Econometric Theory, **19**, 2003, 254–279. Printed in the United States of America.
DOI: 10.1017/S0266466603192031

MULTISTEP PREDICTION IN AUTOREGRESSIVE PROCESSES

CHING-KANG ING
National Taipei University

In this paper, two competing types of multistep predictors, i.e., plug-in and direct predictors, are considered in autoregressive (AR) processes. When a working model AR(k) is used for the h-step prediction with $h > 1$, the plug-in predictor is obtained from repeatedly using the fitted (by least squares) AR(k) model with an unknown future value replaced by their own forecasts, and the direct predictor is obtained by estimating the h-step prediction model's coefficients directly by linear least squares. Under rather mild conditions, asymptotic expressions for the mean-squared prediction errors (MSPEs) of these two predictors are obtained in stationary cases. In addition, we also extend these results to models with deterministic time trends. Based on these expressions, performances of the plug-in and direct predictors are compared. Finally, two examples are given to illustrate that some stationary case results on these MSPEs can not be generalized to the nonstationary case.

1. INTRODUCTION

Assume that observations $x_1, \ldots, x_n$ are generated from the following stationary autoregressive (AR) process:

$$x_t = \sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t, \qquad t = \ldots, -1, 0, 1, \ldots, \tag{1}$$

where $1 \le p < \infty$ is some unknown positive integer, $a_p \neq 0$, $\{\varepsilon_t\}$ is a sequence of (unobservable) independent and identically distributed (i.i.d.) random noises, each with mean 0 and variance σ^2, and the polynomial $A(z) = 1 - a_1 z - \cdots - a_p z^p \neq 0$, for $|z| \le 1$. These assumptions ensure that (see Brockwell and Davis, 1987, Theorem 3.1.1) x_t has a one-sided infinite moving-average representation:

$$x_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i},$$

The author is deeply grateful to the co-editor Pentti Saikkonen and two referees for their helpful suggestions and comments on a previous version of this paper. Address correspondence to Ching-Kang Ing, Department of Statistics, National Taipei University, No. 69, Sec. 2, Chan Kao N. Road, Taipei 104, Taiwan, R.O.C.; e-mail: cking@mail.ntpu.edu.tw.

  0266-4666/03 $12.00

where $b_0 + b_1 z + b_2 z^2 + \cdots = A^{-1}(z)$, $b_0 = 1$, and $b_k^2 \le \nu e^{-\beta k}$ for some $\nu, \beta > 0$. When a working model AR (k) is used for multistep prediction, a natural predictor of x_{n+h}, $h \ge 1$, can be obtained as follows. First, the parameters in the working model are estimated by least squares. Then, the predictor, $\hat{x}_{n+h}(k)$, is given by repeatedly using the fitted model with an unknown future value replaced by their own forecasts. In the following, $\hat{x}_{n+h}(k)$ will be referred to as the plug-in predictor (see also Kunitomo and Yamamoto, 1985; Bhansali, 1997). More specifically, let $\hat{\mathbf{a}}_n(1,k) = (\hat{a}_{1,n}(1,k), \ldots, \hat{a}_{k,n}(1,k))'$ denote the least-squares estimators of the AR parameters in the working model, where $\hat{\mathbf{a}}_n(1,k)$ satisfies

$$\sum_{j=k}^{n-1} \mathbf{x}_j(k)\mathbf{x}_j'(k)\hat{\mathbf{a}}_n(1,k) = \sum_{j=k}^{n-1} \mathbf{x}_j(k)x_{j+1}$$

with $\mathbf{x}_j(k) = (x_j, \ldots, x_{j-k+1})'$ for $j \ge k$. Then, for $h \ge 1$, the plug-in predictor is given by

$$\hat{x}_{n+h}(k) = \mathbf{x}_n'(k)\hat{A}_n^{h-1}(k)\hat{\mathbf{a}}_n(1,k), \tag{2}$$

where, with I_m and $\mathbf{0}_m$, respectively, denoting an identity matrix and a vector of zeros of dimension m,

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

and $\hat{A}_n^0(k) = I_k$.

When $k \ge p$ and $\{x_t\}$ is a Gaussian process, it can be shown that (e.g., Brockwell and Davis, 1987, Chs. 8 and 10) $\hat{\mathbf{a}}_n(1,k)$ is asymptotically equivalent to the maximum likelihood estimator (MLE) of $(a_1, \ldots, a_k)'$ with $a_i = 0$ for $i > p$. (Note that for $k \ge p$, the coefficient vector $(a_1, \ldots, a_k)'$ equals $\mathbf{a}(1,k) = R^{-1}(k)(\gamma_1, \ldots, \gamma_k)'$, where $R(k) = E\{[\mathbf{x}_t(k) - E(\mathbf{x}_t(k))][\mathbf{x}_t(k) - E(\mathbf{x}_t(k))]'\}$ and $\gamma_j = E\{[x_t - E(x_t)][x_{t-j} - E(x_{t-j})]\}$.) Therefore, by the invariance property of the MLE, the plug-in estimator,

$$\hat{\mathbf{a}}_n(h,k) = \hat{A}_n^{h-1}(k)\hat{\mathbf{a}}_n(1,k),$$

is asymptotically equivalent to the MLE of

$$\mathbf{a}(h,k) = (a_1(h,k), \ldots, a_k(h,k))' = A^{h-1}(k)\mathbf{a}(1,k),$$

where $k \ge p$, $A^0(k) = I_k$, and

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

Observe that $\mathbf{a}(h,k)$ with $k \geq p$ is also the coefficient vector of the h-step prediction formula corresponding to model (1), namely,

$$x_{t+h} = \sum_{j=1}^{k} a_j(h,k) x_{t+1-j} + \sum_{j=0}^{h-1} b_j \varepsilon_{t+h-j}. \tag{3}$$

These facts suggest that when $\{x_t\}$ is Gaussian, $\hat{x}_{n+h}(k)$ with $k \geq p$ may perform well compared to the other predictors of the same AR order. However, because it is difficult for the practitioner to know whether $k \geq p$, or $\{x_t\}$ is Gaussian, the advantage of $\hat{x}_{n+h}(k)$ can not be guaranteed. In this situation, the direct predictor, suggested by Findley (1984), is also frequently used as an alternative. The direct predictor, $\check{x}_{n+h}(k)$, is obtained through a linear least-squares regression of x_{t+h} on $x_t, \ldots, x_{t-k+1}$, i.e.,

$$\check{x}_{n+h}(k) = \mathbf{x}_n'(k) \check{\mathbf{a}}_n(h,k), \tag{4}$$

where

$$\check{\mathbf{a}}_n(h,k) = \arg \min_{(c_1,\ldots,c_k) \in R^k} \sum_{j=k}^{n-h} (x_{j+h} - c_1 x_j - \cdots - c_k x_{j-k+1})^2.$$

Viewing the preceding two predictors, it is natural to ask which one is better. When $h = 1$, it is obvious that these two predictors are identical. But for $h \geq 2$, this question has not been completely resolved. Recently, Bhansali (1997) has assumed that the process used in estimation and that for prediction are independent (we call this independent-realization prediction), and has compared the mean-squared prediction errors (MSPEs) of the plug-in and direct predictors under non-Gaussian settings. He shows that the plug-in predictor has a smaller MSPE, provided $k \geq p$ and the terms of order $o_p(1/n)$ are ignored. Although independent-realization prediction is considered in most related literature (e.g., Shibata, 1980; Bhansali, 1996), it is not a natural assumption for time series analysis, because when new observations of a time series become available, they are not independent of the previous data. Although it would seem to be more convincing to have a theoretical comparison between these two multistep predictors obtained without this unnatural assumption, to the author's knowledge, no conclusive results have been given in the existing literature.

To compare the performances of the plug-in and direct predictors, in Section 2, we consider the following (same-realization) multistep MSPEs:

$$\mathrm{MSPE}P_h(k) = E(x_{n+h} - \hat{x}_{n+h}(k))^2 \tag{5}$$

and

$$\mathrm{MSPE}D_h(k) = E(x_{n+h} - \check{x}_{n+h}(k))^2 \tag{6}$$

for $h \geq 1$. Under rather mild conditions, expressions for $\mathrm{MSPE}P_h(k)$ and $\mathrm{MSPE}D_h(k)$, up to terms of order $1/n$, are obtained in the situation where

$k \geq p$ (see Theorems 1 and 2, which follows). These expressions show that for $h \geq 2$,

$$\lim_{n\to\infty} n\{\mathrm{MSPE}D_h(k) - \mathrm{MSPE}P_h(k)\} > 0, \tag{7}$$

a conclusion that coincides with that of Bhansali (1997) for the independent-realization setting. In addition, we also show in Section 2 that when $k \geq p$ is fixed, the advantage of the plug-in predictor with respect to the direct predictor increases with h.

Theorem 3 in Section 2 considers the difference between $\mathrm{MSPE}P_h(k)$ and $\mathrm{MSPE}P_h(k+1)$ and that between $\mathrm{MSPE}D_h(k)$ and $\mathrm{MSPE}D_h(k+1)$ for $k \geq p$. When $k \geq p$, it is shown in Theorem 3 that

$$\lim_{n\to\infty} n[\mathrm{MSPE}P_h(k+1) - \mathrm{MSPE}P_h(k)] > 0$$

and

$$\lim_{n\to\infty} n[\mathrm{MSPE}D_h(k+1) - \mathrm{MSPE}D_h(k)] > 0.$$

Therefore, the (strict) parsimony principle (Findley and Wei, 2002, Sect. 3), which asserts that if two competing models are both correctly specified, then the larger estimated model has a larger limiting (second-order) MSPE, still holds for multistep predictions. It is also worth noting that although these differences are of order $1/n$, and are often ignored by the practitioner, they play important roles in order selection problems (see the discussion following Theorem 3 for more details). At the end of Section 2, attention is focused on the misspecified case. In contrast to the result obtained in (7), if $k < p$, then except on a zero-volume subset of the parameter space,

$$\lim_{n\to\infty} (\mathrm{MSPE}P_h(k) - \mathrm{MSPE}D_h(k)) > 0;$$

see Theorem 4 in Section 2 for more details.

