_i^{-1}(1)\{\sum_{j=2}^{i-1}\mathbf{s}_j(1)\varepsilon_{j+1}\}N_i\sum_{j=2}^{i-1}N_j\varepsilon_{j+1}}{\sum_{j=2}^{i-1}N_j^2} \\ = & \ (\frac{1}{n}\sum_{i=m_h}^{n-h} s_i^* G_i^*) + o(1)\ a.s., \end{aligned} \tag{B.22}$$

where $G_i^* = M_h(1)\Gamma^{-1}(1)\prod_{j=1}^4 L_{j,i}^*$.

Now,

$$\sum_{i=m_h}^{n-h} s_i^* G_i^* = \sum_{i=m_h}^{n-h} G_i^* \sum_{j=i-g_i}^{i} c_{i-j}\varepsilon_j = \sum_{j=m_h-g_{m_h}}^{n-h} M_j^*\varepsilon_j, \tag{B.23}$$

where $M_j^* = \sum_{i=j}^{\min\{f(j),n-h\}} G_i^* c_{i-j}$, $f(j)$ is the smallest integer i such that $i - g_i \geq j$, and G_i^* is set to zero for $i < m_h$. By taking M in $L_{l,j}^*, l = 1,2,3,4$ large enough, M_j^* is $\sigma(\varepsilon_{j-1}, \varepsilon_{j-2}, \cdots)$-measurable. Therefore, Lai and Wei's (1982, Lemma 2) strong law of large numbers for martingales implies

$$\sum_{j=m_h-g_{m_h}}^{n-h} M_j^*\varepsilon_j = o\{(\sum_{j=m_h-g_{m_h}}^{n-h} M_j^{*^2})^{1/2}[\log(\sum_{j=m_h-g_{m_h}}^{n-h} M_j^{*^2})]^{\iota_3}\} + O(1)\ a.s., \tag{B.24}$$

where $\iota_3 > 1/2$. In addition, (B.20) yields

$$M_n^* = o((\log n)^{2\theta_2}(\log\log n)^2) \ a.s. \tag{B.25}$$

Consequently, (B.17) follows from (B.22)-(B.25). $\square$

The following simple facts will be used in the proof of Theorem 3.1.

Lemma B.3. *Let $\{z_n\}$ be a sequence of real numbers.*

(i) If $z_n \geq 0$, $n^{-1}\sum_{j=1}^n z_j = O(1)$, and for some $\xi > 1$, $\liminf_{n\to\infty} \nu_n/n^\xi > 0$, then

$$\sum_{j=1}^n \frac{z_j}{\nu_j} = O(1).$$

(ii) If $n^{-1}\sum_{j=1}^n z_j = o(1)$, then

$$\sum_{j=1}^n \frac{z_j}{j} = o(\log n).$$

PROOF. Omitted.

PROOF OF THEOREM 3.1. We only prove the case $k \geq 2$ since the proof of the case $k = 1$ is similar. By Chow (1965) and an analogy with (3.8) of Ing (2004),

$$\begin{aligned} \text{APE}P_{n,h}(k) - \sum_{i=m_h}^{n-h}(\eta_{i,h})^2 &= \sum_{i=m_h}^{n-h}\left\{\mathbf{x}_i'(k)\hat{L}_{i,h}(k)\left(\hat{\mathbf{a}}_i(1,k) - \mathbf{a}(k)\right)\right\}^2(1+o(1)) \\ &+ O(1) \quad \text{a.s.} \end{aligned} \tag{B.26}$$

Straightforward calculations give

$$\begin{aligned} &\sum_{i=m_h}^{n-h}\left\{\mathbf{x}_i'(k)(\hat{L}_{i,h}(k) - L_h(k))\left(\hat{\mathbf{a}}_i(1,k) - \mathbf{a}(k)\right)\right\}^2 \\ = &\sum_{i=m_h}^{n-h}\frac{1}{i}\left\{\mathbf{x}_i'(k)(\hat{L}_{i,h}(k) - L_h(k))D_i'(k)\hat{R}_i^{-1}(k)\frac{1}{\sqrt{i}}D_i(k)\sum_{j=k}^{i-1}\mathbf{x}_j(k)\varepsilon_{j+1}\right\}^2. \end{aligned} \tag{B.27}$$

