ACCUMULATED PREDICTION ERRORS, INFORMATION CRITERIA AND OPTIMAL FORECASTING FOR AUTOREGRESSIVE TIME SERIES

(RUNNING HEAD: MODEL SELECTION AND OPTIMAL FORECASTING)

Ching-Kang Ing∗
Academia Sinica and National Taiwan University

Abstract

The predictive capability of a modification of Rissanen's accumulated prediction error (APE) criterion, APE_{δ_n}, is investigated in infinite-order autoregressive (AR(∞)) models. Instead of accumulating squares of sequential prediction errors from the beginning, APE_{δ_n} is obtained by summing these squared errors from stage $n\delta_n$, where n is the sample size and $0 < \delta_n < 1$ may depend on n. Under certain regularity conditions, an asymptotic expression is derived for the mean-squared prediction error (MSPE) of an AR predictor with order determined by APE_{δ_n}. This expression shows that the prediction performances of APE_{δ_n} can vary dramatically depending on the choice of δ_n. Another interesting finding is that when δ_n approaches 1 at a certain rate, APE_{δ_n} can achieve asymptotic efficiency in most practical situations. An asymptotic equivalence between APE_{δ_n} and an information criterion with a suitable penalty term is also established from the MSPE point of view. It offers a new perspective for comparing the information- and prediction-based model selection criteria in AR(∞) models. Finally, we provide the first asymptotic efficiency result for the case when the underlying AR(∞) model is allowed to degenerate to a finite autoregression.

Key words and phrases: Accumulated prediction errors, Asymptotic equivalence, Asymptotic efficiency, Information criterion, Order selection, Optimal forecasting

---

⋆Institute of Statistical Science, Academia Sinica, 128, Academia Rd. Sec. 2, Taipei 115, Taiwan, R.O.C. E-mail: cking@stat.sinica.edu.tw

1. INTRODUCTION

In the past two decades, investigations on the accumulated prediction error (APE) (Rissanen, 1986) and its variations have attracted considerable attention among researchers from various disciplines. Prior to the early 1990s, a large number of studies focused on its consistency in selecting regression or time series models (e.g., Wax, 1988, Hannan et al., 1989, Hemerly and Davis, 1989, Wei, 1992, and Speed and Yu, 1993). However, since proving consistency requires assuming that the true model is included among the family of candidate models (which is rather difficult to justify in practice), recent research has focused more on understanding its statistical properties under possible model misspecification (e.g., Kavalieris, 1989, Wei, 1992, West, 1996, McCracken, 2000, Findley, 2005, Inoue and Kilian, 2005, among others). While a much deeper understanding of APE in cases of a misspecified model has been gained from these recent efforts, APE's prediction performance after model selection still remains unclear. This motivated the present study.

To select a model for the realization of a stationary time series, it is common to assume that the realization comes from an autoregressive moving-average (ARMA) process whose AR and MA orders are known to lie within prescribed finite intervals. Then a model selection procedure is used to select orders within these intervals and thereby determine a model for the data. However, as pointed out by Shibata (1980), Goldenshluger and Zeevi (2001) and Ing and Wei (2005), this assumption can rarely be justified in practice, and the less stringent assumption is that the time series data are observations from a linear stationary process. Following this idea, we assume in the sequel that observations $x_1, \cdots, x_n$ are generated by an AR(∞) process $\{x_t\}$, where

$$x_t + \sum_{i=1}^{\infty} a_i x_{t-i} = e_t, t = 0, \pm 1, \pm 2, \cdots, \tag{1.1}$$

with the characteristic polynomial $A(z) = 1 + \sum_{i=1}^{\infty} a_i z^i \neq 0$ for all $|z| \leq 1$ and $\{e_t\}$ being a sequence of independent random noises satisfying $E(e_t) = 0$ and $E(e_t^2) = \sigma^2$ for all t. To predict future observations, we consider a family of approximation models $\{AR(1), \cdots, AR(K_n)\}$, where the maximal order K_n is allowed to tend ∞ as n does in order to reduce approximation errors. In this framework, the APE value of model AR(k), $1 \leq k \leq K_n$, is given by

$$\text{APE}(k) = \sum_{i=m}^{n-1} (x_{i+1} - \hat{x}_{i+1}(k))^2, \tag{1.2}$$

where $\hat{x}_{i+1}(k) = -\mathbf{x}_i'(k)\hat{\mathbf{a}}_i(k)$, $\mathbf{x}_i(k) = (x_i, \cdots, x_{i-k+1})'$, $\hat{\mathbf{a}}_i(k)$ satisfies

$$-\hat{R}_i(k)\hat{\mathbf{a}}_i(k) = \frac{1}{i - K_n} \sum_{j=K_n}^{i-1} \mathbf{x}_j(k) x_{j+1}, \tag{1.3}$$

with

$$\hat{R}_i(k) = \frac{1}{i - K_n} \sum_{j=K_n}^{i-1} \mathbf{x}_j(k)\mathbf{x}_j'(k), \tag{1.4}$$

and $m \geq K_n + 1$ is the first integer j such that $\hat{\mathbf{a}}_j(K_n)$ is uniquely defined. As observed, APE(k) measures the performance of AR(k) when it is used for sequential predictions. Recently, a modification of APE,

$$\text{APE}_{\delta_n}(k) = \sum_{i=n\delta_n}^{n-1} (x_{i+1} - \hat{x}_{i+1}(k))^2, \tag{1.5}$$

with $0 < \delta_n < 1$ depending on n, has also been considered by several authors, e.g., West (1996), McCracken (2000) and Inoue and Kilian (2005). Since APE_{δ_n} includes the original APE as a special case, this paper focuses on APE_{δ_n}. As will be shown later, the performance of APE_{δ_n} can vary dramatically depending on the choice of δ_n.

In view of (1.5), it is natural to predict the next observation x_{n+1} using $\hat{x}_{n+1}(\hat{k}_{n,\delta_n})$, where

$$\hat{k}_{n,\delta_n} = \arg \min_{1 \leq k \leq K_n} \text{APE}_{\delta_n}(k). \tag{1.6}$$

This type of prediction, targeting future values of the observed time series, is referred to as a *same-realization prediction*. On the other hand, if the process used in estimation (or model selection) and that for prediction are independent, then it is called an *independent-realization prediction* (see Shibata, 1980, Bhansali, 1986, Karagrigoriou, 1997, and Schorfheide, 2005). For differences between these two types of predictions in various time series models, see Kunitomo and Yamamoto (1985), Ing (2001, 2003) and Ing and Wei (2003, 2005). The prediction performance of APE_{δ_n} after order selection is assessed using the mean-squared prediction error (MSPE) $q_n(\hat{k}_{n,\delta_n})$, where

$$q_n(k) = E\left(x_{n+1} - \hat{x}_{n+1}(k)\right)^2. \tag{1.7}$$

There are three interrelated issues addressed in this paper. The first one focuses on the asymptotic expression for $q_n(\hat{k}_{n,\delta_n})$. To deal with this problem, we derive an upper bound for the probability $P(\hat{k}_{n,\delta_n} = k)$ based on a new decomposition of APE_{δ_n} and some moment inequalities established in Appendix A; see (3.4) and Lemmas A.6-A.9. Motivated by Ing and Wei (2005), a condition, (3.7), is also introduced to handle the complicated dependent structures among the selected orders, estimated parameters and future observations. (Note that this difficulty does not exist for independent-realization predictions.) Consequently, an asymptotic expression for $q_n(\hat{k}_{n,\delta_n})$ is obtained in Theorem 1 when δ_n is bounded away from 1. A series of examples is given after Theorem 1 to illustrate its implications. In particular, it is shown in Example 1 that when the AR coefficients $\{a_i\}$ decay exponentially (which includes, but is not limited to, the ARMA(p,q) model with $q > 0$ as a special case) and δ_n satisfies $\log \delta_n^{-1} = o(\log n)$, APE_{δ_n} is asymptotically efficient in the sense that its (second-order) MSPE, $q_n(\hat{k}_{n,\delta_n}) - \sigma^2$, is ultimately not greater than $\min_{1\leq k\leq K_n} q_n(k) - \sigma^2$, the (second-order) MSPE of the best predictor among $\{\hat{x}_{n+1}(1), \cdots, \hat{x}_{n+1}(K_n)\}$. For the exact definition of *asymptotic efficiency*, see (2.3). However, if $\{a_i\}$ decay algebraically, Example 3 points out that APE_{δ_n} is no longer asymptotically efficient if δ_n is bounded away from 1. To alleviate this difficulty, Theorem 2 (also in Section 3) allows δ_n to converge to 1 at a suitable rate and offers a theoretical justification for the proposed modification. In light of Theorem 2, we were able to find a δ_n such that the corresponding APE_{δ_n} is asymptotically efficient in both exponential- and algebraic-decay cases, as detailed in Examples 4 and 5.