In Section 3, asymptotic expressions for the MSPEs of the plug-in and direct predictors are obtained in the situation where a deterministic time trend is included in model (1) (see Corollaries 1 and 2). Various comparisons of these MSPEs are also performed at the end of Section 3. In Section 4, two examples are presented illustrating some interesting phenomena associated with the plug-in and direct predictors with nonstationary AR models. The first example shows that the left-hand side of (7) (with $k = 1$ and $h \geq 2$) will converge to 0 in the random walk case. Therefore, the advantage of the plug-in predictor vanishes even when $k \geq p$. The second example shows that the strict parsimony principle mentioned previously does not hold for the plug-in predictor in certain nonstationary AR (2) models. Proofs for the results in Sections 2 and 3 are deferred to Appendixes A and B, respectively.

2. ASYMPTOTIC EXPRESSIONS FOR MSPE$P_h(k)$ AND MSPE$D_h(k)$ IN STATIONARY AR MODEL

In the discussion that follows, we shall frequently use the following condition on the distribution function of $\varepsilon_1, F(\cdot)$. For some positive numbers α, η, and M,

$$|F(x) - F(y)| \leq M|x-y|^{\alpha}, \tag{8}$$

provided $|x - y| < \eta$. Condition (8) is fulfilled by any distribution function with bounded density. Some distribution functions with unbounded density also satisfy (8). For more details on this condition, see Findley and Wei (2002, Sect. 4). When data are generated from model (1), Theorems 1 and 2 deal with asymptotic expressions for MSPE$P_h(k)$ and MSPE$D_h(k)$ when $k \geq p$, respectively. Their proofs are given in Appendix A.

THEOREM 1. *Assume that observations $x_1, \ldots, x_n$ are generated from model (1) with ε_1 satisfying (8) and*

$$E(|\varepsilon_1|^{\theta_h}) < \infty, \tag{9}$$

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

$$n\left(\text{MSPE}P_h(k) - \sum_{j=0}^{h-1} b_j^2 \sigma^2\right) = f_{1,h}(k) + O(n^{-1/2}), \tag{10}$$

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

$$f_{1,h}(k) = \operatorname{tr}(R(k) L_h(k) R^{-1}(k) L_h'(k))\sigma^2.$$

Remark 1. Condition (8) is used to ensure that for any $q > 0$,

$$E\|\hat{R}_n^{-1}(k)\|^q = O(1) \tag{11}$$

holds for n sufficiently large, where

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

and $\|G\| = \sup_{\|\mathbf{z}\| \leq 1} \|G\mathbf{z}\|$ for matrix G (see Lemma A.1 in Appendix A for more details). Equality (11) was first proven by Fuller and Hasza (1981) for stationary Gaussian processes. For various generalizations of their result, see Bhansali and Papangelou (1991), Papangelou (1994), Ing and Wei (2002), and Findley and Wei (2002). By directly assuming that (11) holds, Kunitomo and Yamamoto (1985, Corollary 6) have established (10) with θ_h in (9) replaced by $\max\{32, 16h\}$. Obviously, their assumptions are more stringent than those used in Theorem 1.

THEOREM 2. *Let the assumptions of Theorem 1 hold with θ_h in (9) replaced by* $8 + \delta$, *for some* $\delta > 0$. *Then, for* $k \geq p$ *and* $h \geq 1$,

$$n\left(\mathrm{MSPE}D_h(k) - \sum_{j=0}^{h-1} b_j^2 \sigma^2\right) = f_{2,h}(k) + O(n^{-1/2}), \tag{12}$$

where

$$f_{2,h}(k) = \mathrm{tr}\left[R^{-1}(k)\mathrm{cov}\left(\sum_{j=0}^{h-1} b_j \mathbf{x}_{k+j}(k)\right)\right]\sigma^2,$$

and for random vector $\mathbf{z}$, $\mathrm{cov}(\mathbf{z}) = E[(\mathbf{z} - E(\mathbf{z}))(\mathbf{z} - E(\mathbf{z}))']$.

A comparison between the first terms on the right-hand side of (10) and (12), $f_{1,h}(k)$ and $f_{2,h}(k)$, respectively, is given as Proposition 3.2 by Bhansali (1997), which shows that for $k \geq p \geq 1$ and $h \geq 2$,

$$f_{2,h}(k) > f_{1,h}(k). \tag{13}$$

This result and Theorems 1 and 2 together imply that when the model is correctly specified, the plug-in predictor has an advantage over the direct predictor, provided the terms of order $O(n^{-3/2})$ are ignored. For example, assume $h = 2$ and $k \geq p \geq 1$. Then,

$$f_{2,2}(k) = \{k + (k+2)a_1^2\}\sigma^2 \tag{14}$$

and

$$f_{1,2}(k) = \{(k+2)a_1^2 + k - 1 + a_k^2\}\sigma^2. \tag{15}$$

(Recall that $a_k = 0$ for $k > p$ and that $|a_p| < 1$ in the stationary case.) Hence, for $k \geq p$,

$$n(\mathrm{MSPE}D_2(k) - \mathrm{MSPE}P_2(k)) = (1 - a_k^2)\sigma^2 + O(n^{-1/2}). \tag{16}$$

Moreover, let $k \geq p \geq 1$ be fixed. Then, it can be shown that $f_{2,h}(k) - f_{1,h}(k)$ is a nondecreasing function of h. Therefore, the advantage of the plug-in predictor with respect to the direct predictor becomes more significant as h increases. To see this, observe that

$$\begin{aligned} f_{1,h+1}(k) &= \mathrm{tr}(R(k)L_{h+1}(k)R^{-1}(k)L'_{h+1}(k))\sigma^2 \\ &= \mathrm{tr}(L_h(k)R^{-1}(k)L'_h(k)A'(k)R(k)A(k))\sigma^2 \\ &\quad + 2\mathrm{tr}\left(b_h \sum_{j=0}^{h-1} b_j A^{h-j}(k)\right)\sigma^2 + b_h^2 k\sigma^2 \\ &= f_{1,h}(k) - \mathbf{e}'_k L_h(k)R^{-1}(k)L'_h(k)\mathbf{e}_k\sigma^4 \\ &\quad + 2\mathrm{tr}\left(b_h \sum_{j=0}^{h-1} b_j A^{h-j}(k)\right)\sigma^2 + b_h^2 k\sigma^2, \end{aligned}$$

where the third equality follows from $A'(k)R(k)A(k) = R(k) - \sigma^2\mathbf{e}_k\mathbf{e}_k'$, with k-dimensional vector $\mathbf{e}_k = (1, \mathbf{0}_{k-1}')'$. In addition, one also has

$$f_{2,h+1}(k) = f_{2,h}(k) + 2\mathrm{tr}\left(b_h \sum_{j=0}^{h-1} b_j A^{h-j}(k)\right)\sigma^2 + b_h^2 k\sigma^2.$$

As a result,

$$f_{2,h+1}(k) - f_{1,h+1}(k) = f_{2,h}(k) - f_{1,h}(k) + \mathbf{e}_k' L_h(k) R^{-1}(k) L_h'(k)\mathbf{e}_k\sigma^4, \qquad \mathbf{(17)}$$

which gives the asserted property.

The following result establishes that the second-order terms of MSPE$P_h(k)$ and MSPE$D_h(k)$ are strictly increasing functions of k in the correctly specified case.

THEOREM 3.

(i) *Assume $h \geq 1$ and $k \geq p$. Then,*

$$f_{1,h}(k+1) > f_{1,h}(k), \qquad \mathbf{(18)}$$

provided

$$b_{h-1} \neq 0 \qquad \mathbf{(19)}$$

or

$$\mathbf{l}^* \neq \mathbf{0}_{k+1}, \qquad \mathbf{(20)}$$

where with the convention that $b_j = 0$ for $j < 0$, $\mathbf{l}^* = (\sum_{i=0}^{h-1} b_{h-1-k-i} b_i, \ldots, \sum_{i=0}^{h-1} b_{h-1-i} b_i)'$ is a $(k+1)$-dimensional vector.

(ii) *Assume $h \geq 1$ and $k \geq p$. Then,*

$$f_{2,h}(k+1) = f_{2,h}(k) + \sum_{l=0}^{h-1} b_l^2\sigma^2 > f_{2,h}(k). \qquad \mathbf{(21)}$$

Remark 2. When $1 \leq h \leq 5$, it can be shown that either (19) or (20) holds for all $k \geq p$. In addition, if $h - 1 \leq k$, then the $(k + 2 - h)$th entry of $\mathbf{l}^*$ is $b_0^2 = 1$. In these cases, (18) holds without extra constraints on the parameter space. In general, (19) (with $h \geq 2$) and (20) only exclude at most a zero-volume subset in the parameter space.

Under the assumptions of Theorem 1 and (20) (or (19)), we conclude from (10) and (18) that for $h \geq 1$ and $k \geq p$,

$$\lim_{n\to\infty} n\{\mathrm{MSPE}P_h(k+1) - \mathrm{MSPE}P_h(k)\} > 0. \qquad \mathbf{(22)}$$

Similarly, when the assumptions of Theorem 2 hold, (12) and (21) also yield

$$\lim_{n\to\infty} n\{\mathrm{MSPE}D_h(k+1) - \mathrm{MSPE}D_h(k)\} = \sum_{l=0}^{h-1} b_l^2 \sigma^2 > 0 \tag{23}$$

for $h \geq 1$ and $k \geq p$.