By Lai and Wei (1982) and (3.1) and (3.2) of Lai and Wei (1983), we have

$$\|\hat{\alpha}_n(k-1) - \alpha(k-1)\| \ = \ O\left((\frac{\log n}{n})^{1/2}\right) \ a.s., \tag{B.28}$$

$$\|L^*_{n,h}(k) - \bar{L}_h(k)\| = O\left((\frac{\log n}{n})^{1/2} \right) \; a.s., \tag{B.29}$$

$$\|\hat{L}_{n,h}(k)\| = O(1) \; a.s., \tag{B.30}$$

$$\|\mathbf{x}_n(k)/\sqrt{n}\| = O((\log\log n)^{1/2}) \; a.s. \tag{B.31}$$

In addition, by Wei (1987, Lemma 1), the law of the iterated logarithm, and (3.3) of Lai and Wei (1983),

$$\|\frac{1}{\sqrt{n}} D_n(k) \sum_{j=k}^{n-1} \mathbf{x}_j(k)\varepsilon_{j+1}\| = o((\log n)^{\delta}(\log\log n)^{1/2}) \; a.s., \tag{B.32}$$

where $\delta > 1/\alpha$. As a result, by (A.11), (A.12), (B.3), (B.27)-(B.32), and the fact that $N_n/\sqrt{n} = O((\log\log n)^{1/2})$ a.s., one obtains

$$\sum_{i=m_h}^{n-h} \left\{ \mathbf{x}'_i(k)(\hat{L}_{i,h}(k) - L_h(k))\,(\hat{\mathbf{a}}_i(1,k) - \mathbf{a}(k)) \right\}^2 = O(1) \; a.s. \tag{B.33}$$

Armed with (B.2), (B.3), and the fact that $\|\hat{R}_n^{*^{-1}}(k)\| = O(\log\log n)$ a.s. (which is given after (B.16)), it can be shown that

$$\|\hat{R}_n^{-1}(k) - \hat{R}_n^{*^{-1}}(k)\| = o(\frac{(\log\log n)^2}{n^{\eta}}) \; a.s., \tag{B.34}$$

where $\eta > 0$ is some positive constant. Since (A.11) yields for some $C_1 > 0$, $\|D_i(k)L'_h(k)\mathbf{x}_i(k)\| = \|\bar{L}'_h(k)D_i(k)\mathbf{x}_i(k)\| \le C_1(\|\mathbf{s}_i(k-1)\| + |N_i/\sqrt{i}|)$, we obtain

$$\begin{aligned}
&\sum_{i=m_h}^{n-h} \frac{1}{i} \left\{ \mathbf{x}'_i(k)L_h(k)D'_i(k)(\hat{R}_i^{-1}(k) - \hat{R}_i^{*^{-1}}(k))\frac{1}{\sqrt{i}}D_i(k)\sum_{j=k}^{i-1}\mathbf{x}_j(k)\varepsilon_{j+1} \right\}^2 \\
\le\; & C_2 \sum_{i=m_h}^{n-h} \frac{1}{i} \left\{ (\|\mathbf{s}_i(k-1)\| + |\frac{N_i}{\sqrt{i}}|)\|\hat{R}_i^{-1}(k) - \hat{R}_i^{*^{-1}}(k)\| \|\frac{1}{\sqrt{i}}D_i(k)\sum_{j=k}^{i-1}\mathbf{x}_j(k)\varepsilon_{j+1}\| \right\}^2 \\
=\; & O(1) \; a.s.,
\end{aligned} \tag{B.35}$$

where $C_2 > 0$ is some positive constant independent of n and the equality follows from (B.32), (B.34), $N_n/\sqrt{n} = O((\log\log n)^{1/2})$ a.s., $(1/n)\sum_{j=k}^{n-1}\|\mathbf{s}_j(k-1)\| = O(1)$ a.s., and (i) of Lemma B.3.