The second issue concerns the performances of the information criterion and its relation to APE_{δ_n} from the same-realization prediction point of view. The value of the information criterion for model AR(k) is defined by

$$IC_{P_n}(k) = \log \hat{\sigma}_n^2(k) + \frac{P_n k}{n}, \tag{1.8}$$

where $P_n > 1$ is a positive number (possibly) depending on n, log denotes the natural logarithm,

$$\hat{\sigma}_n^2(k) = \frac{1}{N} \sum_{t=K_n}^{n-1} \left(x_{t+1} + \hat{\mathbf{a}}_n'(k)\mathbf{x}_t(k) \right)^2, \tag{1.9}$$

and $N = n - K_n$. Note that AIC (Akaike, 1974), BIC (Schwarz, 1978) and HQ (Hannan and Quinn, 1979) correspond to IC_{P_n} with $P_n = 2, \log n$ and $c \log_2 n$, respectively, where $c > 2$ and $\log_2 n = \log(\log n)$. (1.8) is referred to as an AIC-like criterion if P_n is independent of n, and as a BIC-like criterion if $P_n \to \infty$ and $P_n = o(n)$. Theorem 3 (Section 4) gives an asymptotic expression for $q_n(\hat{k}_{n,P_n})$, where

$$\hat{k}_{n,P_n} = \arg \min_{1 \le k \le K_n} IC_{P_n}(k). \tag{1.10}$$

This result extends Corollary 1 if Ing and Wei (2005), which only focuses on the MSPE of AIC-like criteria. An interesting implication of Theorem 3 is that HQ is asymptotically efficient in the exponential-decay case whereas BIC is not; see Examples 7 and 8 in Section 4. While both HQ and BIC are known to be consistent in the finite-order AR model (Hannan and Quinn, 1979), these examples show that their prediction performances can remarkably differ in the AR(∞) case. Based on Theorems 1-3, an asymptotic equivalence between IC_{P_n} and APE_{δ_n}, with δ_n and P_n satisfying (4.15), is given at the end of Section 4; see (4.16). This type of equivalence, which concentrates on the two criteria's predictive capabilities, is a somewhat-different idea from the one considered in Kavalieris (1989), Hannan et al. (1989), and Wei (1992), which aimed to establish an algebraic connection between the two criteria. For a more detailed discussion, see Section 4.

The third issue we are interested in is a long-standing unresolved problem concerning time series model selection. When (1.1) does not degenerate to an AR model of finite order, Ing and Wei (2005) recently showed that AIC satisfies (2.3), and hence is asymptotically efficient for same-realization predictions. (For a related result in independent-realization settings, see Shibata, 1980.) However, if the *truly infinite order* assumption is violated, then, as mentioned previously, the BIC-like criteria (e.g., HQ and BIC) are consistent, but AIC, which tends to choose an overfitting model, does not possess this property (Shibata, 1976). Moreover, since Theorem 4 (Section 5) shows that BIC-like criteria can achieve (2.3) in the finite-order case, it becomes very challenging to determine a criterion for an optimal prediction when (1.1) is allowed to degenerate to a finite autoregression. To tackle this dilemma, in Section 5, we first consider an important special case where $\{a_i\}$ either decay exponentially or are zero for all but a finite number of i. Theorem 5 of Section 5 obtains an interesting result that $IC_{P_n}(k)$, with $P_n \to \infty$ and $P_n = o(\log n)$ and $APE_{\delta_n}(k)$, with $\delta_n^{-1} \to \infty$ and $\log \delta_n^{-1} = o(\log n)$, can simultaneously achieve asymptotic efficiency over these two types of AR processes. However, if the case where $\{a_i\}$ decay algebraically is also included, then the criteria proposed by Theorem 5 fail to preserve the same optimality. A two-stage procedure, (5.1), which is a hybrid between AIC and a BIC-like criterion, is provided as a remedy. Its validity is justified theoretically in Theorem 6 of Section 5. Concluding remarks are given in Section 6. For ease of reading, the proofs of the results in Sections 3-5 are deferred to Appendices A-C, respectively.

2. PRELIMINARY RESULTS

In this section, some preliminary results on the MSPE of AIC (and its variants) are introduced. We begin with a list of assumptions which are used throughout this paper.

(K.1) Let $\{ x_t \}$ be a linear process satisfying (1.1) with $A(z) = 1+a_1z+a_2z^2+\cdots \neq 0$ for $|\, z \,| \leq 1$. Furthermore, let coefficients $\{ a_i \}$ obey $\sum_{i=1}^{\infty} |i^{1/2}a_i| < \infty$.

(K.2) Let the distribution function of e_t be denoted by F_t. There are two arbitrarily small positive numbers, α and δ_0^*, and one arbitrarily large positive number, C_0, such that for all $t = \cdots, -1, 0, 1, \cdots$ and $|\, x-y \,| < \delta_0^*$,

$$|\, F_t(x) - F_t(y) \,| \leq C_0 \; |\, x-y \,|^{\alpha} \,.$$

(K.3) $\sup_{-\infty<t<\infty} E \,|\, e_t \,|^s < \infty, \qquad s = 1, 2, \cdots.$

(K.4) The maximal order K_n satisfies

$$C_l \leq \frac{K_n^{2+\delta_1^*}}{n} \leq C_u,$$

where δ_1^*, C_l, and C_u are some prescribed positive numbers.

(K.5) $a_n \neq 0$ for infinitely many n.

First note that the MSPE of $\hat{x}_{n+1}(k)$, $q_n(k)$ (see (1.7)), can be expressed as

$$\sigma^2 + E\; (\, \mathbf{f}(k) + \mathcal{S}(k) \,)^2 \,, \tag{2.1}$$

where

$$\begin{aligned}
\mathbf{f}(k) &= \mathbf{x}_n'(k)\, \hat{R}_n^{-1}(k)\, \frac{1}{N} \sum_{j=K_n}^{n-1} \mathbf{x}_j(k)\, e_{j+1,k}, \\
e_{j+1,k} &= x_{j+1} + \sum_{l=1}^{k} a_l(k)\, x_{j+1-l}, \\
(\, a_1(k), \cdots, a_k(k) \,)' &= \mathbf{a}(k) = \arg \min_{(c_1,\cdots,c_k)' \in R^k} E \left(x_{k+1} + \sum_{l=1}^{k} c_l\, x_{k+1-l} \right)^2,
\end{aligned}$$

and

$$\mathcal{S}(k) \;=\; \sum_{i=1}^{\infty} (\, a_i - a_i(k) \,)\; x_{n+1-i}$$

with $a_i(k) = 0$ for i $> k$. To simplify the notation, $\mathbf{a}(k)$ is sometimes viewed as an infinite-dimensional vector with undefined entries set to zero. Ing and Wei (2003, Theorem 3) obtained an asymptotic expression for $q_n(k) - \sigma^2$, which holds uniformly for all $1 \leq k \leq K_n$. This result is summarized in the following proposition.

Proposition 1. *Assume that* (K.1)–(K.4) *hold. Then,*

$$\lim_{n\to\infty} \max_{1\leq k\leq K_n} \left| \frac{q_n(k) - \sigma^2}{L_n(k)} - 1 \right| = 0, \tag{2.2}$$

where

$$L_n(k) = \frac{k\sigma^2}{N} + \|\, \mathbf{a} - \mathbf{a}(k) \,\|_R^2,$$

and for an infinite-dimensional vector $\mathbf{d} = (\, d_1, d_2, \cdots)'$,

$$\|\mathbf{d}\|_R^2 = \sum_{i\leq i,j\leq\infty} d_i d_j \gamma_{i-j}$$

with $\gamma_{i-j} = E(x_i x_j)$. *We also note that* $\|\, \mathbf{a} - \mathbf{a}(k) \,\|_R^2 = E(\mathcal{S}^2(k))$ *decreases as* k *increases.*

If one attempts to find an order k whose corresponding predictor, $\hat{x}_{n+1}(k)$, has the minimal MSPE, then some data-driven order selection criteria are needed. An order selection criterion, $\hat{k}_n$, is said to be asymptotically efficient if $\hat{x}_{n+1}(\hat{k}_n)$ satisfies

$$\limsup_{n\to\infty} \frac{q_n(\hat{k}_n) - \sigma^2}{\min\limits_{1\leq k\leq K_n} q_n(k) - \sigma^2} \leq 1, \tag{2.3}$$

where $1 \leq \hat{k}_n \leq K_n$. Inequality (2.3) says that the (second-order) MSPE of the predictor with order determined by an asymptotically efficient criterion is ultimately not greater than that of the best predictor among $\{\hat{x}_{n+1}(1), \cdots, \hat{x}_{n+1}(K_n)\}$. In view of (2.2), (2.3) is equivalent to

$$\limsup_{n\to\infty} \frac{q_n(\hat{k}_n) - \sigma^2}{L_n(k_n^*)} \leq 1, \tag{2.4}$$

where $L_n(k_n^*) = \min_{1\leq k\leq K_n} L_n(k)$.