Kunitomo and Yamamoto (1985, Theorem 4) obtain that $f_{1,h}(k+1) \geq f_{1,h}(k)$ when $k \geq p$ and $h \geq 2$ (see also (A.26)). However, from the order selection point of view, Kunitomo and Yamamoto's result is not sharp enough. To see this, first consider the following multistep generalizations of accumulated prediction errors (APEs) based on sequential plug-in and direct predictors:

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

and

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

respectively, where m is the smallest positive number such that $\hat{\mathbf{a}}_i(h,k)$ and $\check{\mathbf{a}}_i(h,k)$ are well defined for all $i \geq m$. Note that the APE with $h = 1$ was first proposed by Rissanen (1986) for the purpose of use for determining p. For $h = 1$, Hemerly and Davis (1989), under certain conditions, prove that $\hat{k}_n = \arg\min_{1\leq k\leq K} \mathrm{APE}P_{n,1}(k) = \arg\min_{1\leq k\leq K} \mathrm{APE}D_{n,1}(k)$ is a strongly consistent estimator of p in the sense that

$$P\left(\lim_{n\to\infty} \hat{k}_n = p\right) = 1.$$

Here, K is a prescribed upper bound of p. Under the conditions similar to those used in Hemerly and Davis (1989), Ing (2002) further shows that for $h > 1$

$$P\left(\lim_{n\to\infty} \frac{1}{\sigma^2 \log n} \{\mathrm{APE}P_{n,h}(k+1) - \mathrm{APE}P_{n,h}(k)\} = f_{1,h}(k+1) - f_{1,h}(k)\right) = 1 \tag{24}$$

and

$$P\left(\lim_{n\to\infty} \frac{1}{\sigma^2 \log n} \{\mathrm{APE}D_{n,h}(k+1) - \mathrm{APE}D_{n,h}(k)\} = f_{2,h}(k+1) - f_{2,h}(k)\right) = 1, \tag{25}$$

as $k \geq p$. As is clear from (24) (or (25)), without the strict inequality, (18) (or (21)), it does not seem possible for $\hat{k}^{(h)}_{P,n} = \arg\min_{1\leq k\leq K} \mathrm{APE}P_{n,h}(k)$ (or $\hat{k}^{(h)}_{D,n} = \arg\min_{1\leq k\leq K} \mathrm{APE}D_{n,h}(k)$) with $h \geq 2$ to be strongly consistent. For more details on the statistical properties of $\mathrm{APE}P_{n,h}(k)$ and $\mathrm{APE}D_{n,h}(k)$, see Ing (2002).

The following theorem compares $\mathrm{MSPE}P_h(k)$ and $\mathrm{MSPE}D_h(k)$ in the situation where $k < p$.

THEOREM 4. *Assume that observations $x_1, \ldots, x_n$ are generated from model (1) with $1 \le p \le \infty$. (Here, $p = \infty$ means that $\{x_t\}$ is truly an $\mathrm{AR}(\infty)$ process. In this case, $\sum_{i=1}^{\infty} |a_i| < \infty$ is assumed instead of $a_p \neq 0$.) Also assume that ε_1 satisfies (8) and $E(|\varepsilon_1|^{4h+2+\delta}) < \infty$, for some $\delta > 0$. Then, for $h \ge 2$ and $1 \le k < p$,*

$$\lim_{n\to\infty} (\mathrm{MSPE}P_h(k) - \mathrm{MSPE}D_h(k)) \\ = (\mathbf{a}_D(h,k) - \mathbf{a}(h,k))' R(k) (\mathbf{a}_D(h,k) - \mathbf{a}(h,k)), \tag{26}$$

where $\mathbf{a}_D(h,k) = R^{-1}(k)(\gamma_h, \ldots, \gamma_{h+k-1})'$ and $\mathbf{a}(h,k)$ is defined in Section 1. Moreover, if

$$\gamma_s \neq \gamma_s^*, \tag{27}$$

for at least one positive integer s satisfying $h \le s \le k + h - 1$, then for $h \ge 2$ and $1 \le k < p$,

$$\lim_{n\to\infty} (\mathrm{MSPE}P_h(k) - \mathrm{MSPE}D_h(k)) > 0, \tag{28}$$

where $\gamma_j^* = \gamma_j$ for $1 \le j \le k$ and $\gamma_j^* = \mathbf{a}_D(1,k)'(\gamma_{j-1}^*, \ldots, \gamma_{j-k}^*)'$ for $j \ge k + 1$.

Remark 3. Because p is allowed to be infinite, Theorem 4 covers the causal and invertible autoregressive moving average (ARMA) process as a special case. When (27) is violated, $\gamma_{\max\{h,k+1\}}, \ldots, \gamma_{k+h-1}$ must be functions of $\gamma_0, \ldots, \gamma_k$. Therefore, (27) only excludes at most a zero-volume subset in the parameter space. In particular, (27) holds automatically in the situation where $p \le k + h - 1$ and $h \le k + 1$ (e.g., $h = 2$ and $k = p - 1$), because if (27) is violated in this situation, then $a_{k+1} = \cdots = a_p = 0$, which contradicts the assumption that $a_p \neq 0$.

Theorem 4 and Remark 3 assert that in the misspecified case, the direct predictor is superior to the plug-in predictor for almost all points in the parameter space. On the other hand, we have found examples showing that when (27) is violated, the plug-in predictor can be better than the direct predictor in the sense of (7). However, to obtain a more conclusive result for the situation where (27) does not hold, it is necessary to compare $\mathrm{MSPE}P_h(k)$ and $\mathrm{MSPE}D_h(k)$, up to terms of order $1/n$. Because this comparison is quite complicated, we do not pursue these details in this paper. For a related result on an asymptotic expression for $\mathrm{MSPE}P_h(k)$, up to terms of order $1/n$, in the misspecified case, see Kunitomo and Yamamoto (1985, Theorem 3).

3. EXTENSIONS TO STATIONARY AR MODELS AROUND DETERMINISTIC TIME TRENDS

This section investigates performances of the plug-in and direct predictors in stationary AR models around deterministic time trends. Assume that observations $x_1, \ldots, x_n$ come from the following model:

$$x_t = \sum_{i=0}^{q_0} \beta_i t^i + \sum_{i=1}^{p} a_i x_{t-i} + \varepsilon_t, \qquad t = \ldots, -1, 0, 1, \ldots, \tag{29}$$

where q_0 is a nonnegative integer, β_i's are unknown real numbers, and p, a_i's and ε_t are defined in model (1). We also assume that q_0 in model (29) is known to simplify the discussion. With an argument similar to that used in Theorem 3.1.1 of Brockwell and Davis (1987) and some algebraic manipulations, x_t can be expressed as

$$x_t = \sum_{i=0}^{q_0} \beta_i^* t^i + y_t, \tag{30}$$

where β_i^*'s are functions of β_i's and a_i's and

$$y_t = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i},$$

with b_i's defined in Section 1, is a stationary AR(p) process satisfying model (1). For example, if $q_0 = 1$ (linear time trend), then

$$\beta_0^* = \frac{\beta_0}{1 - \sum_{i=1}^{p} a_i} - \beta_1 \frac{\sum_{i=1}^{p} i a_i}{\left(1 - \sum_{i=1}^{p} a_i\right)^2}$$

and

$$\beta_1^* = \frac{\beta_1}{1 - \sum_{i=1}^{p} a_i}.$$

Under model (29), the plug-in predictor of x_{n+h}, $\dot{x}_{n+h}(k)$, has the following representation:

$$\dot{x}_{n+h}(k) = \tilde{\mathbf{x}}_n'(k)\dot{\mathbf{a}}_n(h,k) = \tilde{\mathbf{x}}_n'(k)\hat{B}_n^{h-1}(k)\dot{\mathbf{a}}_n(1,k), \tag{31}$$

where $\tilde{\mathbf{x}}_t(k) = (1, t+1, \ldots, (t+1)^{q_0}, x_t, \ldots, x_{t-k+1})'$ is a (q_0+1+k)-dimensional vector, $\dot{\mathbf{a}}_n(h,k) = \hat{B}_n^{h-1}(k)\dot{\mathbf{a}}_n(1,k)$,

$$\begin{aligned}\dot{\mathbf{a}}_n(1,k) &= (\hat{\beta}_{0,n}, \ldots, \hat{\beta}_{q_0,n}, \dot{a}_{1,n}(1,k), \ldots, \dot{a}_{k,n}(1,k))' \\ &= \left(\sum_{j=k}^{n-1} \tilde{\mathbf{x}}_j(k)\tilde{\mathbf{x}}_j'(k)\right)^{-1} \sum_{j=k}^{n-1} \tilde{\mathbf{x}}_j(k) x_{j+1},\end{aligned}$$

and

$$\hat{B}_n(k) = \begin{pmatrix} B_1 & \hat{B}_{2,n}(k) \\ \mathbf{0}_{k\times(q_0+1)} & \dot{A}_n(k) \end{pmatrix}$$

is a $(q_0+1+k)\times(q_0+1+k)$ matrix. Here,

$$B_1 = \begin{pmatrix} 1 & C_0^1 & C_0^2 & \ldots & C_0^{q_0} \\ 0 & C_1^1 & C_1^2 & & \vdots \\ \vdots & \ddots & C_2^2 & & \vdots \\ \vdots & & \ddots & \ddots & C_{q_0-1}^{q_0} \\ 0 & \ldots & \ldots & 0 & C_{q_0}^{q_0} \end{pmatrix}$$

is a $(q_0+1)\times(q_0+1)$ upper triangular matrix with $C_l^r = r!/(l!(r-l)!)$, $0 \le l \le r$,

$$\hat{B}_{2,n}(k) = \begin{pmatrix} \hat{\beta}_{0,n} & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ \hat{\beta}_{q_0,n} & 0 & \ldots & 0 \end{pmatrix}$$

is a $(q_0+1)\times k$ matrix, $\mathbf{0}_{k\times(q_0+1)}$ is a $k\times(q_0+1)$ matrix of zeros, and $\dot{A}_n(k)$ is the matrix $\hat{A}_n(k)$ (see Section 1) with $\hat{\mathbf{a}}_n(1,k)$ replaced by $(\dot{a}_{1,n}(1,k), \ldots, \dot{a}_{k,n}(1,k))'$. The corresponding direct predictor, $\ddot{x}_{n+k}(k)$, is

$$\ddot{x}_{n+h}(k) = \tilde{\mathbf{x}}_n'(k)\ddot{\mathbf{a}}_n(h,k), \tag{32}$$

where

$$\ddot{\mathbf{a}}_n(h,k) = \left(\sum_{j=k}^{n-h} \tilde{\mathbf{x}}_j(k)\tilde{\mathbf{x}}_j'(k)\right)^{-1} \sum_{j=k}^{n-h} \tilde{\mathbf{x}}_j(k) x_{j+h}.$$

Corollary 1 establishes an asymptotic expression for the MSPE of $\dot{x}_{n+h}(k)$, $k \ge p$, under non-Gaussian settings.