By (A.11) and some algebraic manipulations,

$$\begin{aligned} & \sum_{i=m_h}^{n-h} \frac{1}{i} \left\{ \mathbf{x}_i'(k) L_h(k) D_i'(k) \hat{R}_i^{*^{-1}}(k) \frac{1}{\sqrt{i}} D_i(k) \sum_{j=k}^{i-1} \mathbf{x}_j(k) \varepsilon_{j+1} \right\}^2 \\ = \;& (I) + (II) + (III), \end{aligned}$$

where

$$\begin{aligned} (I) &= \sum_{i=m_h}^{n-h} \left\{ \mathbf{s}_i'(k-1) M_h(k-1) \hat{\Gamma}_i^{-1}(k-1) \frac{1}{i} \sum_{j=k}^{i-1} \mathbf{s}_j(k-1) \varepsilon_{j+1} \right\}^2, \\ (II) &= (\sum_{j=0}^{h-1} b_j)^2 \sum_{i=m_h}^{n-h} \frac{N_i^2 (\sum_{j=k}^{i-1} N_j \varepsilon_{j+1})^2}{(\sum_{j=k}^{i-1} N_j^2)^2}, \\ (III) &= 2(\sum_{j=0}^{h-1} b_j) \sum_{i=m_h}^{n-h} \frac{F_{i,k}}{i}. \end{aligned}$$

According to (B.29) and analogies with (A.1) and Theorem 3.1 of Ing (2004),

$$\begin{aligned} (I) &= \sum_{i=m_h}^{n-h} \left\{ \mathbf{s}_i'(k-1) \hat{M}_{i,h}(k-1) \hat{\Gamma}_i^{-1}(k-1) \frac{1}{i} \sum_{j=k}^{i-1} \mathbf{s}_j(k-1) \varepsilon_{j+1} \right\}^2 + o(\log n) \ a.s. \\ &= f_{1,h}(k-1) \log n + o(\log n) \ a.s. \end{aligned}$$

By Wei (1987, Theorem 4),

$$(II) = 2(\sum_{j=0}^{h-1} b_j)^2 \sigma^2 \log n + o(\log n) \ a.s.$$

In view of Lemma B.2 and (ii) of Lemma B.3, one obtains

$$(III) = o(\log n) \ a.s.$$

As a result,

$$\begin{aligned} & \sum_{i=m_h}^{n-h} \frac{1}{i} \left\{ \mathbf{x}_i'(k) L_h(k) D_i'(k) \hat{R}_i^{*^{-1}}(k) \frac{1}{\sqrt{i}} D_i(k) \sum_{j=k}^{i-1} \mathbf{x}_j(k) \varepsilon_{j+1} \right\}^2 \\ = \;& \{2(\sum_{j=0}^{h-1} b_j)^2 \sigma^2 + f_{1,h}(k-1)\} \log n + o(\log n) \ a.s. \qquad \text{(B.36)} \end{aligned}$$

Consequently, (3.6) follows from (B.26), (B.33), (B.35), (B.36), and the Cauchy-Schwarz inequality. □

To analyze APE$D_{n,h}(k)$, Lemma B.4 is required.

Lemma B.4. *Let the assumptions of Theorem 3.1 hold. Then,*

$$\sum_{i=m_h}^{n-h} \frac{x_i^2(\sum_{j=k}^{i-h} x_j \eta_{j,h})^2}{(\sum_{j=k}^{i-h} x_j^2)^2} = 2(\sum_{j=0}^{h-1} b_j)^2 \sigma^2 \log n + o(\log n) \qquad a.s. \tag{B.37}$$

PROOF. Following arguments similar to those used in the proofs of Lemma 2 and Theorem 1 of Ing and Sin (2007), one obtains

$$\liminf_{n\to\infty} \frac{\log\log n}{n^2} \sum_{j=1}^{n-1} x_j^2 > 0 \text{ a.s.,} \tag{B.38}$$

and

$$x_n = O((n \log\log n)^{1/2}) \text{ a.s.} \tag{B.39}$$

By Borel-Cantelli lemma,

$$\varepsilon_n = o(n^{1/2}) \text{ a.s.} \tag{B.40}$$

In addition, it is not difficult to show that for $\theta > 1/2$ and $l \geq 1$,

$$\frac{1}{n^\theta} \sum_{j=1}^{n-l} \varepsilon_j \varepsilon_{j+l} = o(1) \text{ a.s.} \tag{B.41}$$