When $\hat{k}_n$ is determined by FPE (Akaike, 1969), Mallow's C_p (Mallows, 1973), S_p (Hocking, 1976), AIC or $S_n(k)$ (Shibata, 1980), Ing and Wei (2005, Theorem 2) gave an asymptotic expression for $q_n(\hat{k}_n) - \sigma^2$; see, also, Proposition 2 below. Values of S_n, FPE, S_p and C_p for model AR(k) are defined by

$$\begin{aligned} S_n(k) &= (N + 2k)\hat{\sigma}_n^2(k), \\ \mathrm{FPE}(k) &= \left(\frac{n+k}{n-k} \right) \hat{\sigma}_n^2(k), \\ \mathrm{S}_p(k) &= (1 + \frac{k}{N-k-1})\hat{\hat{\sigma}}_n^2(k) \end{aligned}$$

and

$$C_p(k) = N\hat{\sigma}_n^2(k) - (N - 2k)\hat{\hat{\sigma}}_n^2(K_n),$$

respectively, where $\hat{\sigma}_n^2(k)$ is defined in (1.9) and

$$\hat{\hat{\sigma}}_n^2(k) = \left(\frac{N}{N-k} \right) \hat{\sigma}_n^2(k).$$

For later reference, we also define

$$\begin{aligned} \hat{k}_n^S &= \arg \min_{1 \le k \le K_n} S_n(k), \\ \hat{k}_n^A &= \arg \min_{1 \le k \le K_n} \mathrm{IC}_2(k), \\ \hat{k}_n^F &= \arg \min_{1 \le k \le K_n} \mathrm{FPE}(k), \\ \hat{k}_n^{\mathrm{S}_p} &= \arg \min_{1 \le k \le K_n} \mathrm{S}_p(k), \end{aligned}$$

and

$$\hat{k}_n^C = \arg \min_{1 \le k \le K_n} C_p(k).$$

It is worth noting that the main difficulty in analyzing the same-realization MSPE after order selection is that one must face the complicated dependent structures among the selected orders, estimated parameters and future observations. To tackle this difficulty, Ing and Wei (2005) imposed the following assumption on $L_n(k)$.

(K.6). For every exponent $\xi > 0$, there is a nonnegative exponent $\theta = \theta(\xi)$, $0 \le \theta < 1$, such that for all large n and all $k \in A_{n,\theta} = \{ k : 1 \le k \le K_n, \ | k - k_n^* | \ge {k_n^*}^{\theta} \}$,

$$ {k_n^*}^{\xi} \frac{N \left(L_n(k) - L_n \left(k_n^* \right) \right)}{| k - k_n^* |} \ge \bar{C} > 0, \tag{2.5}$$

where $\bar{C}$ is some positive number independent of n and ξ.

If $\{x_t\}$ is an AR process of finite order, then (2.5) automatically holds . When $a_i \neq 0$ for infinitely many i, Examples 1 and 2 of that same paper (Ing and Wei, 2005) also show that (2.5) is fulfilled in the following cases: (a) the exponential-decay case,

$$C_1 k^{-\theta_1} e^{-\beta k} \leq \| \mathbf{a} - \mathbf{a}(k) \|_R^2 \leq C_2 k^{\theta_1} e^{-\beta k}, \tag{2.6}$$

where $C_2 \geq C_1 > 0$, $\theta_1 \geq 0$, and $\beta > 0$ (note that (K.1) yields that (2.6) is equivalent to $C_1^* k^{-\theta_1} e^{-\beta k} \leq \sum_{i \geq k} a_i^2 \leq C_2^* k^{\theta_1} e^{-\beta k}$ for some $C_2^* \geq C_1^* > 0$); and (b) the algebraic-decay case,

$$(C_3 - M_1 k^{-\xi_1}) k^{-\beta} \leq \| \mathbf{a} - \mathbf{a}(k) \|_R^2 \leq (C_3 + M_1 k^{-\xi_1}) k^{-\beta}, \tag{2.7}$$

where $C_3, M_1 > 0$, $\xi_1 \geq 2$ and $\beta > 1 + \delta_1^*$ (note that δ_1^* is defined in (K.4)). These facts reveal that (2.5) is quite reasonable from both practical and theoretical points of view, since it includes the ARMA model (which is the most used short-memory time series model by far) and the AR(∞) model with algebraically decaying coefficients (which is of much theoretical interest in the context of model selection) as special cases. Now, Proposition 2 is stated as follows.

Proposition 2. *Assume that* (K.1)–(K.6) *hold. Then*

$$\lim_{n \to \infty} \frac{q_n(\hat{k}_n) - \sigma^2}{L_n(k_n^*)} = 1,$$

where $\hat{k}_n = \hat{k}_n^S, \hat{k}_n^A, \hat{k}_n^F, \hat{k}_n^{S_p}$, *or* $\hat{k}_n^C$.

As an immediate consequence of Proposition 2, we obtain that AIC, FPE, $S_n(k)$, $\mathrm{S}_p(k)$ and C_p are all asymptotically efficient in the sense of (2.3).

3. MSPE of APE$_{\delta_n}$ in AR(∞) processes

This section provides asymptotic expressions for $q_n(\hat{k}_{n,\delta_n}) - \sigma^2$, where $\hat{k}_{n,\delta_n}$ is defined in (1.6). Without loss of generality, in the rest of this paper, $n\delta_n$ is assumed to be a positive integer. First note that for $0 < \delta_n < 1$,

$$\mathrm{APE}_{\delta_n}(k) = \sum_{i=n\delta_n}^{n-1} (x_{i+1} + \mathbf{x}_i'(k) \hat{\mathbf{a}}_i(k))^2 = \sum_{i=n\delta_n}^{n-1} \{e_{i+1} + \hat{e}_{i,k} + (e_{i+1,k} - e_{i+1})\}^2, \tag{3.1}$$

where $\hat{e}_{i,k} = \mathbf{x}_i'(k)(\hat{\mathbf{a}}_i(k) - \mathbf{a}(k))$ and $e_{i+1,k}$ is defined after (2.1). Following Lai and Wei (1982, (2.7)),

$$\begin{aligned}\sum_{i=n\delta_n}^{n-1} \hat{e}_{i,k}^2 &= \sum_{i=n\delta_n}^{n-1} h_i(k) e_{i+1,k}^2 + Q_{n\delta_n}(k) - Q_n(k) + \sum_{i=n\delta_n}^{n-1} h_i(k)\hat{e}_{i,k}^2 \\ &- 2\sum_{i=n\delta_n}^{n-1} (1 - h_i(k))\hat{e}_{i,k} e_{i+1,k}, \end{aligned} \tag{3.2}$$

where

$$h_i(k) = \mathbf{x}_i'(k)(\sum_{j=K_n}^{i} \mathbf{x}_j(k)\mathbf{x}_j'(k))^{-1}\mathbf{x}_i(k)$$

and

$$Q_i(k) = (\sum_{j=K_n}^{i-1} \mathbf{x}_j(k) e_{j+1,k})'(\sum_{j=K_n}^{i-1} \mathbf{x}_j(k)\mathbf{x}_j'(k))^{-1}(\sum_{j=K_n}^{i-1} \mathbf{x}_j(k) e_{j+1,k}).$$

On substituting (3.2) into (3.1), one obtains

$$\begin{aligned}\mathrm{APE}_{\delta_n}(k) = \sum_{i=n\delta_n}^{n-1} e_{i+1}^2 + \sum_{i=n\delta_n}^{n-1} h_i(k) e_{i+1,k}^2 + Q_{n\delta_n}(k) - Q_n(k) + \sum_{i=n\delta_n}^{n-1} h_i(k)\hat{e}_{i,k}^2 \\ + \quad 2\sum_{i=n\delta_n}^{n-1} h_i(k)\hat{e}_{i,k} e_{i+1,k} + \sum_{i=n\delta_n}^{n-1} (e_{i+1,k} - e_{i+1})^2 + 2\sum_{i=n\delta_n}^{n-1} (e_{i+1,k} - e_{i+1}) e_{i+1}. \end{aligned} \tag{3.3}$$

Define

$$L_n^{(\delta_n)}(k) = \frac{k\sigma^2 \log \delta_n^{-1}}{N(1-\delta_n)} + \|\mathbf{a} - \mathbf{a}(k)\|_R^2,$$

and

$$k_{n,\delta_n}^* = \arg \min_{1 \le k \le K_n} L_n^{(\delta_n)}(k).$$

(Note that $(\log \delta_n^{-1})/(1-\delta_n) > 1$ as $0 < \delta_n < 1$.) As one of the main technical contributions of this paper, we obtain for $k \neq k_{n,\delta_n}^*$,

$$\begin{aligned} & P(\hat{k}_{n,\delta_n} = k) \\ \le \quad & P\left(\frac{1}{N(1-\delta_n)L_n^{(\delta_n)}(k)}\mathrm{APE}_{\delta_n}(k) \le \frac{1}{N(1-\delta_n)L_n^{(\delta_n)}(k)}\mathrm{APE}_{\delta_n}(k_{n,\delta_n}^*)\right) \\ \le \quad & \sum_{l=1}^{12} P\left(|N_l(k)| \ge \frac{1}{12} V_{n,\delta_n}\right), \end{aligned} \tag{3.4}$$

where

$$|N_1(k)| = \frac{\left|\sum_{i=n\delta_n}^{n-1} h_i(k)e_{i+1,k}^2 - k\sigma^2 \log \delta_n^{-1}\right|}{N(1-\delta_n)L_n^{(\delta_n)}(k)},$$