COROLLARY 1. *Assume that in model (29)* ε_1 *satisfies (8) and for any* $q > 0$,

$$E(|\varepsilon_1|^q) < \infty. \tag{33}$$

Then, for $k \geq p$ *and* $h \geq 1$,

$$n\left(\mathrm{MSPE}\widetilde{P}_h(k) - \sum_{j=0}^{h-1} b_j^2 \sigma^2\right) = C_{q_0,h} + f_{1,h}(k) + O(n^{-1/2}), \tag{34}$$

where

$$\mathrm{MSPE}\widetilde{P}_h(k) = E(x_{n+h} - \dot{x}_{n+h}(k))^2$$

and

$$C_{q_0,h} = \mathbf{1}'_{q_0+1} C^*_{q_0} \mathbf{1}_{q_0+1} \left(\sum_{i=0}^{h-1} b_i\right)^2 \sigma^2,$$

with $\mathbf{1}_{q_0+1}$ *denoting a vector of 1's of dimensional* $q_0 + 1$ *and*

$$C^*_{q_0} = \begin{pmatrix} 1 & \frac{1}{2} & \cdots & \frac{1}{q_0+1} \\ \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{q_0+2} \\ \vdots & \vdots & & \vdots \\ \frac{1}{q_0+1} & \frac{1}{q_0+2} & \cdots & \frac{1}{2q_0+1} \end{pmatrix}^{-1}$$

being a $(q_0 + 1) \times (q_0 + 1)$ *matrix.*

Remark 4. As observed in (34), the (second-order) MSPE of the plug-in predictor is the sum of two terms. By comparing (34) with (10), the first term, $C_{q_0,h}$, can be viewed as the loss due to forecasting the deterministic time trend, $\sum_{i=0}^{q_0} \beta_i^* t^i$, in (30), and the second term, $f_{1,h}(k)$, can be viewed as the loss due to forecasting the stationary process, y_t, in (30). Equality (34) considerably extends (2.5) of Fuller and Hasza (1981), which provides an asymptotic expression for $\mathrm{MSPE}\widetilde{P}_h(k)$ in the situation where $q_0 = 0$ (constant time trend) and ε_i's are independent Gaussian random variables, each with mean 0 and variance σ^2.

Corollary 2 provides an asymptotic expression for the MSPE of $\ddot{x}_{n+h}(k)$ with $k \geq p$.

COROLLARY 2. *Let the assumptions of Corollary 1 hold. Then, for $k \geq p$ and $h \geq 1$,*

$$n\left(\mathrm{MSPE}\widetilde{D}_h(k) - \sum_{j=0}^{h-1} b_j^2 \sigma^2\right) = C_{q_0,h} + f_{2,h}(k) + O(n^{-1/2}), \tag{35}$$

where

$$\mathrm{MSPE}\widetilde{D}_h(k) = E(x_{n+h} - \ddot{x}_{n+h}(k))^2.$$

The (second-order) MSPE of the direct predictor is also the sum of two different losses due to forecasting two different components in model (30), and the loss due to forecasting the deterministic time trend is exactly the same as that of the plug-in predictor. To illustrate the preceding theoretical results, assume that $q_0 = 1$ and $h = 2$. In this case, $C_{q_0,h} = 4(1 + a_1)^2\sigma^2$. Hence, for $k \geq p$,

$$\begin{aligned}\lim_{n\to\infty} n\left(\mathrm{MSPE}\widetilde{P}_2(k) - \sum_{j=0}^{1} b_j^2 \sigma^2\right) &= 4(1+a_1)^2\sigma^2 + f_{1,2}(k)\\ &= 4(1+a_1)^2\sigma^2\\ &\quad + \{(k+2)a_1^2 + k - 1 + a_k^2\}\sigma^2\end{aligned}$$

and

$$\begin{aligned}\lim_{n\to\infty} n\left(\mathrm{MSPE}\widetilde{D}_2(k) - \sum_{j=0}^{1} b_j^2 \sigma^2\right) &= 4(1+a_1)^2\sigma^2 + f_{2,2}(k)\\ &= 4(1+a_1)^2\sigma^2 + \{(k+2)a_1^2 + k\}\sigma^2.\end{aligned}$$

In general, one has for $k \geq p$ and $h \geq 2$,

$$\lim_{n\to\infty} n(\mathrm{MSPE}\widetilde{D}_h(k) - \mathrm{MSPE}\widetilde{P}_h(k)) = f_{2,h}(k) - f_{1,h}(k) > 0. \tag{36}$$

Therefore, the plug-in predictor is still better than the direct predictor in the sense of (36). In addition, by Corollaries 1 and 2 and Theorem 3, when $k \geq p$ and $h \geq 1$, the strict parsimony principle still holds for the plug-in and direct predictors in stationary AR models with deterministic time trends.

The following corollary (Corollary 3) compares the performances of $\dot{x}_{n+h}(k)$ and $\ddot{x}_{n+h}(k)$ with $k < p$. It shows that if (27) holds, then for $h > 1$ and $k < p$, the direct predictor is better than the plug-in predictor, a conclusion that coincides with that of Theorem 4 for the zero mean process. Like Theorem 4, Corollary 3 also allows p in model (29) to equal ∞. In this case, $\sum_{i=1}^{\infty}|a_i| < \infty$ is assumed instead of $a_p \neq 0$.

COROLLARY 3. *Assume that in model (29) $1 \le p \le \infty$ and ε_1 satisfies (8) and (33). Then, for $h \ge 2$ and $1 \le k < p$,*

$$\lim_{n\to\infty} (\mathrm{MSPE}\widetilde{P}_h(k) - \mathrm{MSPE}\widetilde{D}_h(k)) = (\mathbf{a}_D(h,k) - \mathbf{a}(h,k))' R(k) \times (\mathbf{a}_D(h,k) - \mathbf{a}(h,k)).$$

Moreover, if (27) holds, then for $h \ge 2$ and $1 \le k < p$,

$$\lim_{n\to\infty} (\mathrm{MSPE}\widetilde{P}_h(k) - \mathrm{MSPE}\widetilde{D}_h(k)) > 0.$$

4. SOME AMBIGUITIES ASSOCIATED WITH PLUG-IN AND DIRECT PREDICTORS IN NONSTATIONARY AR MODELS

Although the advantage of the plug-in predictor with respect to the direct predictor is guaranteed in the situation where the underlying process is stationary and $k \ge p$, this is not the case for the random walk model,

$$x_t = x_{t-1} + \varepsilon_t, \tag{37}$$

even if the assumed model is correctly specified. Throughout this section, we assume that $\{\varepsilon_t\}$ is a sequence of independent Gaussian random variables, each with mean 0 and variance σ^2.

PROPOSITION 1. *Let observations $x_1, \dots, x_n$ come from model (37) with the initial condition $x_0 = 0$. Then for $h \ge 1$,*

$$\lim_{n\to\infty} n(\mathrm{MSPE}P_h(1) - h\sigma^2) = 2h^2\sigma^2$$

and

$$\lim_{n\to\infty} n(\mathrm{MSPE}D_h(1) - h\sigma^2) = 2h^2\sigma^2.$$

Consequently,

$$\lim_{n\to\infty} n(\mathrm{MSPE}D_h(1) - \mathrm{MSPE}P_2(h)) = 0.$$

The proof of Proposition 1 is similar to that of Corollary 1 of Ing (2001). We omit the details. In the rest of this section, we consider the following nonstationary AR(2) model:

$$x_t = x_{t-2} + \varepsilon_t. \tag{38}$$

Proposition 2, which follows, shows that under model (38) the plug-in two-step-ahead predictor obtained from the AR(2) model and that obtained from the AR(3) model have the same prediction efficiency. (Recall that in the stationary AR(2) model, Remark 2 and (22) show that the plug-in two-step-ahead predictor obtained from the AR(2) model is strictly better than that obtained from the AR(3) model.)

PROPOSITION 2. *Let observations $x_1, \ldots, x_n$ come from model (38) with the initial condition $x_i = 0$ for $i = 0, -1$. Then,*

$$\lim_{n\to\infty} n(\mathrm{MSPE}P_2(3) - \sigma^2) = 4\sigma^2,$$

$$\lim_{n\to\infty} n(\mathrm{MSPE}P_2(2) - \sigma^2) = 4\sigma^2,$$

and

$$\lim_{n\to\infty} n(\mathrm{MSPE}P_2(3) - \mathrm{MSPE}P_2(2)) = 0.$$

Remark 5. A complete verification of Proposition 2, which requires a rather subtle argument, is not presented in this paper, but it is available from the author. We also note that the ambiguity encountered with plug-in predictors stated in Proposition 2 does not exist for their corresponding direct predictors because it can be shown that

$$\lim_{n\to\infty} n(\mathrm{MSPE}D_2(3) - \mathrm{MSPE}D_2(2)) = \sigma^2.$$

The main purpose of this section is to show that some stationary case results on MSPEs of the plug-in and direct predictors can not be generalized to nonstationary cases. However, to gain a deeper understanding of these predictors' performances in nonstationary cases, extensions of Proposition 1 (or 2) to more general nonstationary AR models are needed. Further investigation into this question would be interesting.