(B.38)-(B.41) together imply

$$\sum_{i=m_h}^{n-h} \frac{x_i^2(\sum_{j=k}^{i-h} x_j \eta_{j,h})^2}{(\sum_{j=k}^{i-h} x_j^2)^2} = (\sum_{j=0}^{h-1} b_j)^2 \sum_{i=m_h}^{n-h} \frac{x_i^2(\sum_{j=k}^{i-1} x_j \varepsilon_{j+1})^2}{(\sum_{j=k}^{i-1} x_j^2)^2} + O(1) \qquad \text{a.s.} \tag{B.42}$$

Consequently, (B.37) follows from (B.42) and the fact that

$$\sum_{i=m_h}^{n-h} \frac{x_i^2(\sum_{j=k}^{i-1} x_j \varepsilon_{j+1})^2}{(\sum_{j=k}^{i-1} x_j^2)^2} = 2\sigma^2 \log n + o(\log n) \qquad \text{a.s.,}$$

which is guaranteed by (2.15) of Ing and Sin (2007). □

PROOF OF THEOREM 3.2. We only prove the case of $k \geq 2$ since the proof of the case of $k = 1$ is similar. By the same reasoning as in (B.26), we have

$$\begin{aligned} \text{APE}D_{n,h}(k) - \sum_{i=m_h}^{n-h} \eta_{i,h}^2 = &= (1+o(1)) \sum_{i=m_h}^{n-h} \left\{ \mathbf{x}_i'(k) V_{i-h}(k) \left(\sum_{j=k}^{i-h} \mathbf{x}_j(k) \eta_{j,h} \right) \right\}^2 \\ &+ O(1) \quad \text{a.s.} \end{aligned} \quad \text{(B.43)}$$

Observe that

$$\begin{aligned} &\sum_{i=m_h}^{n-h} \left\{ \mathbf{x}_i'(k) V_{i-h}(k) \left(\sum_{j=k}^{i-h} \mathbf{x}_j(k) \eta_{j,h} \right) \right\}^2 \\ &= \sum_{i=m_h}^{n-h} \frac{1}{i} \left\{ \mathbf{x}_i'(k) \bar{D}_i'(k) \bar{R}_{i,h}^{-1}(k) \left(\frac{1}{\sqrt{i}} \bar{D}_i(k) \sum_{j=k}^{i-h} \mathbf{x}_j(k) \eta_{j,h} \right) \right\}^2 . \end{aligned} \quad \text{(B.44)}$$

According to (B.38), (B.39) and arguments similar to those used to obtain (B.32) and (B.34),

$$\left\| \frac{1}{\sqrt{n}} \bar{D}_n(k) \sum_{j=k}^{n-h} \mathbf{x}_j(k) \eta_{j,h} \right\| = o((\log n)^{\delta} (\log \log n)^{1/2}) \text{ a.s.},$$

and

$$\| \bar{R}_{n,h}^{-1}(k) - \bar{R}_{n,h}^{*^{-1}}(k) \| = O(n^{-\eta} (\log \log n)^2) \text{ a.s.},$$

where $\delta > 1/\alpha$ and $\eta > 0$. These facts and the reasoning similar to that used in (B.35) yield

$$\sum_{i=m_h}^{n-h} \frac{1}{i} \left\{ \mathbf{x}_i'(k) \bar{D}_i'(k) (\bar{R}_{i,h}^{-1}(k) - \bar{R}_{i,h}^{*^{-1}}(k)) \left(\frac{1}{\sqrt{i}} \bar{D}_i(k) \sum_{j=k}^{i-h} \mathbf{x}_j(k) \eta_{j,h} \right) \right\}^2 = O(1) \text{ a.s.} \quad \text{(B.45)}$$