$$|N_2(k)| = \frac{\left|\sum_{i=n\delta_n}^{n-1} h_i(k_{n,\delta_n}^*)e_{i+1,k_{n,\delta_n}^*}^2 - k_{n,\delta_n}^*\sigma^2 \log \delta_n^{-1}\right|}{N(1-\delta_n)L_n^{(\delta_n)}(k)},$$

$$|N_3(k)| = \frac{|Q_{n\delta_n}(k) - k\sigma^2|}{N(1-\delta_n)L_n^{(\delta_n)}(k)}, \qquad |N_4(k)| = \frac{|Q_{n\delta_n}(k_{n,\delta_n}^*) - k_{n,\delta_n}^*\sigma^2|}{N(1-\delta_n)L_n^{(\delta_n)}(k)},$$

$$|N_5(k)| = \frac{|Q_n(k) - k\sigma^2|}{N(1-\delta_n)L_n^{(\delta_n)}(k)}, \qquad |N_6(k)| = \frac{|Q_n(k_{n,\delta_n}^*) - k_{n,\delta_n}^*\sigma^2|}{N(1-\delta_n)L_n^{(\delta_n)}(k)},$$

$$|N_7(k)| = \frac{\left|\sum_{i=n\delta_n}^{n-1} h_i(k)\hat{e}_{i,k}^2\right|}{N(1-\delta_n)L_n^{(\delta_n)}(k)}, \qquad |N_8(k)| = \frac{\left|\sum_{i=n\delta_n}^{n-1} h_i(k_{n,\delta_n}^*)\hat{e}_{i,k_{n,\delta_n}^*}^2\right|}{N(1-\delta_n)L_n^{(\delta_n)}(k)},$$

$$|N_9(k)| = \frac{2\left|\sum_{i=n\delta_n}^{n-1} h_i(k)\hat{e}_{i,k}e_{i+1,k}\right|}{N(1-\delta_n)L_n^{(\delta_n)}(k)}, |N_{10}(k)| = \frac{2\left|\sum_{i=n\delta_n}^{n-1} h_i(k_{n,\delta_n}^*)\hat{e}_{i,k_{n,\delta_n}^*}e_{i+1,k_{n,\delta_n}^*}\right|}{N(1-\delta_n)L_n^{(\delta_n)}(k)},$$

$$|N_{11}(k)| = \frac{\left|\sum_{i=n\delta_n}^{n-1} \left\{(e_{i+1,k} - e_{i+1})^2 - (e_{i+1,k_{n,\delta_n}^*} - e_{i+1})^2 - \|\mathbf{a} - \mathbf{a}(k)\|_R^2 + \|\mathbf{a} - \mathbf{a}(k_{n,\delta_n}^*)\|_R^2\right\}\right|}{N(1-\delta_n)L_n^{(\delta_n)}(k)},$$

$$|N_{12}(k)| = \frac{\left|\sum_{i=n\delta_n}^{n-1} (e_{i+1,k} - e_{i+1,k_{n,\delta_n}^*})e_{i+1}\right|}{N(1-\delta_n)L_n^{(\delta_n)}(k)}, \text{ and } V_{n,\delta_n}(k) = \frac{L_n^{(\delta_n)}(k) - L_n^{(\delta_n)}(k_{n,\delta_n}^*)}{L_n^{(\delta_n)}(k)}.$$

By (3.4), Chebyshev's inequality, and moment bounds for $|N_i|, i = 1, \cdots, 12$ (to be established in Appendix A), an upper bound for $P(\hat{k}_{n,\delta_n} = k)$ can be obtained. This upper bound plays an important role in verifying the main results of this section, Theorems 1 and 2. When δ_n is bounded away from 1, Theorem 1 below provides an asymptotic expression for $q_n(\hat{k}_{n,\delta_n}) - \sigma^2$.

Theorem 1. *Assume that (K.1)–(K.5) hold and δ_n satisfies*

$$\limsup_{n\to\infty} \delta_n < 1 \tag{3.5}$$

and

$$\liminf_{n\to\infty} n^{\theta_3}\delta_n > 0, \tag{3.6}$$

where $0 < \theta_3 < \delta_1^*/(2+\delta_1^*)$ *(recall that* δ_1^* *is defined in (K.4)). Moreover, assume that the following conditions hold:*

(i) for every exponent $\xi > 0$*, there is a nonnegative exponent* $0 \le \theta = \theta(\xi) < 1$ *and a positive number* $M = M(\xi) > 0$ *such that*

$$\liminf_{n\to\infty} \min_{k \in A^{(\delta_n)}_{n,\theta,M}} \left(k^*_{n,\delta_n}\right)^{\xi} \frac{N\left(L_n^{(\delta_n)}(k) - L_n^{(\delta_n)}(k^*_{n,\delta_n})\right)}{\frac{\log \delta_n^{-1}}{1-\delta_n}\left|k - k^*_{n,\delta_n}\right|} > 0, \tag{3.7}$$

where

$$A^{(\delta_n)}_{n,\theta,M} = \left\{k:\ 1 \le k \le K_n,\ \left|k - k^*_{n,\delta_n}\right| \ge M\left(k^*_{n,\delta_n}\right)^{\theta}\right\};\ \textit{and}$$

(ii)

$$\lim_{n\to\infty} \frac{\log \delta_n^{-1}}{(k^*_{n,\delta_n})^{\eta}} = 0, \tag{3.8}$$

where η *is some positive number satisfying* $\eta = 1$ *if* $\theta = 0$ *and* $0 < \eta < 1-\theta$ *if* $0 < \theta < 1$*, and* $0 \le \theta = \theta(\xi) < 1$ *is obtained from (i) when* ξ *is limited to the open interval* $(0, \min\{1/2, \{(2+\delta_1^*)(1-\theta_3)/2\} - 1\})$.

Then,

$$\lim_{n\to\infty} \frac{q_n(\hat{k}_{n,\delta_n}) - \sigma^2}{L_n(k^*_{n,\delta_n})} = 1. \tag{3.9}$$

Remark 1. If for any $\xi > 0$, (3.7) holds for $\theta = 0$ and $M = 1$, then it can be shown that (3.9) is still valid without condition (3.8). When δ_n decreases to 0 at a polynomial rate, this finding can be used to illustrate the deficiency of APE_{δ_n} in situations where the AR coefficients decay exponentially fast; see Example 2 below for more details. □

Remark 2. Since (K.5) is assumed, it is not difficult to see that $k^*_{n,\delta_n} \to \infty$ as $n \to \infty$. Therefore, when $0 < \delta_n = \delta < 1$ is independent of n, (3.8) automatically holds. □

The following examples help gain further insights into Theorem 1.

Example 1. Assume that the AR coefficients satisfy

$$C_1 e^{-\beta k} \leq \|\mathbf{a} - \mathbf{a}(k)\|_R^2 \leq C_2 e^{-\beta k}, \tag{3.10}$$

where $0 < C_1 \leq C_2 < \infty$ and $\beta > 0$. (3.10) is fulfilled by any causal and invertible ARMA(p, q) model with $q > 0$. In this example, we shall show how to choose δ_n to attain (2.3) under (3.10).

Let δ_n satisfy (3.5) and

$$\log \delta_n^{-1} = o(\log n), \tag{3.11}$$

which guarantees (3.6). It is easy to see that (3.5) and (3.11) are satisfied by $0 < \delta_n = \delta < 1$, or $\delta_n^{-1} \to \infty$, with $\delta_n^{-1}/n^{\nu} \to 0$ for any $\nu > 0$. (3.10) implies that for some $C_1 > 0$,

$$\frac{1}{\beta} \log n - \frac{1}{\beta} \log \left(\frac{\log \delta_n^{-1}}{1 - \delta_n} \right) - C_1 \leq k_{n,\delta_n}^* \leq \frac{1}{\beta} \log n - \frac{1}{\beta} \log \left(\frac{\log \delta_n^{-1}}{1 - \delta_n} \right) + C_1, \tag{3.12}$$

and for any $\xi > 0$, (3.7) holds for $\theta = 0$ and some $M > 0$. Therefore, condition (i) of Theorem 1 follows and θ in condition (ii) of Theorem 1 can be chosen to be 0. Moreover, since (3.11) and (3.12) yield $\log \delta_n^{-1}/k_{n,\delta_n}^* \to 0$, condition (ii) of Theorem 1 is ensured. Consequently, (3.9) holds for those values of δ_n which satisfy (3.5) and (3.11). This result, Proposition 2 and the fact that

$$\lim_{n \to \infty} \frac{L_n(k_n^*)}{L_n(k_{n,\delta_n}^*)} = 1 \tag{3.13}$$

(which is guaranteed by (3.10)-(3.12)) lead to the conclusion that APE$_{\delta_n}$, with δ_n satisfying (3.5) and (3.11), is asymptotically efficient under (3.10). □