REFERENCES

Bhansali, R.J. (1996) Asymptotically efficient autoregressive model selection for multistep prediction. *Annals of the Institute of Statistical Mathematics* 48, 577–602.

Bhansali, R.J. (1997) Direct autoregressive predictors for multistep prediction: order selection and performance relative to the plug in predictors. *Statistica Sinica* 7, 425–449.

Bhansali, R.J. & F. Papangelou (1991) Convergence of moments of least squares estimators for the coefficients of an autoregressive process of unknown order. *Annals of Statistics* 19, 1155–1162.

Brockwell, P.J. & R.A. Davis (1987) *Time Series: Theory and Methods*. New York: Springer-Verlag.

Findley, D.F. (1984) On some ambiguities associated with the fitting of ARMA models to time series. *Journal of Time Series Analysis* 5, 213–225.

Findley, D.F. & C.Z. Wei (2002) AIC, overfitting principles, and boundedness of moments of inverse matrices for vector autoregressive models fit to weakly stationary time series. *Journal of Multivariate Analysis*, forthcoming.

Fuller, W.A. & D.P. Hasza (1981) Properties of predictors for autoregressive time series. *Journal of American Statistical Association* 76, 155–161.

Hamilton, J.D. (1994) *Time Series Analysis*. Princeton: Princeton University Press.

Hemerly, E.M. & M.H.A. Davis (1989) Strong consistency of the PLS criterion for order determination of autoregressive processes. *Annals of Statistics* 17, 941–946.
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.
Ing, C.K. (2002) Selecting optimal multistep predictors in autoregressive processes. Manuscript.
Ing, C.K. & C.Z. Wei (2002) On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis*, forthcoming.
Kunitomo, N. & T. Yamamoto (1985) Properties of predictors in misspecified autoregressive time series. *Journal of the American Statistical Association* 80, 941–950.
Papangelou, F. (1994) On a distributional bound arising in autoregressive model fitting. *Journal of Applied Probability* 31, 401–408.
Rissanen, J. (1986) Order estimation by accumulated prediction errors. In *Essays in Time Series and Applied Processes,* special volume 23A of *Journal of Applied Probability*, 55–61.
Schwarz, G. (1978) Estimating of dimension of a model. *Annals of Statistics* 6, 461–464.
Shibata, R. (1980) Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* 8, 147–164.
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.

APPENDIX A

In the following, we use C to denote a generic positive constant independent of the sample size n. To verify Theorems 1 and 2, we begin with the following three lemmas, Lemmas A.1, A.2, and A.3, which are immediate consequences of Lemma 1, Theorem 2, and Lemma 4 of Ing and Wei (2002), respectively.

LEMMA A.1. *Let the assumptions of Theorem 1 hold with θ_h in (9) replaced by 2. Then, for any positive integer k and any $q > 0$,*

$$E(\|\hat{R}_n^{-1}(k)\|^q) \leq C$$

holds for sufficiently large n.

LEMMA A.2. *Let the assumptions of Theorem 1 hold with θ_h in (9) replaced by* $\max\{2q_1,4\}$*, for some $q_1 > 0$. Then, for any positive integer k and any $0 < q < q_1$,*

$$E(\|\hat{R}_n^{-1}(k) - R^{-1}(k)\|^q) \leq Cn^{-q/2}$$

holds for sufficiently large n.

LEMMA A.3. *Assume that $\{x_t\}$ satisfies model (1) with $E(|\varepsilon_1|^q) < \infty$ for some $q \geq 2$. Then, for any positive integer k,*

$$E\left(\left\|\frac{1}{\sqrt{n}}\sum_{j=k}^{n-1}\mathbf{x}_j(k)\varepsilon_{j+1}\right\|^q\right) \leq C$$

holds for $n \geq k+1$, where for vector $\mathbf{d} = (d_1,\ldots,d_k)'$, $\|\mathbf{d}\|^2 = \sum_{i=1}^k d_i^2$.

Proof of Theorem 1. In view of (2) and (3),

$$x_{n+h} - \hat{x}_{n+h}(k) = \sum_{j=0}^{h-1} b_j \varepsilon_{n+h-j} + \mathbf{x}_n'(k) A^{h-1}(k) \mathbf{a}(1,k) - \mathbf{x}_n'(k) \hat{A}_n^{h-1}(k) \hat{\mathbf{a}}_n(1,k). \qquad \textbf{(A.1)}$$

Some algebraic manipulations yield

$$\mathbf{x}_n'(k) A^{h-1}(k) \mathbf{a}(1,k) - \mathbf{x}_n'(k) \hat{A}_n^{h-1}(k) \hat{\mathbf{a}}_n(1,k) = P_h(k) + c_n, \qquad \textbf{(A.2)}$$

where $P_h(k) = -\mathbf{x}_n'(k) L_h(k)(\hat{\mathbf{a}}_n(1,k) - \mathbf{a}(1,k))$ and $c_n = 0$ for $h = 1$ and

$$|c_n| \leq C \sum_{l=2}^{h} \|\hat{\mathbf{a}}_n(1,k) - \mathbf{a}(1,k)\|^l \|\mathbf{x}_n(k)\| \qquad \textbf{(A.3)}$$

for $h \geq 2$. By Lemmas A.1 and A.3, Hölder's inequality, and (9), one has for $h \geq 2$ and all $l = 2, \ldots, h$,

$$\begin{aligned} E(\|\hat{\mathbf{a}}_n(1,k) - \mathbf{a}(1,k)\|^{2(l+1)}) &\leq E(\|\hat{R}_n^{-1}(k)\|^{2(l+1)q})^{1/q} E\left(\left\| \frac{1}{n-k} \sum_{j=k}^{n-1} \mathbf{x}_j \varepsilon_{j+1} \right\|^{2(l+1)q/(q-1)} \right)^{(q-1)/q} \\ &= O(n^{-l-1}), \end{aligned} \qquad \textbf{(A.4)}$$

where $q \geq \lfloor 2(h+1)/\delta \rfloor + 1$, and for real number a, $\lfloor a \rfloor$ denotes the largest integer $\leq a$. Moreover, because x_t can be expressed as

$$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j},$$

with b_j satisfying $b_j^2 \leq \nu e^{-\beta j}$ for some $\nu, \beta > 0$, by Lemma 2 of Wei (1987),

$$E(\|\mathbf{x}_n(k)\|^{2(l+1)}) \leq C \qquad \textbf{(A.5)}$$

for $l \leq h$. Now by (A.3)–(A.5) and Hölder's inequality,

$$E(c_n^2) = O(n^{-2}) \qquad \textbf{(A.6)}$$

for $h \geq 2$.

To deal with $P_h(k)$, first define

$$P_h^*(k) = -\mathbf{x}_n^{*\prime}(k) L_h(k) R^{-1}(k) \frac{1}{n-k} \sum_{j=k}^{g_n} \mathbf{x}_j(k) \varepsilon_{j+1}, \qquad \textbf{(A.7)}$$

where

$$\mathbf{x}_n^*(k) = \left(\sum_{j=0}^{(1/\beta)\log n} b_j \varepsilon_{n-j}, \ldots, \sum_{j=0}^{(1/\beta)\log n} b_j \varepsilon_{n-k+1-j} \right)'$$

and $g_n = n - (2/\beta)\log n - 2k - 2$. If one can show that

$$E(nP_h^{*^2}(k)) = \operatorname{tr}(R(k) L_h(k) R^{-1}(k) L_h'(k)) \sigma^2 + O(n^{-1/2}) \qquad \textbf{(A.8)}$$

and

$$E\{n(P_h^2(k) - P_h^{*^2}(k))\} = O(n^{-1/2}), \tag{A.9}$$

then the desired result (10) follows from (A.1), (A.2), (A.6), (A.8), (A.9), and the Cauchy–Schwarz inequality.

To obtain (A.8), first observe that

$$E(nP_h^{*^2}(k)) = \left(1 + O\left(\frac{\log n}{n}\right)\right)\operatorname{tr}(R^{-1}(k)L_h'(k)R^*(k)L_h(k))\sigma^2, \tag{A.10}$$

where $R^*(k) = E(\mathbf{x}_n^*(k)\mathbf{x}_n^{*'}(k))$. Simple algebra gives

$$\operatorname{tr}(R^{-1}(k)L_h'(k)R^*(k)L_h(k)) = \operatorname{tr}(R^{-1}(k)L_h'(k)R(k)L_h(k)) + O\left(\sum_{l=(1/\beta)\log n}^{\infty} b_l^2\right). \tag{A.11}$$

Because $\sum_{l=(1/\beta)\log n}^{\infty} b_l^2 = O(1/n)$, (A.8) follows from this, (A.10), and (A.11).