Now,

$$\sum_{i=m_h}^{n-h} \frac{1}{i} \left\{ \mathbf{x}_i'(k) \bar{D}_i'(k) \bar{R}_{i,h}^{*^{-1}}(k) \left(\frac{1}{\sqrt{i}} \bar{D}_i(k) \sum_{j=k}^{i-h} \mathbf{x}_j(k) \eta_{j,h} \right) \right\}^2 = (I) + (II) + (III), \quad \text{(B.46)}$$

where

$$\begin{aligned} (I) &= \sum_{i=m_h}^{n-h} \{ \mathbf{s}_i'(k-1) [\sum_{j=k}^{i-h} \mathbf{s}_j(k-1) \mathbf{s}_j'(k-1)]^{-1} \sum_{j=k}^{i-h} \mathbf{s}_j(k-1) \eta_{j,h} \}^2, \\ (II) &= \sum_{i=m_h}^{n-h} \frac{x_i^2 (\sum_{j=k}^{i-h} x_j \eta_{j,h})^2}{(\sum_{j=k}^{i-h} x_j^2)^2}, \end{aligned}$$

and

$$(III) = \sum_{j=m_h}^{n-h} \frac{\bar{F}_i(k)}{i}$$

with $\bar{F}_i(k)$ defined in (A.23). By analogy with Theorem 3.2 of Ing (2004),

$$(I) = f_{2,h}(k-1)\log n + o(\log n) \qquad \text{a.s.} \tag{B.47}$$

According to Lemma B.4,

$$(II) = 2\sigma^2(\sum_{j=0}^{h-1} b_j)^2 \log n + o(\log n) \qquad \text{a.s.} \tag{B.48}$$

By reasoning similar to that used in the proof of Lemma B.2,

$$\sum_{j=m_h}^{n-h} \bar{F}_i(k) = o(n) \qquad \text{a.s.,}$$

and hence

$$(III) = o(\log n) \qquad \text{a.s.} \tag{B.49}$$

Consequently, (3.8) follows from (B.43)-(B.49). □

Lemma B.5. *Let the assumptions of Theorem 3.1 hold. Then, for* $1 \leq k < p_h$ *and* $h \geq 1$*,*

$$\sum_{i=m_h}^{n-h} [\mathbf{x}_i'(\check{\mathbf{a}}_i(h,k) - \tilde{\mathbf{a}}(h,k))]^2 = o(n) \textit{ a.s.,} \tag{B.50}$$

where

$$\tilde{\mathbf{a}}(h,k) = U_{k\times(k-1)} \left\{ \sum_{j=0}^{h-1} \alpha_{h-j}(k-1) \right\} + (1,0,\cdots,0)', \tag{B.51}$$

$$U_{k\times(k-1)} = \begin{pmatrix} 1 & 0 & \cdots & \cdots & 0 \\ -1 & 1 & \ddots & & \vdots \\ 0 & -1 & \ddots & \ddots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 0 & 0 & \cdots & 0 & -1 \end{pmatrix}$$

is a $k \times (k-1)$ *matrix, and* $\alpha_l(k-1) = \lim_{t\to\infty} \alpha_l^{(t)}(k-1)$*, with*

$$\alpha_l^{(t)}(k-1) = \arg \min_{(f_1,\cdots,f_{k-1})' \in R^{k-1}} E(s_{t+l} - f_1 s_t - \cdots - f_{k-1} s_{t-k+2})^2.$$

PROOF. First observe that $x_{t+h} = \eta_{t,h} + x_t + (\sum_{j=0}^{h-1} S_M^{h-1-j}(p)\alpha(p))' \mathbf{s}_t(p)$. This gives

$$\mathbf{x}_i'(k)(\check{\mathbf{a}}_i(h,k) - \tilde{\mathbf{a}}(h,k)) = W_{i,1}(k) + W_{i,2}(k), \tag{B.52}$$

where

$$W_{i,1}(k) = \mathbf{x}_i'(k) V_{i-h}(k) \sum_{j=k}^{i-h} \mathbf{x}_j(k) \eta_{j,h},$$

and

$$W_{i,2}(k) = \mathbf{x}_i'(k) V_{i-h}(k) \sum_{j=k}^{i-h} \mathbf{x}_j(k) \tilde{\varepsilon}_j(h,k),$$