Example 2. This example is given to indicate that if δ_n decays to 0 at a polynomial rate, then APE$_{\delta_n}$ cannot be asymptotically efficient even in the exponential-decay case. More specifically, assume that

$$\delta_n = C_1 n^{-\theta_3}, \tag{3.14}$$

where $C_1 > 0$ and $0 < \theta_3 < \delta_1^*/(2 + \delta_1^*)$, and the AR coefficients obey a special case of (3.10),

$$\|\mathbf{a} - \mathbf{a}(k)\|_R^2 = C_2 e^{-\beta k}(1 + O(k^{-1})), \tag{3.15}$$

where C_2 and β are some positive numbers. These assumptions yield that for any $\xi > 0$, (3.7) holds for $\theta = 0$ and $M = 1$, and hence, by Remark 1, (3.9) follows. Since under (3.14) and (3.15),

$$\liminf_{n\to\infty} \frac{L_n(k_{n,\delta_n})}{L_n(k_n^*)} > 1, \tag{3.16}$$

APE_{δ_n}, with δ_n satisfying (3.14), fails to achieve (2.3) in the exponential-decay case. □

Example 3. This example investigates prediction performances of APE_{δ_n} in the algebraic-decay case (2.7). When (2.7), (3.5) and (3.6) are assumed, the same argument as the one in Example 2 of Ing and Wei (2005) yields that

$$k^*_{n,\delta_n} = \left(\frac{(1-\delta_n)NC_3\beta}{(\log \delta_n^{-1})\sigma^2}\right)^{1/(\beta+1)} + O(1), \tag{3.17}$$

and for any $\xi > 0$, (3.7) holds for $1 - \min\{\xi, 1\} < \theta < 1$ and some $M > 0$. These facts, together with (3.6), guarantee that conditions (i) and (ii) of Theorem 1 hold. As a result, (3.9) is ensured by Theorem 1. Moreover, since (2.7), (3.5) and (3.6) also imply (3.16), APE_{δ_n} is not asymptotically efficient in this case. □

While Example 3 shows that APE_{δ_n} with δ_n bounded away from 1 cannot be asymptotically efficient in the algebraic-decay case, we have found that as $\delta_n \to 1$,

$$\lim_{n\to\infty} \frac{L_n(k^*_{n,\delta_n})}{L_n(k_n^*)} = 1 \tag{3.18}$$

is always true. This observation and Theorem 1 led us to ask whether the difficulty of APE_{δ_n} mentioned in Example 3 can be alleviated by letting $\delta_n \to 1$ at a suitable rate. This question is answered in Theorem 2 and some examples following it.

Theorem 2. *Assume that (K.1)–(K.5) hold and δ_n satisfies $0 < \delta_n < 1$ and $\lim_{n\to\infty} \delta_n = 1$. Also assume that condition (i) of Theorem 1 holds. Moreover, (3.9) follows if k^*_{n,δ_n} and δ_n meet one of the following conditions:*

(i) $\lim_{n\to\infty} k^*_{n,\delta_n}/n^{\theta_3} = 0$ for any $\theta_3 > 0$ and $(1-\delta_n)^{-1} = O(k^{*\xi_2}_{n,\delta_n})$ for some $0 < \xi_2 < 1/2$; or

(ii) $(1-\delta_n)^{-1} = O(k^{*\xi_2}_{n,\delta_n})$ for some $0 < \xi_2 < \min\{1/2, \delta_1^*/2\}$.

This result, (3.18) and Proposition 2 together imply that APE_{δ_n} is asymptotically efficient and equivalent to AIC.

In light of Theorem 2, the following examples demonstrate how to choose δ_n such that the resulting APE_{δ_n} is asymptotically efficient in both exponential- and algebraic-decay cases.

Example 4. Assume that the AR coefficients obey (2.6). Although Example 1 shows that when θ_1 in (2.6) is equal to 0, APE_{δ_n}, with δ_n satisfying (3.5) and (3.11), is asymptotically efficient, it is unclear whether this result still holds for $\theta_1 > 0$. Fortunately, this difficulty can be bypassed by letting

$$\delta_n = 1 - C_1(\log n)^{-r}, \tag{3.19}$$

with $C_1 > 0$ and $0 < r < 1/2$. First note that when (2.6) is true, the same argument as in Example 1 of Ing and Wei (2005) yields that for some $C_2 > 0$,

$$\frac{1}{\beta}\log n - C_2 \log_2 n \le k^*_{n,\delta_n} \le \frac{1}{\beta}\log n + C_2 \log_2 n, \tag{3.20}$$

and for any $\xi > 0$, (3.7) holds for any $0 < \theta < 1$ and some $M > 0$. Therefore, condition (i) of Theorem 1 follows. Moreover, since condition (i) of Theorem 2 is ensured by (3.19) and (3.20), (3.9) is now guaranteed by Theorem 2. Consequently, APE_{δ_n}, with δ_n satisfying (3.19), attains asymptotic efficiency under (2.6). □

Example 5. This example shows that if

$$\delta_n = 1 - C_3(\log n)^{-r}, \tag{3.21}$$

where C_3 and r are any positive numbers, then the corresponding APE_{δ_n} is asymptotically efficient under the algebraic-decay case (2.7). To see this, first note that following the same line of reasoning as in Example 2 of Ing and Wei (2005), (3.17) is still valid here and for any $\xi > 0$, (3.7) holds for $1 - \min\{\xi, 1\} < \theta < 1$ and some $M > 0$. In addition, since condition (ii) of Theorem 2 is ensured by (3.17) and (3.21), the desired result follows from Theorem 2. □

Examples 4 and 5 suggest that to achieve asymptotic efficiency through APE_{δ_n} in both exponential- as well as algebraic-decay cases, δ_n can be chosen to satisfy (3.19). However, the question of how to determine C_1 and r in (3.19) seems difficult to answer from a finite sample point of view. Further investigations in this direction are still needed. We close this section with two remarks concerning the performances of APE_{δ_n} in finite-order AR models and for independent-realization predictions.

Remark 3. When (1.1) degenerates to an AR(p_0) model with $1 \le p_0 < \infty$, it can be shown that $\hat{k}_{n,\delta_n}$, with $\liminf_{n\to\infty} \delta_n > 0$, which tends to choose an overfitting model, is not a consistent estimator of p_0 (see, e.g., Inoue and Kilian, 2005). On the other hand, if $\delta_n \to 0$ at a certain rate, then (C.5) of Appendix C yields that the corresponding APE_{δ_n} is consistent. Since these results and Theorem 2 offer totally different suggestions for choosing δ_n, it becomes very challenging to achieve asymptotic efficiency through APE_{δ_n} when (1.1) is allowed to degenerate to a finite autoregression. In Section 5, some selection criteria to remedy this difficulty are proposed. □

Remark 4. The APE_{δ_n} described in Theorem 2 is also asymptotically efficient for independent-realization predictions (for the definition of asymptotic efficiency in independent-realization settings, see Shibata, 1980, Bhansali, 1986, and Karagrigoriou, 1997). As argued in Ing and Wei (2005, (5.37)), this property can be ensured by showing that

$$\text{p-lim}_{n\to\infty} \frac{L_n(\hat{k}_{n,\delta_n})}{L_n(k_n^*)} - 1 = 0. \tag{3.22}$$

To obtain (3.22), first note that Lemmas A.6-A.9 of Appendix A ensure that for $i = 1, \cdots, 12$

$$\text{p-lim}_{n\to\infty} \max_{1\le k\le K_n} |N_i(k)| = 0,$$

which further implies that

$$\text{p-lim}_{n\to\infty} \frac{L_n^{(\delta_n)}(\hat{k}_{n,\delta_n})}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1 = 0. \tag{3.23}$$

Since as $\delta_n \to 1$,

$$\max_{1\le k\le K_n} \frac{L_n^{(\delta_n)}(k)}{L_n(k)} \to 1, \tag{3.24}$$

(3.22) follows from (3.18), (3.23) and (3.24). □

4. MSPE of IC_{P_n} in AR(∞) processes.

In this section, prediction performances of the information criterion, $IC_{P_n}(k)$ (see (1.8)), are investigated. When $P_n > 1$ is independent of n, Ing and Wei (2005, Corollary 1) obtained an asymptotic expression for $q_n(\hat{k}_{n,P_n}) - \sigma^2$, where $\hat{k}_{n,P_n}$ (see (1.9)) is the minimizer of $IC_{P_n}(k)$, with $1 \le k \le K_n$ and K_n satisfying (K.4).