For the proof of (A.9), write

$$\begin{aligned}
\sqrt{(n-k)}(P_h(k) - P_h^*(k)) = {} & -(\mathbf{x}_n(k) - \mathbf{x}_n^*(k))' L_h(k)\hat{R}^{-1}(k)\frac{1}{\sqrt{(n-k)}}\sum_{j=k}^{n-1}\mathbf{x}_j(k)\varepsilon_{j+1} \\
& - \mathbf{x}_n^{*'}(k)L_h(k)(\hat{R}^{-1}(k) - R^{-1}(k))\frac{1}{\sqrt{(n-k)}}\sum_{j=k}^{n-1}\mathbf{x}_j(k)\varepsilon_{j+1} \\
& - \mathbf{x}_n^{*'}(k)L_h(k)R^{-1}(k)\frac{1}{\sqrt{(n-k)}}\sum_{j=g_n+1}^{n-1}\mathbf{x}_j(k)\varepsilon_{j+1} \\
\equiv {} & (\mathrm{I}) + (\mathrm{II}) + (\mathrm{III}).
\end{aligned} \tag{A.12}$$

By Hölder's inequality, Lemmas A.1 and A.2, an analogy with Lemma A.3, Lemma 2 of Wei (1987), and (9), the following equalities hold:

$$E\{(\mathrm{I})^2\} = O\left(\frac{1}{n}\right), \tag{A.13}$$

$$E\{(\mathrm{II})^2\} = O\left(\frac{1}{n}\right), \tag{A.14}$$

and

$$E\{(\mathrm{III})^2\} = O\left(\frac{\log n}{n}\right). \tag{A.15}$$

These equalities and (A.8) further imply that

$$E\{\sqrt{n}P_h^*(k)(\mathrm{I})\} = O(n^{-1/2}) \tag{A.16}$$

and

$$E\{\sqrt{n}P_h^*(k)(\mathrm{II})\} = O(n^{-1/2}). \tag{A.17}$$

In view of (A.13)–(A.17), (A.9) is guaranteed if we can show that

$$E\{\sqrt{n}P_h^*(k)(\mathrm{III})\} = O(n^{-1/2}). \tag{A.18}$$

Rewrite (III) as

$$\begin{aligned}
&- \mathbf{x}_n^{*\prime}(k)L_h(k)R^{-1}(k)\,\frac{1}{\sqrt{(n-k)}}\sum_{j=g_n+1}^{v_n-1}\mathbf{x}_j(k)\varepsilon_{j+1}\\
&- \mathbf{x}_n^{*\prime}(k)L_h(k)R^{-1}(k)\,\frac{1}{\sqrt{(n-k)}}\sum_{j=v_n}^{n-1}\mathbf{x}_j^*(k)\varepsilon_{j+1}\\
&- \mathbf{x}_n^{*\prime}(k)L_h(k)R^{-1}(k)\,\frac{1}{\sqrt{(n-k)}}\sum_{j=v_n}^{n-1}(\mathbf{x}_j(k)-\mathbf{x}_j^*(k))\varepsilon_{j+1}\\
&\qquad \equiv (\mathrm{IV}) + (\mathrm{V}) + (\mathrm{VI}),
\end{aligned} \tag{A.19}$$

where $v_n = n - (1/\beta)\log n - k - 1$. By an analogy with Lemma A.3 and Lemma 2 of Wei (1987),

$$E\left\| \frac{1}{\sqrt{(n-k)}}\sum_{j=v_n}^{n-1}(\mathbf{x}_j(k)-\mathbf{x}_j^*(k))\varepsilon_{j+1}\right\|^4 = O\left(\frac{(\log n)^2}{n^4}\right).$$

This and the Cauchy–Schwarz inequality yield

$$E((\mathrm{VI})^2) = O\left(\frac{\log n}{n^2}\right)$$

and hence

$$E\{\sqrt{n}P_h^*(k)(\mathrm{VI})\} = O\left(\frac{\sqrt{\log n}}{n}\right). \tag{A.20}$$

By independence, one also has

$$E(\sqrt{n}P_h^*(k)(\mathrm{IV})) = E(\sqrt{n}P_h^*(k)(\mathrm{V})) = 0. \tag{A.21}$$

Consequently, (A.18) follows from (A.19)–(A.21). ■

Proof of Theorem 2. Rearranging the order of summation, we have

$$\begin{aligned}
x_{n+h} - \check{x}_{n+h}(k) - \sum_{j=0}^{h-1} b_j\varepsilon_{n+h-j} &= -\mathbf{x}_n'(k)\hat{R}_n^{-1}(h,k)\,\frac{1}{n-k-h+1}\sum_{j=k}^{n-h}\mathbf{x}_j(k)\sum_{l=0}^{h-1}b_l\varepsilon_{j+h-l}\\
&= Q_h(k),
\end{aligned} \tag{A.22}$$

where

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

and

$$Q_h(k) = -\mathbf{x}_n'(k)\hat{R}_n^{-1}(h,k)\,\frac{1}{n-k-h+1}\sum_{j=k}^{n-1}\left(\sum_{l=0}^{h-1} b_l\,\mathbf{x}_{j+1+l-h}(k)\right)\varepsilon_{j+1}.$$

Note that in this proof, $\mathbf{x}_j(k)$ is set to $\mathbf{0}_k$, provided $j < k$ or $j > n-h$. By arguments similar to those used for verifying (A.8) and (A.9),

$$E(nQ_h^2(k)) = \operatorname{tr}\left\{R^{-1}(k)\operatorname{cov}\left(\sum_{l=0}^{h-1} b_l\,\mathbf{x}_{k+l}(k)\right)\right\}\sigma^2 + O(n^{-1/2}). \tag{A.23}$$

Consequently, (12) follows from (A.22) and (A.23). ■

Proof of Theorem 3. We first prove (18). By (D.3) of Kunitomo and Yamamoto (1985),

$$R^{-1}(k+1) = \begin{pmatrix} R^{-1}(k) & \mathbf{0}_k \\ \mathbf{0}_k' & 0 \end{pmatrix} + \frac{1}{\sigma^2}\,\mathbf{l}\mathbf{l}', \tag{A.24}$$

where $\mathbf{l}' = (a_k, \ldots, a_1, -1)$ is a $(k+1)$-dimensional vector. Also observe that

$$L_h(k+1) = \begin{pmatrix} L_h(k) & \boldsymbol{T}^* \\ \mathbf{0}_k' & b_{h-1} \end{pmatrix}, \tag{A.25}$$

where $\boldsymbol{T}^* = \sum_{j=0}^{h-2} b_j A^{h-2-j}(k)\mathbf{e}_k^*$ and $\mathbf{e}_k^* = (0,\ldots,0,1)'$ is a k-dimensional vector. Equalities (A.24) and (A.25) yield

$$f_{1,h}(k+1) = f_{1,h}(k) + \mathbf{l}'L_h'(k+1)R(k+1)L_h(k+1)\mathbf{l}. \tag{A.26}$$

Therefore, if (19) ($b_{h-1} \neq 0$) holds, then (18) follows from (A.26), $L_h(k+1)\mathbf{l} \neq \mathbf{0}_{k+1}$, and the positive definiteness of $R(k+1)$.

On the other hand, assume that (20) ($\mathbf{l}^* \neq \mathbf{0}_{k+1}$) holds. Notice that for $j \geq 0$,

$$A^j(k+1) = R^{-1}(k+1)E_j(k+1), \tag{A.27}$$

where

$$E_j(k+1) = \begin{pmatrix} r_j & r_{j-1} & \cdots & r_{j-k} \\ \vdots & \vdots & & \vdots \\ r_{k+j} & r_{k+j-1} & \cdots & r_j \end{pmatrix}.$$

One also has

$$E_j(k+1)\mathbf{l} = -\sigma^2(b_{j-k}, \ldots, b_j)'$$

for $j \geq 0$. These in turn imply that the second term on the right-hand side of (A.26) equals $\sigma^4\mathbf{l}^{*\prime}R^{-1}(k+1)\mathbf{l}^*$, which is positive according to (20) and the positive definiteness of $R^{-1}(k+1)$. As a result, (18) follows.

To show (21), first observe that

$$f_{2,h}(k+1) = \operatorname{tr}\left(R^{-1}(k+1)\sum_{l_1=0}^{h-1}\sum_{l_2=0}^{h-1} b_{l_1} b_{l_2} E_{|l_1-l_2|}(k+1)\right)\sigma^2$$

$$= \sigma^2 \sum_{l_1=0}^{h-1}\sum_{l_2=0}^{h-1} b_{l_1} b_{l_2} \operatorname{tr}(A^{|l_1-l_2|}(k+1)), \tag{A.28}$$

where the second equality is ensured by (A.27). Because $\operatorname{tr}(A^j(k+1)) = \operatorname{tr}(A^j(k))$ for $j \geq 1$ (see also Kunitomo and Yamamoto, 1985), (D.1), this fact and (A.28) give

$$f_{2,h}(k+1) = f_{2,h}(k) + \sum_{l=0}^{h-1} b_l^2 \sigma^2 > f_{2,h}(k),$$

as asserted. ■

The following technical lemma is needed for verifying Theorem 4. Its proof can be found in Ing and Wei (2002, Lemma 3).