with $\tilde{\varepsilon}_j(h,k) = (\sum_{l=0}^{h-1} S_M^{h-1-l}(p)\alpha(p))' \mathbf{s}_j(p) - (\sum_{l=0}^{h-1} \alpha_{h-l}(k-1))' \mathbf{s}_j(k-1)$. An argument similar to that used in the proof of Theorem 3.2 yields

$$\sum_{i=m_h}^{n-h} W_{i,1}^2(k) = O(\log n) \text{ a.s.}$$

In view of this and (B.52), (B.50) is ensured by

$$\sum_{i=m_h}^{n-h} W_{i,2}^2(k) = o(n) \text{ a.s.} \tag{B.53}$$

It is straightforward to see that

$$\begin{aligned} W_{n,2}^2(k) &\le (\|\mathbf{s}_n(k-1)\| + |x_n n^{-1/2}|)^2 \|\bar{R}_{n,h}^{-1}(k)\|^2 \\ &\times \|n^{-1} \sum_{j=k}^{n-h} (\mathbf{s}_j(k-1), x_j n^{-1/2})' \tilde{\varepsilon}_j(h,k)\|^2. \end{aligned} \tag{B.54}$$

Following the same reasonings as for (B.4) and (B.12), we have, for some $\iota > 0$,

$$\|n^{-1} \sum_{j=k}^{n-h} (\mathbf{s}_j(k-1), x_j n^{-1/2})' \tilde{\varepsilon}_j(h,k)\| = o(n^{-\iota}) \text{ a.s.} \tag{B.55}$$

By (B.38) and an argument similar to that used to prove (B.3),

$$\|\bar{R}_{n,h}^{-1}(k)\| = O(\log\log n) \text{ a.s.} \tag{B.56}$$

In addition, (B.1) guarantees

$$\sum_{j=m_h}^{n-h} \|\mathbf{s}_j(k-1)\|^2 = o(n) \text{ a.s.} \tag{B.57}$$

Consequently, (B.53) follows from (B.54)-(B.57) and (B.39). □

PROOF OF THEOREM 3.3. By an analogy with (B.43),

$$\begin{aligned}
\mathrm{APE}D_{n,h}(k) &= \sum_{i=m_h}^{n-h} \{\eta_{i,h} + \mathbf{x}_i'(p+1)\left(\mathbf{a}(h,p+1) - \check{\mathbf{a}}_i(h,k)\right)\}^2 \\
&= \sum_{i=m_h}^{n-h} \eta_{i,h}^2 + (1+o(1)) \sum_{i=m_h}^{n-h} \{\mathbf{x}_i'(p+1)\left(\mathbf{a}(h,p+1) - \check{\mathbf{a}}_i(h,k)\right)\}^2 \\
&+ O(1) \quad \text{a.s.,}
\end{aligned} \tag{B.58}$$

where $\check{\mathbf{a}}_i(h,k)$'s are viewed as $(p+1)$-dimensional vectors with undefined entries set to 0. Direct calculations yield

$$\begin{aligned}
&\sum_{i=m_h}^{n-h} \{\mathbf{x}_i'(p+1)\left(\mathbf{a}(h,p+1) - \check{\mathbf{a}}_i(h,k)\right)\}^2 \\
=\ & (\mathbf{a}(h,p+1) - \tilde{\mathbf{a}}(h,k))' V_{n-h}(k) (\mathbf{a}(h,p+1) - \tilde{\mathbf{a}}(h,k)) \\
-\ & 2\sum_{i=m_h}^{n-h} \mathbf{x}_i'(p+1)\left(\mathbf{a}(h,p+1) - \tilde{\mathbf{a}}(h,k)\right) \mathbf{x}_i'(p+1)\left(\check{\mathbf{a}}_i(h,k) - \tilde{\mathbf{a}}(h,k)\right) \\
+\ & \sum_{i=m_h}^{n-h} [\mathbf{x}_i'(\check{\mathbf{a}}_i(h,k) - \tilde{\mathbf{a}}(h,k))]^2,
\end{aligned} \tag{B.59}$$