Theorem 3 below extends Ing and Wei's result to the case where P_n is allowed to tend to ∞ with n. To introduce Theorem 3, we need to define

$$k^*_{n,P_n} = \arg \min_{1 \le k \le K_n} L_{n,P_n}(k), \tag{4.1}$$

where

$$L_{n,P_n}(k) = \frac{(P_n - 1)k\sigma^2}{N} + \|\mathbf{a} - \mathbf{a}(k)\|^2_R. \tag{4.2}$$

Theorem 3. *Let (K.1)–(K.5) hold and P_n satisfy*

$$\liminf_{n\to\infty} P_n > 1, \tag{4.3}$$

and

$$P_n = O(n^{\theta_3}), \tag{4.4}$$

where $0 < \theta_3 < \delta^*_1/(4 + 2\delta^*_1)$. Moreover, assume that the following conditions hold:

(i) for every exponent $\xi > 0$, there is a nonnegative exponent $0 \le \theta = \theta(\xi) < 1$ and a positive number $M = M(\xi) > 0$ such that

$$\liminf_{n\to\infty} \min_{k \in A_{P_n,\theta,M}} (k^*_{n,P_n})^{\xi} \frac{N\left(L_{n,P_n}(k) - L_{n,P_n}(k^*_{n,P_n})\right)}{(P_n - 1)|k - k^*_{n,P_n}|} > 0, \tag{4.5}$$

where $A_{P_n,\theta,M} = \{k : 1 \le k \le K_n, |k - k^*_{n,P_n}| \ge M(k^*_{n,P_n})^{\theta}\}$; and

(ii)

$$\lim_{n\to\infty} \frac{P_n - 1}{(k^*_{n,P_n})^{\eta}} = 0, \tag{4.6}$$

where η is some positive number satisfying $\eta = 1$ if $\theta = 0$ and $0 < \eta < 1 - \theta$ if $\theta > 0$, and $0 \le \theta = \theta(\xi) < 1$ is obtained from (i) when ξ is limited to the open interval $(0, \{\delta^*_1/(4 + 2\delta^*_1)\} - \theta_3)$.

Then,

$$\lim_{n\to\infty} \frac{q_n(\hat{k}_{n,P_n}) - \sigma^2}{L_n(k^*_{n,P_n})} = 1. \tag{4.7}$$

Remark 5. If in (4.6), $\theta = 0$ and $M = 1$, then it can be shown that (4.7) is still valid without condition (4.6). This result can be applied to verify that BIC is not asymptotically efficient in the exponential-decay case; see Example 7 for more details. □

Remark 6. Since (K.5) implies that $k^*_{n,P_n} \to \infty$ as $n \to \infty$, (4.6) is not needed when $\limsup_{n\to\infty} P_n < \infty$. □

The following examples illustrate implications of Theorem 3. Special emphasis is placed on comparing the predictive capabilities of three well-known information criteria, AIC, HQ and BIC, in various situations.

Example 6. Assume that the AR coefficients satisfy (3.10). As mentioned previously, (3.10) is fulfilled by any causal and invertible ARMA(p,q) model with $q > 0$. We shall show in this example that when P_n satisfies (4.3) and $P_n = o(\log n)$, then the corresponding information criterion (including AIC and HQ as special cases) is asymptotically efficient. By the same reasoning as in Example 1, it follows that

$$k^*_{n,P_n} = \frac{\log n + \log(P_n - 1)}{\beta} + O(1), \tag{4.8}$$

and for any $\xi > 0$, (4.5) holds for $\theta = 0$ and some $M > 0$. These results and the restriction, $P_n = o(\log n)$, further imply (4.6). According to Theorem 3, (4.7) follows. Moreover, the claimed result is ensured by observing that (3.13) is still valid if k^*_{n,δ_n} in the denominator is replaced by k^*_{n,P_n}. □

Example 7. This example illustrates that an information criterion cannot be asymptotically efficient in the exponential-decay case when the weight for penalizing the number of regressors in the model is "too strong". To see this, let (3.15) hold and

$$P_n = C_1(\log n)^{C_2}, \tag{4.9}$$

for some $C_1, C_2 > 0$. Under these assumptions, we obtain (4.8) and that for any $\xi > 0$, (4.5) holds true for $\theta = 0$ and $M = 1$. By Remark 5, (4.7) follows. Moreover, since (3.15), (4.8) and (4.9) yield

$$\liminf_{n\to\infty} \frac{L_n(k^*_{n,P_n})}{L_n(k^*_n)} > 1, \tag{4.10}$$

it is concluded that $IC_{P_n}(k)$, with P_n satisfying (4.9), is not asymptotically efficient. One important implication of this example is that BIC is not asymptotically efficient in the algebraic-decay case. □

Example 8. Consider the algebraic-decay case (2.7). Let P_n satisfy (4.3) and

$$P_n = O\left((\log n)^{C_1}\right), \tag{4.11}$$

for some $C_1 > 0$. By an argument similar to that used in Example 3, it follows that

$$k^*_{n,P_n} = \left(\frac{NC_3\beta}{(P_n-1)\sigma^2}\right)^{1/(\beta+1)} + O(1), \tag{4.12}$$

and for any $\xi > 0$, (4.5) holds for $1-\min\{\xi, 1\} < \theta < 1$ and some $M > 0$. In addition, (4.6) is ensured by (4.11) and (4.12). As a result, (4.7) follows from Theorem 3. Moreover, it can be shown that under (2.7),

$$\lim_{n\to\infty} \frac{L_n(k^*_{n,P_n})}{L_n(k^*_n)} = 1 \tag{4.13}$$

if $\lim_{n\to\infty} P_n = 2$, and

$$\limsup_{n\to\infty} \frac{L_n(k^*_{n,P_n})}{L_n(k^*_n)} > 1 \tag{4.14}$$

if $\lim_{n\to\infty} P_n \neq 2$. (4.7), (4.13) and (4.14) imply that AIC is asymptotically efficient in the algebraic-decay case (2.7), whereas HQ, BIC and any information criterion with $\lim_{n\to\infty} P_n \neq 2$ are not. □

Before leaving this section, we note that when the conditions imposed by Theorems 1 and 3 (or Theorems 2 and 3) hold and

$$\lim_{n\to\infty} \frac{\log \delta_n^{-1}}{(1-\delta_n)(P_n-1)} = 1, \tag{4.15}$$

then

$$\lim_{n\to\infty} \frac{E\left(x_{n+1} - \hat{x}_{n+1}(\hat{k}_{n,P_n})\right)^2 - \sigma^2}{E\left(x_{n+1} - \hat{x}_{n+1}(\hat{k}_{n,\delta_n})\right)^2 - \sigma^2} = 1 \tag{4.16}$$

is obtained. As observed, (4.16) leads to an asymptotic equivalence between APE_{δ_n} and IC_{P_n} from a same-realization prediction point of view. For a related result, Wei (1992, Theorem 4.2.2), under (1.1) and certain moment conditions on e_t (which can be verified for the normal distribution), established an algebraic connection between BIC and APE,

$$\log\left(\frac{\text{APE}(k)}{n}\right) = \text{BIC}(k) + o\left(\frac{\log n}{n}\right) \text{ a.s.}, \tag{4.17}$$

where k is a positive integer and *fixed* with n. Therefore, except for the $o(\log n/n)$ term, the logarithm of APE$(k)/n$ is (a.s.) identical to BIC(k). Hannan et al. (1989) also obtained (4.17) in a stationary AR(p_0) model with $p_0 < \infty$ and $k \geq p_0$ (the correctly specified case). However, the equivalence introduced by (4.16) seems to be more relevant in situations where the two criteria's predictive capabilities after order selection are compared.

5. Optimal prediction for possibly degenerate AR(∞) processes

This section deals with optimal prediction problems in situations where the underlying AR(∞) process can degenerate to an AR process of finite order. We first adopt (K.5$'$) to replace the truly infinite-order assumption (K.5).

(K.5$'$): The AR coefficients satisfy either

(i) $a_{p_0} \neq 0$ for some unknown $1 \leq p_0 < \infty$ and $a_l = 0$ for all $l \geq p_0 + 1$; or

(ii) (3.10).

(Note that (K.1) guarantees that (3.10) is equivalent to $C_1 e^{-\beta k} \leq \sum_{i \geq k} a_i^2 \leq C_2 e^{-\beta k}$ for some $0 < C_1 \leq C_2 < \infty$.) From a practical point of view, (K.5$'$) is reasonably flexible because it contains any causal and invertible ARMA(p,q) model, with $p + q \geq 1$, as a special case. Before tackling order selection problems under (K.5$'$), a preliminary result is needed, which shows that APE$_{\delta_n}$ and IC_{P_n}, with δ_n and P_n satisfying certain conditions, are asymptotically efficient in finite-order cases.