LEMMA A.4. *Assume that* $\{x_t\}$ *satisfies model (1) with* $E(|\varepsilon_1|^{2q}) < \infty$ *for some* $q \geq 2$. *Then, for any pair of positive integers* (h, k) *with* $h \geq 1$ *and* $k < p$,

$$E\left(\left\| \frac{1}{\sqrt{n}} \sum_{j=k}^{n-h} \mathbf{x}_j(k)\{\mathbf{x}_j'(p)(\mathbf{a}(h,p) - \mathbf{a}_D(h,k))\}\right\|^q\right) \leq C$$

holds for sufficiently large n, where $\mathbf{a}_D(h,k)$*, defined in Theorem 4, is now viewed as a p-dimensional vector with undefined entries set to 0.*

Proof of Theorem 4. For $k < p$,

$$x_{n+h} - \check{x}_{n+h}(k) = \sum_{j=0}^{h-1} b_j \varepsilon_{n+h-j} + \mathbf{x}_n'(p)(\mathbf{a}(h,p) - \mathbf{a}_D(h,k))$$

$$- \mathbf{x}_n'(k)(\check{\mathbf{a}}_n(h,k) - \mathbf{a}_D(h,k)), \tag{A.29}$$

where $\mathbf{a}(h,p)$ is defined in Section 1 and $\mathbf{a}_D(h,k)$ in this proof is sometimes viewed as a p-dimensional vector with undefined entries set to 0. Express the last term on the right-hand side of (A.29) as

$$\mathbf{x}_n'(k)\hat{R}_n^{-1}(h,k)\,\frac{1}{n-h-k-1}\sum_{j=k}^{n-h}\mathbf{x}_j(k)\left\{\sum_{i=0}^{h-1} b_i \varepsilon_{j+h-i} + \mathbf{x}_j'(p)(\mathbf{a}(h,p) - \mathbf{a}_D(h,k))\right\}. \tag{A.30}$$

According to Hölder's inequality, the expectation of the square of (A.30) is bounded by

$$\left(E\left\| d_n \sum_{j=k}^{n-h}\mathbf{x}_j(k)\left\{\sum_{i=0}^{h-1} b_i \varepsilon_{j+h-i} + \mathbf{x}_j'(p)(\mathbf{a}(h,p) - \mathbf{a}_D(h,k))\right\}\right\|^{3q/(q-1)}\right)^{2(q-1)/3q}$$

$$\times (E\|\mathbf{x}_n(k)\|^{6q/(q-1)})^{(q-1)/3q}(E\|\hat{R}_n^{-1}(h,k)\|^{2q})^{1/q}, \tag{A.31}$$

where $d_n = (n-h-k-1)^{-1}$ and $q > 1 + (6/\delta)$. By Lemma A.4 and an analogy with Lemma A.3, the first component of (A.31) is of order $O(1/n)$, provided

$$E\{|\varepsilon_1|^{6+\delta}\} < \infty. \tag{A.32}$$

Moreover, by Lemma 2 of Wei (1987) and Lemma A.1, the remaining components of (A.31) are both of order $O(1)$ if (A.32) holds. Consequently, (A.31) is of order $O(1/n)$ because (A.32) is guaranteed by the assumption that $E\{|\varepsilon_1|^{4h+2+\delta}\} < \infty$ for some $\delta > 0$. This fact, the Cauchy–Schwarz inequality, and (A.29) further yield

$$\begin{aligned} E\{(x_{n+h} - \check{x}_{n+h}(k))^2\} - \sigma^2 \sum_{j=0}^{h-1} b_j^2 &= (\mathbf{a}(h,p) - \mathbf{a}_D(h,k))' R(p)(\mathbf{a}(h,p) - \mathbf{a}_D(h,k)) \\ &\quad + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{A.33}$$

For MSPE$P_h(k)$, we have

$$\begin{aligned} x_{n+h} - \hat{x}_{n+h}(k) &= \sum_{j=0}^{h-1} b_j \varepsilon_{n+h-j} + \mathbf{x}_n'(p)(\mathbf{a}(h,p) - \mathbf{a}_D(h,k)) + \mathbf{x}_n'(k)(\mathbf{a}_D(h,k) - \mathbf{a}(h,k)) \\ &\quad - \mathbf{x}_n'(k)(\hat{\mathbf{a}}_n(h,k) - \mathbf{a}(h,k)). \end{aligned} \tag{A.34}$$

By arguments similar to those used for proving (A.6) and (A.33), the expectation of the square of the last term on the right-hand side of (A.34) is of order $O(1/n)$. This result and the fact that the first three terms on the right-hand side of (A.34) are pairwise uncorrelated imply that

$$\begin{aligned} E(x_{n+h} - \hat{x}_{n+h}(k))^2 - \sigma^2 \sum_{j=0}^{h-1} b_j^2 &= (\mathbf{a}(h,p) - \mathbf{a}_D(h,k))' R(p)(\mathbf{a}(h,p) - \mathbf{a}_D(h,k)) \\ &\quad + (\mathbf{a}_D(h,k) - \mathbf{a}(h,k))' R(k) \\ &\quad \times (\mathbf{a}_D(h,k) - \mathbf{a}(h,k)) + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \tag{A.35}$$

In view of (A.33) and (A.35), (26) follows. To show (28), first observe that for $k < p$,

$$\mathbf{a}(h,k) = R^{-1}(k)(\gamma_h^*, \ldots, \gamma_{k+h-1}^*).$$

This identity, the definition of $\mathbf{a}_D(h,k)$, condition (27), and the positive definiteness of $R^{-1}(k)$ yield the desired property. ■

Remark A.1. By an argument similar to that used for verifying Theorem 3 of Ing and Wei (2002), $O(1/\sqrt{n})$ in (A.33) and (A.35) can be weakened to $O(1/n)$ under a slightly stronger moment condition. However, because the main concern of Theorem 4 is with the terms of order $O(1)$, pursuing this more delicate result does not seem necessary.

APPENDIX B

The h-step prediction model corresponding to model (29) is

$$
\begin{aligned}
x_{t+h} &= \tilde{\mathbf{x}}_t'(k)\mathbf{b}(h,k) + \sum_{i=0}^{h-1} b_i\,\varepsilon_{t+h-i} \\
&= \tilde{\mathbf{x}}_t'(k)B^{h-1}(k)(\beta_0,\ldots,\beta_{q_0},a_1,\ldots,a_k)' + \sum_{i=0}^{h-1} b_i\,\varepsilon_{t+h-i}, \qquad \textbf{(B.1)}
\end{aligned}
$$

where $k \geq p$,

$$
B(k) = \begin{pmatrix} B_1 & B_2(k) \\ \mathbf{0}_{k\times(q_0+1)} & A(k) \end{pmatrix},
$$

$B_2(k)$ with $k \geq p$ is the matrix $\hat{B}_{2,n}(k)$ with $\hat{\beta}_{i,n}$ replaced by β_i for $i = 0,\ldots,q_0$ (note that for $k < p$, the definition of $B_2(k)$ is slightly different; see the proof of Corollary 3, which follows, for details), and $\mathbf{b}(h,k) = B^{h-1}(k)(\beta_0,\ldots,\beta_{q_0},a_1,\ldots,a_k)'$ for $k \geq p$. Inspired by Chapter 16 of Hamilton (1994), which deals with parameter estimation and hypothesis testing problems under model (29) with $q_0 = 1$, we consider a linear transformation of $\tilde{\mathbf{x}}_t(k)$, $G(k)\tilde{\mathbf{x}}_t(k)$. Here,

$$
G(k) = \begin{pmatrix} I_{q_0+1} & \mathbf{0}_{(q_0+1)\times k} \\ T(k) & I_k \end{pmatrix}
$$

is a $(q_0+1+k)\times(q_0+1+k)$ matrix, and

$$
T(k) = \begin{pmatrix} T_{1,0} & \cdots & T_{1,q_0} \\ \vdots & & \vdots \\ T_{k,0} & \cdots & T_{k,q_0} \end{pmatrix}
$$

is a $k\times(q_0+1)$ matrix with $T_{i,j} = -\sum_{l=j}^{q_0}(-i)^{l-j}C_{l-j}^{l}\beta_l^*$ for $i = 1,\ldots,k$ and $j = 0,\ldots q_0$, and β_l^* being defined in (30). Observe that $G(k)\tilde{\mathbf{x}}_t(k) = \tilde{\mathbf{y}}_t(k) = (1, t+1,\ldots,(t+1)^{q_0}, y_t,\ldots,y_{t-k+1})'$, where y_t is defined in (30), and that

$$
(\beta_0,\ldots,\beta_{q_0})' = (\beta_0^*,\ldots,\beta_{q_0}^*)' + T'(k)\mathbf{a}(1,k) \qquad \textbf{(B.2)}
$$

for $k \geq p$.

Define the normalized sample covariance matrix $\widetilde{R}_n(k)$ as

$$
\begin{aligned}
\widetilde{R}_n(k) &= \frac{1}{n-k}\sum_{j=1}^{n-1} V_n(k)G(k)\tilde{\mathbf{x}}_j(k)\tilde{\mathbf{x}}_j'(k)G'(k)V_n'(k) \\
&= \frac{1}{n-k}\sum_{j=1}^{n-1} V_n(k)\tilde{\mathbf{y}}_j(k)\tilde{\mathbf{y}}_j'(k)V_n'(k),
\end{aligned}
$$

where

$$V_n(k) = \begin{pmatrix} D_n & \mathbf{0}_{(q_0+1)\times k} \\ \mathbf{0}_{k\times(q_0+1)} & I_k \end{pmatrix}$$

with $D_n = \operatorname{diag}(1, (n+1)^{-1}, \ldots, (n+1)^{-q_0})$ being a diagonal matrix of dimension $q_0 + 1$. The following two lemmas deal with moment properties of $\tilde{R}_n^{-1}(k)$.

LEMMA B.1. *Let the assumptions of Corollary 1 hold. Then, for any positive integer k and any $q > 0$,*

$$E(\|\tilde{R}_n^{-1}(k)\|^q) \le C$$

holds for sufficiently large n.

LEMMA B.2. *Let the assumptions of Corollary 1 hold. Then, for any positive integer k and any $q > 0$,*

$$E(\|\tilde{R}_n^{-1}(k) - \tilde{R}^{-1}(k)\|^q) \le Cn^{-q/2}$$

holds for sufficiently large n, where

$$\tilde{R}(k) = \begin{pmatrix} C_{q_0}^{*^{-1}} & \mathbf{0}_{(q_0+1)\times k} \\ \mathbf{0}_{k\times(q_0+1)} & R(k) \end{pmatrix}$$

with $C_{q_0}^*$ defined in Corollary 1 and $R(k)$ defined in Section 1.