where $\tilde{\mathbf{a}}(h,k)$ is defined in (B.51) (note that the $\tilde{\mathbf{a}}(h,k)$'s in the first two terms on the right hand side of (B.59) are viewed as $(p+1)$-dimensional vectors with undefined entries set to 0). By Lai and Wei (1983, (3.2)),

$$\liminf_{n\to\infty} n^{-1} V_{n-h}(k) > 0 \quad \text{a.s.} \tag{B.60}$$

Consequently, (3.9) follows from (B.58)-(B.60), (B.50), the Cauchy-Schwarz inequality, and the fact that $\mathbf{a}(h,p+1) - \tilde{\mathbf{a}}(h,k) \neq \mathbf{0}$. □

PROOF OF THEOREM 3.4. Theorem 3.4 follows immediately from Theorems 3.1-3.3. The details are omitted. □

APPENDIX C

In this appendix, we sketch the proof of Theorem 4.1. Applying an argument used in the proof of Wei (1992, Theorem 3.5), it can be shown that for $k < p_h$,

$$\liminf_{n\to\infty} \hat{\sigma}^2_{D,n}(h,k) - \hat{\sigma}^2_{D,n}(h,p_h) > 0 \ a.s. \tag{C.1}$$

Armed with the probability results obtained in Appendix B, we also have for $k_1 \geq p_1$ and $k_2 \geq p_h$,

$$|\hat{\sigma}^2_{P,n}(h,k_1) - \hat{\sigma}^2_{D,n}(h,k_2)| = o(\log n/n) \ a.s., \tag{C.2}$$

$$|\hat{\sigma}^2_{P,n}(h,k_1) - \hat{\sigma}^2_{P,n}(h,p_1)| = o(\log n/n) \ a.s., \tag{C.3}$$

$$|\hat{\sigma}^2_{D,n}(h,k_2) - \hat{\sigma}^2_{D,n}(h,p_h)| = o(\log n/n) \ a.s. \tag{C.4}$$

In addition, it can be shown that for $k_1 \geq p_1$ and $k_2 \geq p_h$,

$$\begin{aligned} & tr\left\{ (\sum_{j=k}^{n-h} \mathbf{x}_j(k)\mathbf{x}_j'(k))\ddot{L}_{h,n}(k)(\sum_{j=k}^{n-h} \mathbf{x}_j(k)\mathbf{x}_j'(k))^{-1}\ddot{L}_{h,n}'(k) \right\} \tilde{\sigma}^2_n C_n \\ = \ & \{\sigma^2(\sum_{j=0}^{h-1} b_j)^2 + f_{1,h}(k_1-1)\}C_n + o\left(C_n\right) \ a.s., \end{aligned} \tag{C.5}$$

and

$$\begin{aligned} & tr\left\{ (\sum_{j=k}^{n-h} \mathbf{x}_j(k)\mathbf{x}_j'(k))^{-1}(\sum_{j=k}^{n-2h+1} \mathbf{z}_j(k)\mathbf{z}_j'(k)) \right\} \tilde{\sigma}^2_n C_n \\ = \ & \{\sigma^2(\sum_{j=0}^{h-1} b_j)^2 + f_{2,h}(k_2-1)\}C_n + o\left(C_n\right) \ a.s. \end{aligned} \tag{C.6}$$

Consequently, the asymptotic efficiency of $(\hat{O}_n, \hat{M}_n)$ follows from (C.1)-(C.6).

REFERENCES

CHAN, N. H. AND WEI, C. Z. (1988): ”Limiting distribution of least squares estimates of unstable autoregressive processes,” *Annals of Statistics*, 16, 367-401.

CHEVILLON, G. AND HENDRY, D. F. (2005): ”Non-parametric direct multi-step estimation for forecasting economic processes,” *International Journal of Forecasting*, 21, 201-218.

CHOW, Y. S. (1965): ”Local convergence of martingale and the law of large numbers,” *Annals of Mathematical Statistics*, 36, 552-558.

FINDLEY, D. F. (1983): ”On the use of multiple models for multi-period forecasting,” *Proceeding of the Business and Statistics Section, American Statistical Association*, 528-531.