Theorem 4. *Assume that (K.1)–(K.4) and (i) of (K.5$'$) hold. Then, (2.3) holds for $\hat{k}_n = \hat{k}_{n,\delta_n}$ and $\hat{k}_{n,P_n}$, where δ_n satisfies $\delta_n^{-1} \to \infty$ and (3.6), and P_n satisfies $P_n \to \infty$ and $P_n/n \to 0$.*

Remark 7. Since Theorem 4 adopts $\{\text{AR}(1), \cdots, \text{AR}(K_n)\}$ as the set of candidate models, where $K_n \to \infty$ at a certain rate, the true model AR(p_0) is included asymptotically. Zheng and Loh (1997) also took this approach and showed that $\hat{k}_0 = \arg\min_{1 \leq k \leq K_n} IC_{P_n}(k)$ is a consistent estimator of p_0 under the assumptions that $K_n^2/n \to 0$, $P_n/K_n \to \infty$ and $P_n/n \to 0$. While their conditions on e_t were weaker than those in Theorem 4, they did not evaluate the (same-realization) prediction efficiencies of the proposed information criteria. Moreover, the limitation of $P_n/K_n \to \infty$ is cumbersome when (ii) of (K.5$'$) is simultaneously taken into account. To achieve optimal prediction in this latter situation, one needs to justify the validity of IC_{P_n}, with P_n tending to infinity more slowly than K_n; see the discussion below for details. It seems difficult to attain this goal based on Zheng and Loh's result due to the limitation mentioned above. □

When (ii) of (K.5′) holds, Example 6 points out that IC_{P_n}, with $P_n = o(\log n)$, possesses asymptotic efficiency. (Note that in this case, the orders of the optimal prediction models tend to infinity at a $\log n$ rate (see, e.g., Goldenshluger and Zeevi, 2001, and Ing and Wei, 2005. (K.4), requiring K_n to grow faster than $\log n$, guarantees that these models are ultimately included among the candidate family.) On the other hand, if (i) of (K.5′) is true, then Theorem 4 shows that IC_{P_n}, with $P_n \to \infty$ and $P_n/n \to 0$, is asymptotically efficient under (K.1)-(K.4). These results suggest that IC_{P_n}, with $P_n \to \infty$, $P_n = o(\log n)$ and K_n satisfying (K.4), can simultaneously achieve (2.3) over the two types of AR processes defined in (i) and (ii) of (K.5′). According to Example 1 and Theorem 4, APE_{δ_n}, with $\delta_n^{-1} \to \infty$, $\log \delta_n^{-1} = o(\log n)$ and K_n satisfying (K.4), also has this property. These discussions are now summarized in the following theorem.

Theorem 5. *Assume that (K.1)-(K.4) and (K.5′) hold. Then, (2.3) holds for* $\hat{k}_n = \hat{k}_{n,\delta_n}$ *and* $\hat{k}_{n,P_n}$*, where* δ_n *satisfies* $\delta_n^{-1} \to \infty$ *and* $\log \delta_n^{-1} = o(\log n)$*, and* P_n *satisfies* $P_n \to \infty$ *and* $P_n = o(\log n)$*.*

While Theorem 5 seems satisfactory for practical purposes, the question of how (2.3) is achieved in a more general case that allows the AR coefficients to decay algebraically still attracts much theoretical interest. As can be seen in Examples 3 and 8, the criteria given by Theorem 5 fail to preserve asymptotic efficiency when (2.7) is added into (K.5′). Therefore, we propose using an alternative criterion that chooses order $\hat{k}_n^{(\iota)}$:

$$\hat{k}_n^{(\iota)} = \hat{k}_{n,2} I_{\{\hat{k}_{n,P_n} \neq \hat{k}_{n^\iota, P_{n^\iota}}\}} + \hat{k}_{n,P_n} I_{\{\hat{k}_{n,P_n} = \hat{k}_{n^\iota, P_{n^\iota}}\}}, \tag{5.1}$$

where $0 < \iota < 1$, $P_n \to \infty$, $\hat{k}_{n^\iota, P_{n^\iota}} = \arg\min_{1 \le k \le K_{n^\iota}} IC_{P_{n^\iota}}(k)$ and

$$IC_{P_{n^\iota}}(k) = \log \hat{\sigma}^2_{n^\iota}(k) + \frac{P_{n^\iota} k}{n^\iota},$$

with $\hat{\sigma}^2_{n^\iota}(k) = (1/N_\iota) \sum_{j=K_{n^\iota}}^{n^\iota - 1} (x_{j+1} + \hat{\mathbf{a}}'_{n^\iota}(k) \mathbf{x}_j(k))^2$, $N_\iota = n^\iota - K_{n^\iota}$,

$$\hat{\mathbf{a}}_{n^\iota}(k) = -\hat{R}^{-1}_{\iota, n^\iota}(k)(1/N_\iota) \sum_{j=K_{n^\iota}}^{n^\iota - 1} \mathbf{x}_j(k) x_{j+1},$$

and $\hat{R}_{\iota, n^\iota}(k) = (1/N_\iota) \sum_{j=K_{n^\iota}}^{n^\iota - 1} \mathbf{x}_j(k) \mathbf{x}'_j(k)$ (note that without loss of generality, n^ι and K_{n^ι} are assumed to be positive integers). As observed, (5.1) is a hybrid selection procedure that combines together AIC and a BIC-like criterion. If the true order is finite, then it is expected that the orders selected by the BIC-like criterion at stages n^ι and n will ultimately be the same due to consistency. On the other hand, when the true order is infinite, an interesting result is derived for which it is nearly impossible for the BIC-like criterion to choose the same order at these different stages; see Appendix C for more details. Therefore, it is reasonable to adopt $\hat{k}_{n,2}$ (the order selected by AIC) if IC_{P_n} and $IC_{P_{n^\iota}}$ determine different orders, and $\hat{k}_{n,P_n}$ (the order selected by the BIC-like criterion) otherwise. Theorem 6 justifies the validity of $\hat{k}_n^{(\iota)}$.

Theorem 6. *Let (K.1)–(K.4) and (K.6) hold and ι and P_n in (5.1) satisfy $0 < \iota < 1$, $P_n \to \infty$, $P_n = O(n^{\iota_1})$, with $0 < \iota_1 < \delta_1^*/(2+\delta_1^*)$, and $P_n/P_{n^\iota}^{\nu} = O(1)$ for some $\nu > 0$. Further, assume that the AR coefficients meet one of the following conditions:*

(i) (i) of (K.5′); or

(ii) For every exponent $\xi > 0$, there is a nonnegative exponent $0 \le \theta = \theta(\xi) < 1$ and a positive number $M = M(\xi) > 0$ such that

$$\liminf_{n\to\infty} \min_{k\in A_{P_n,\theta,M}} (k^*_{n,P_n})^{\xi} \frac{L_{n,P_n}(k) - L_{n,P_n}(k^*_{n,P_n})}{L_{n,P_n}(k^*_{n,P_n})} > 0, \tag{5.2}$$

with $A_{P_n,\theta,M}$ (defined in Theorem 3) satisfying

$$A^C_{P_{n^\iota},\theta,M} \cap A^C_{P_n,\theta,M} = \emptyset \tag{5.3}$$

for all sufficiently large n. Here, $\emptyset$ denotes the empty set,

$$A^C_{P_n,\theta,M} = \{k : 1 \le k \le K_n, k \notin A_{P_n,\theta,M}\}$$

and

$$A^C_{P_{n^\iota},\theta,M} = \{k : 1 \le k \le K_{n^\iota}, k \notin A_{P_{n^\iota},\theta,M}\},$$

with $A_{P_{n^\iota},\theta,M} = \{k : 1 \le k \le K_{n^\iota}, |k - k^*_{n^\iota,P_{n^\iota}}| \ge (k^*_{n^\iota,P_{n^\iota}})^{\theta}\}$. (Note that (5.3) implies that $a_l \ne 0$ for infinitely many l.)

Then, (2.3) holds for $\hat{k}_n^{(\iota)}$.

As an application of Theorem 6, it is shown in Example 9 below that $\hat{k}_n^{(\iota)}$, $0 < \iota < 1$, is asymptotically efficient when the true model is: (i) an AR process of finite order, (ii) an AR(∞) process with coefficients satisfying (3.10) (the exponential-decay case); or (iii) an AR(∞) process with coefficients satisfying (2.7) (the algebraic-decay case). To simplify the discussion, let

$$P_n = C_1 (\log n)^{C_2}, \tag{5.4}$$

for some $C_1, C_2 > 0$. Note that (5.4) satisfies all requirements for P_n imposed by Theorem 6.

Example 9. Assume that either (K.5$'$) or (2.7) holds. To show that $\hat{k}_n^{(\iota)}$, $0 < \iota < 1$, is asymptotically efficient in this situation, in view of Theorem 6, it suffices to show that (5.2) and (5.3) are satisfied by both (ii) of (K.5$'$) (or, equivalently, (3.10)) and (2.7). First, assume that (3.10) is true. Then, by an argument similar to that used in Example 6, we obtain (4.8) and that for any $\xi > 0$, (5.2) holds for $1 - \min\{\xi, 1\} < \theta < 1$ and some $M > 0$. In addition, it is easy to see that (5.3) follows from (4.8), (5.4) and the facts that $0 < \iota < 1$ and $0 \le \theta < 1$.

Next, let (2.7) hold. Reasoning as for Example 8, (4.12) is obtained and for any $\xi > 0$, (5.2) holds for $1 - \min\{\xi, 1\} < \theta < 1$ and some $M > 0$. Moreover, (5.3) follows from (4.12), (5.4), $0 < \iota < 1$ and $0 \le \theta < 1$. Consequently, the desired result is obtained. □

Remark 8. To suggest a suitable combination of ι and P_n in finite samples, one may rely on an extensive simulation study. This is the subject of ongoing research. It is worth noting that based on APE$_{\delta_n}$, we can also construct a two-stage criterion to achieve (2.3) universally over the three types of AR processes mentioned after Theorem 6. However, this criterion seems relatively less attractive compared to $\hat{k}_n^{(\iota)}$, since it gets involved in the trouble of determining twice the number of tuning parameters, namely, ι, δ_n, C_1 and r, where C_1 and r are defined in (3.19). □

6. Concluding remarks.

Recently, APE$_{\delta_n}$ has become very popular among researchers from several disciplines, particularly those required to do a lot of forecasting. While it is of fundamental importance to realize the impact of APE$_{\delta_n}$ on predictions after model selection, discussions directly related to this issue still seem to be lacking. Theorems 1 and 2 of Section 3 are devoted to filling this gap. Under model (1.1), they provide an asymptotic expression for the MSPE of the (least squares) predictor with the order determined by APE$_{\delta_n}$, where δ_n can vary freely over (0,1), and is allowed to tend to 0 or 1 at a suitable rate. In light of this expression, we are able to assess APE$_{\delta_n}$'s predictive performance after model (order) selection in various practical situations. In particular, a series of examples in Section 3 shows that when δ_n is suitably chosen, APE_{δ_n} can achieve asymptotic efficiency in both exponential- and algebraic-decay cases.