Lemmas B.1 and B.2 can be shown by arguments similar to those used for verifying Lemma 1 and Theorem 2 of Ing and Wei (2002), respectively. We omit the details to save space. By analogy with Lemma A.3, Lemma B.3 is obtained as follows.

LEMMA B.3. *Assume that $\{x_t\}$ satisfies model (29) with $E(|\varepsilon_1|^q) < \infty$ for some $q \ge 2$. Then, for any positive integer k,*

$$E\left(\left\| \frac{1}{\sqrt{n}} \sum_{j=k}^{n-1} V_n(k)\tilde{\mathbf{y}}_j(k)\varepsilon_{j+1} \right\|^q\right) \le C$$

holds for $n \ge k + 1$.

Based on (B.2) and some tedious calculations, we obtain the following results, which play important roles in proving Corollary 1.

LEMMA B.4. *For $k \ge p$,*

$$G'^{-1}(k)B(k)G'(k) = \begin{pmatrix} B_1 & \mathbf{0}_{(q_0+1)\times k} \\ \mathbf{0}_{k\times(q_0+1)} & A(k) \end{pmatrix}.$$

Hence, for $k \ge p$ and $h \ge 1$,

$$G'^{-1}(k)B^{h-1}(k)G'(k) = \begin{pmatrix} B_1^{h-1} & \mathbf{0}_{(q_0+1)\times k} \\ \mathbf{0}_{k\times(q_0+1)} & A^{h-1}(k) \end{pmatrix}$$

and

$$\left\| (G'(k)V_n'(k))^{-1}B^{h-1}(k)G'(k)V_n'(k) - \begin{pmatrix} I_{q_0+1} & \mathbf{0}_{(q_0+1)\times k} \\ \mathbf{0}_{k\times(q_0+1)} & A^{h-1}(k) \end{pmatrix} \right\| = O\left(\frac{1}{n}\right).$$

Proof of Corollary 1. By (31), (B.1), Lemma B.4, and some algebraic manipulations, one has

$$E(x_{n+h} - \dot{x}_{n+h}(k))^2 = \sum_{i=0}^{h-1} b_i^2 \sigma^2 + E(\widetilde{P}_h(k) + \tilde{c}_n)^2, \tag{B.3}$$

where $\widetilde{P}_h(k) = -\tilde{\mathbf{x}}_n'(k)\widetilde{L}_h(k)(\dot{\mathbf{a}}_n(1,k) - \mathbf{b}(1,k))$, $\widetilde{L}_h(k) = \sum_{j=0}^{h-1} b_j B^{h-1-j}(k)$, $\tilde{c}_n = 0$ for $h = 1$, and

$$|\tilde{c}_n| \le C \sum_{l=1}^{h-1} \|\dot{\mathbf{a}}_n(1,k) - \mathbf{b}(1,k)\|^l \|\widetilde{R}_n^{-1}(k)\| \\ \times \left\| \frac{1}{n-k} \sum_{j=k}^{n-1} V_n(k)\tilde{\mathbf{y}}_j(k)\varepsilon_{j+1} \right\| \|V_n(k)\tilde{\mathbf{y}}_n(k)\| \tag{B.4}$$

for $h \ge 2$. Because

$$\|\dot{\mathbf{a}}_n(1,k) - \mathbf{b}(1,k)\| \le \|V_n'(k)\|\|G'(k)\|\|\widetilde{R}_n^{-1}(k)\| \left\| \frac{1}{n-k} \sum_{j=k}^{n-1} V_n(k)\tilde{\mathbf{y}}_j(k)\varepsilon_{j+1} \right\| \\ \le C\|\widetilde{R}_n^{-1}(k)\| \left\| \frac{1}{n-k} \sum_{j=k}^{n-1} V_n(k)\tilde{\mathbf{y}}_j(k)\varepsilon_{j+1} \right\|,$$

by Lemmas B.1 and B.3, (B.4), and Hölder's inequality,

$$E(\tilde{c}_n^2) \le C\frac{1}{n^2}. \tag{B.5}$$

According to Lemmas B.1–B.3 and arguments similar to those used for verifying (A.8) and (A.9), one has

$$nE(\widetilde{P}_h^2(k)) = \operatorname{tr}(\widetilde{R}_0(k)\widetilde{L}_h^*(k)\widetilde{R}^{-1}(k)\widetilde{L}_h^{*\prime}(k))\sigma^2 + O(n^{-1/2}), \tag{B.6}$$

where

$$\widetilde{R}_0(k) = E(V_n(k)\tilde{\mathbf{y}}_n(k)\tilde{\mathbf{y}}_n'(k)V_n'(k)) = \begin{pmatrix} \mathbf{1}_{q_0+1}\mathbf{1}_{q_0+1}' & \mathbf{0}_{(q_0+1)\times k} \\ \mathbf{0}_{k\times(q_0+1)} & R(k) \end{pmatrix}$$

and $\widetilde{L}_h^*(k) = (G'(k)V_n'(k))^{-1}\widetilde{L}_h(k)G'(k)V_n'(k)$. By Lemma B.4, it can be shown that

$$\left\| \widetilde{L}_h^*(k) - \begin{pmatrix} \sum_{i=0}^{h-1} b_i I_{q_0+1} & \mathbf{0}_{(q_0+1)\times k} \\ \mathbf{0}_{k\times(q_0+1)} & L_h(k) \end{pmatrix} \right\| = O(n^{-1}).$$

This fact, the definition of $\widetilde{R}_0(k)$, and (B.6) yield

$$nE(\widetilde{P}_h^2(k)) = \mathbf{1}'_{q_0+1} C^*_{q_0} \mathbf{1}_{q_0+1} \left(\sum_{i=0}^{h-1} b_i \right)^2 \sigma^2 + f_{1,h}(k) + O(n^{-1/2}). \quad \textbf{(B.7)}$$

Consequently, (34) follows from (B.3), (B.5), and (B.7). ■

Proof of Corollary 2. Equality (35) can be shown by arguments similar to those used for proving Theorem 2 and Corollary 1. The details are omitted. ■

Proof of Corollary 3. For the plug-in predictor, one has for $k < p$,

$$\begin{aligned} x_{n+h} - \dot{x}_{n+h}(k) - \sum_{i=0}^{h-1} b_i \varepsilon_{n+h-i} &= (\tilde{\mathbf{x}}'_n(p) B^{h-1}(p)\mathbf{b}(1,p) - \tilde{\mathbf{x}}'_n(k) B^{h-1}(k)\mathbf{b}(1,k)) \\ &\quad - (\tilde{\mathbf{x}}'_n(k)\hat{B}_n^{h-1}(k)\dot{\mathbf{a}}_n(1,k) - \tilde{\mathbf{x}}'_n(k)B^{h-1}(k)\mathbf{b}(1,k)), \end{aligned} \quad \textbf{(B.8)}$$

where the definition of $B(k)$ with $k < p$ and that with $k \geq p$ are the same, but the first column of $B_2(k)$, the upper-right $(q_0+1) \times k$ submatrix of $B(k)$, in the case of $k < p$ is $(\beta_0^*, \ldots, \beta_{q_0}^*)' + T'(k)\mathbf{a}(1,k)$ instead of $(\beta_0, \ldots, \beta_{q_0})'$, and $\mathbf{b}(1,k)$ with $k < p$ is a $(q_0 + 1 + k)$-dimensional vector defined by $((\beta_0^*, \ldots, \beta_{q_0}^*) + \mathbf{a}'(1,k)T(k), \mathbf{a}'(1,k))'$. Because it can be shown that Lemma B.4 still holds with $k < p$, one has

$$\tilde{\mathbf{x}}'_n(p)B^{h-1}(p)\mathbf{b}(1,p) - \tilde{\mathbf{x}}'_n(k)B^{h-1}(k)\mathbf{b}(1,k) = \mathbf{y}'_n(p)\mathbf{a}(h,p) - \mathbf{y}'_n(k)\mathbf{a}(h,k), \quad \textbf{(B.9)}$$

where $\mathbf{y}_j(l) = (y_j, \ldots, y_{j-l+1})'$. Moreover, by Lemmas A.4, B.1, and B.3 and an argument similar to that used for showing (B.5),

$$E(\tilde{\mathbf{x}}'_n(k)\hat{B}_n^{h-1}(k)\dot{\mathbf{a}}_n(1,k) - \tilde{\mathbf{x}}'_n(k)B^{h-1}(k)\mathbf{b}(1,k))^2 = O(n^{-1}). \quad \textbf{(B.10)}$$

As a result,

$$\begin{aligned} E(x_{n+h} - \dot{x}_{n+h}(k))^2 - \sum_{j=0}^{h-1} b_j^2 \sigma^2 &= (\mathbf{a}(h,p) - \mathbf{a}_D(h,k))'R(p)(\mathbf{a}(h,p) - \mathbf{a}_D(h,k)) \\ &\quad + (\mathbf{a}_D(h,k) - \mathbf{a}(h,k))'R(k)(\mathbf{a}_D(h,k) - \mathbf{a}(h,k)) \\ &\quad + O\left(\frac{1}{\sqrt{n}}\right) \end{aligned} \quad \textbf{(B.11)}$$

follows from (B.8)–(B.10) and an analogy with (A.35).

Similarly, it also can be shown that

$$\begin{aligned} E(x_{n+h} - \ddot{x}_{n+h}(k))^2 - \sum_{j=0}^{h-1} b_j^2 \sigma^2 &= (\mathbf{a}(h,p) - \mathbf{a}_D(h,k))'R(p)(\mathbf{a}(h,p) - \mathbf{a}_D(h,k)) \\ &\quad + O\left(\frac{1}{\sqrt{n}}\right). \end{aligned} \quad \textbf{(B.12)}$$

Consequently, Corollary 3 follows from (B.11), (B.12), and an argument similar to that used for verifying (28). ■