——— (1984): ”On some ambiguities associated with the fitting of ARMA models to time series,” *Journal of Time Series Analysis*, 5, 213-225.

FULLER, W. A. AND HASZA, D. P. (1981): ”Properties of predictors for autoregressive time series,” *Journal of American Statistical Association*, 76, 155-161.

ING, C. K. (2001): ”A note on mean-squared prediction errors of the least squares predictors in random walk models,” *Journal of Time Series Analysis*, 22, 711-724.

——— (2003): ”Multistep prediction in autoregressive processes,” *Econometric Theory*, 19, 254-279.

——— (2004): ”Selecting optimal multistep predictors for autoregressive processes of unknown order,” *Annals of Statistics*, 32, 693-722.

ING, C. -K. AND SIN, C. Y. (2006): ” Using information criteria to select a model with minimal prediction mean squared error: The case of unstable autoregression,” Technical Report, Institute of Statistical Science, Academia Sinica.

——— (2007): ”On prediction errors in regression models with nonstationary regressors,” in *Time Series and Related Topics: in Memory of Ching-Zong Wei*, ed. by H. C. Ho, C. K. Ing and T. L. Lai. IMS Lecture Notes-Monograph Series, Volume 52, to appear.

ING, C. -K. AND WEI, C. Z. (2003): ”On same-realization prediction in an infinite-order autoregressive process,” *Journal of Multivariate Analysis*, 85, 130-155.

——— (2006). "A maximal moment inequality for long range dependent time series with applications to estimation and model selection," *Statistica Sinica*, 16, 721-740.

KANG, I. B. (2003): "Multi-period forecasting using different models for different horizons: an application to U.S. economic time series data," *International Journal of Forecasting*, 19, 387-400.

LAI, T. L. AND WEI, C. Z. (1982): "Least squares estimates in stochastic regression models with application to identification and control systems," *Annals of Statistics*, 10, 154-166.

——— (1983): "Asymptotic properties of general autoregressive models and strong consistency of least squares estimates of their applications," *Journal of Multivariate Analysis*, 13, 1-23.

LIN, J. L. AND TSAY, R. S (1996): "Cointegration constraint and forecasting: an empirical examination," *Journal of Applied Econometrics*, 11, 519-538.

LIN, J. L. AND WEI, C. Z. (2007): "Forecasting unstable processes," in *Time Series and Related Topics: in Memory of Ching-Zong Wei*, ed. by H. C. Ho, C. K. Ing and T. L. Lai. IMS Lecture Notes-Monograph Series, Volume 52, to appear.

MÓRICZ, F. (1976): "Moment inequalities and the strong laws of large numbers," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 35, 299-314.

MARCELLINO, M, STOCK, J. H. AND WATSON, M. W. (2006): "A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series," *Journal of Econometrics*, 135, 499-526.

RISSANEN, J. (1986): "Order estimation by accumulated prediction errors," in *Essays in Time Series and Allied Processes: Papers in Honour of E. J. Hannan*, ed. by J. Gani and M. P. Priestley. Applied Probability Trust, Sheffield.

SCHORFHEIDE, F. (2005): "VAR forecasting under misspecification," *Journal of Econometrics*, 128, 99-136.

STOCK, J. H. AND WATSON, M. W. (2002): "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, 20, 147-162.

TIAO, G. C. AND TSAY, R. C. (1994): "Some advances in no-linear and adaptive modelling in time-series," *Journal of Forecasting*, 13, 109-131.

WEI, C. Z. (1987): "Adaptive prediction by least squares predictors in stochastic regression models with application to time series," *Annals of Statistics*, 15, 1667-1682.

——— (1992): "On predictive least squares principles," *Annals of Statistics*, 20, 1-42.

WEISS, A. A. (1991): "Multi-step estimation and forecasting in dynamic models," *Journal of Econometrics* , 48, 135-149.

WEST, K. D. (2006): "Forecast evaluation," in *Handbook of Economic Forecasting, Volume 1*, ed. by G. Elliott, C.W.J. Granger and A. Timmermann, Elsevier, North-Holland.

WHITE, H. (2006): "Time-series estimation of the effects of natural experiments," *Journal of Econometrics*, 135, 527-566.