An asymptotic equivalence between APE_{δ_n} and IC_{P_n} is established in Section 4 from a prediction point of view. Since this equivalence can be checked simply through δ_n and P_n (see (4.15)), it offers a new and global perspective for comparing information- and prediction-based model selection criteria in misspecified AR processes. Section 5 provides the first asymptotic efficiency result for the case when model (1.1) is allowed to degenerate to a finite autoregression. Two special features are worth mentioning: (1) We show that some consistent criteria, such as HQ and APE_{δ_n}, with $\delta_n \to 0$ and $\log \delta_n^{-1} = o(\log n)$, can simultaneously attain asymptotic efficiency over finite-order AR models and AR(∞) models with exponentially decaying coefficients, which constitute an important class of AR(∞) models. (2) A new procedure (which is a hybrid between AIC and a BIC-like criterion) is constucted to achieve asymptotic efficiency in more general AR models, which include finite-order AR models and AR(∞) models with exponentially or algebraically decaying coefficients as special cases. The success of this new procedure relies mainly on a two-stage design that allows AIC and a BIC-like criterion to cover each other's weaknesses.

To verify the main results of the present article, (A.1) and (A.2) of Appendix A, which provide qth moment bounds for the inverse of the sample covariance matrix with an increasing dimension, are required to hold for sufficiently large q. By assuming (K.3) (among other conditions), Lemma A.1 of Appendix A (see, also, Ing and Wei, 2003, Theorem 2) guarantees that (A.1) and (A.2) hold for any $q > 0$, and hence is used to meet this requirement. As a result, (K.3) appears in all theorems of this paper because of Lemma A.1. Although (K.3) is rather stronger than is necessary, it is convenient. Note that it is possible to slightly relax (K.3) at the price of greatly reducing the number of candidate models; see Ing and Wei (2005, Section 6) for a related discussion. However, since the benefits are rather limited, the details are not pursued here in order to simplify the discussion. To substantially loosen (K.3), it is necessary to verify (A.1) and (A.2) under much milder moment conditions. Further efforts are still needed to achieve this goal.

As a final remark, we note that the popular fractional integrated ARMA process is excluded by (K.1). Extensions of these results to situations where some long-memory time series models are included are currently being investigated by the author.

Appendix A: Proofs of Theorems 1 and 2

This section begins with Lemma 2 of Wei (1987), which is frequently used later. In the rest of this paper, C is used to denote a generic positive constant independent of sample size n and of any index with an upper (or lower) limit dependent on n.

Lemma A.0. *Let $\{\varepsilon_t, \mathcal{F}_t\}$ be a sequence of martingale differences such that for some $\alpha \geq 2$,*

$$\sup_t E\{|\varepsilon_t|^\alpha | \mathcal{F}_{t-1}\} \leq C \;\; a.s.$$

Let μ_t be $\mathcal{F}_{t-1}$-measurable random variables, $S_{n_1} = \sum_{t=1}^{n_1} \mu_t \varepsilon_t$ and

$$S^*_{n_1} = \max_{1 \leq t \leq n_1} |S_t|.$$

Then,

$$E(S^*_{n_1})^\alpha \leq K E(\sum_{t=1}^{n_1} \mu_t^2)^{\alpha/2},$$

where K depends only on α and C.

Lemmas 1-4 below, quoted from Proposition 1 and Lemmas 1-3 of Ing and Wei (2005), respectively, also play important roles. To introduce these results, we need (K.1′), a condition slightly weaker than (K.1).

(K.1′) Let $\{ x_t \}$ be a linear process satisfying (1.1) with $A(z) = 1 + a_1 z + a_2 z^2 + \cdots \neq 0$ for $| z | \leq 1$. Furthermore, let coefficients $\{ a_i \}$ obey $\sum_{i=1}^{\infty} |a_i| < \infty$.

Lemma A.1. *Assume that* (K.1′), (K.2), (K.3) *and $K_n = O(n^{(1/2)-r})$ hold for some $r > 0$. Then, for any $q > 0$,*

$$\max_{1 \leq k \leq K_n} E \left\| \hat{R}_n^{-1}(k) \right\|^q \leq C \tag{A.1}$$

and

$$\max_{1 \leq k \leq K_n} \frac{E \left\| \hat{R}_n^{-1}(k) - R^{-1}(k) \right\|^q}{\left(\dfrac{k^2}{N} \right)^{q/2}} \leq C \tag{A.2}$$

hold for all sufficiently large n, *where* $\hat{R}_n(k)$ *is defined in (1.4),* $R(k) = E(\mathbf{x}_n(k)\mathbf{x}'_n(k))$ *and* $\mathbf{x}_n(k)$ *is defined after (1.2).*

Lemma A.2. *Assume that* (K.1$'$) *holds and* $\sup_{-\infty<t<\infty} E\left(|e_t|^{2q}\right) < \infty$ *for some* $q \geq 2$. *Let* $\{m_{i,n}\}$, *i=0, 1, 2, be sequences of positive integers satisfying* $m_{2,n} \geq m_{1,n} \geq m_{0,n}$ *for all* $n \geq 1$. *Then, for all* $1 \leq k \leq m_{0,n}$,

$$E\left\| \frac{1}{\sqrt{m_n}} \sum_{j=m_{1,n}}^{m_{2,n}} \mathbf{x}_j(k)(e_{j+1,k} - e_{j+1}) \right\|^q \leq Ck^{q/2}\|\mathbf{a} - \mathbf{a}(k)\|_R^q, \tag{A.3}$$

where $m_n = m_{2,n} - m_{1,n} + 1$, $e_{j+1,k}$ *is defined after (2.1),* $\|\mathbf{a} - \mathbf{a}(k)\|_R^2$ *is defined in Proposition 1 and for a k-dimensional vector* $\mathbf{v} = (v_1, \cdots, v_k)'$, $\|\mathbf{v}\|^2 = \sum_{i=1}^k v_i^2$.

Lemma A.3. *Assume that* (K.1$'$) *holds and* $\sup_{-\infty<t<\infty} E\{|e_t|^q\} < \infty$ *for some* $q \geq 2$. *Let* $\{m_{i,n}\}$, *i= 0, 1, 2, and* $\{m_n\}$ *be defined as in Lemma A.2. Then,*

$$\max_{1\leq k\leq m_{0,n}} (k^{-q/2})E\left\| \frac{1}{\sqrt{m_n}} \sum_{j=m_{1,n}}^{m_{2,n}} \mathbf{x}_j(k)e_{j+1} \right\|^q \leq C. \tag{A.4}$$

Lemma A.4. *Assume that* (K.1), $sup_{-\infty<t<\infty}E\,|\,e_t\,|^{2q} < \infty$, *for some* $q \geq 2$, *and* $K_n = O(n^{1/2})$ *hold. Then, for all* $1 \leq k \leq K_n$,

$$E\left| \left\| \frac{1}{\sqrt{N}} \sum_{j=K_n}^{n-1} \mathbf{x}_j(k)e_{j+1} \right\|_{R^{-1}(k)}^2 - k\sigma^2 \right|^q \leq Ck^{q/2}, \tag{A.5}$$

where for the $k\times k$ *symmetric matrix* A *and k-dimensional vector* $\mathbf{y}$, $\|\,\mathbf{y}\|_A^2 = \mathbf{y}'A\mathbf{y}$.

We also need a modification of Lemma 6 of Ing and Wei (2005).

Lemma A.5. *Assume* (K.1$'$) *and* $sup_{-\infty<t<\infty}E\,|\,e_t\,|^{2q} < \infty$ *for some* $q \geq 2$. *Let* $\{m_{i,n}\}$, *i= 0, 1, 2, and* $\{m_n\}$ *be defined as in Lemma A.2. Then, for all* $1 \leq k, j \leq m_{0,n}$,

$$E\,\left|\, S^2_{m_{1,n},m_{2,n}}(k) - \sigma_k^2 - \left(S^2_{m_{1,n},m_{2,n}}(j) - \sigma_j^2\right)\right|^q \leq Cm_n^{-q/2}\,\|\,\mathbf{a}(j) - \mathbf{a}(k)\,\|_R^q\,, \tag{A.6}$$

where $S^2_{m_{1,n},m_{2,n}}(k) = (1/m_n)\sum_{t=m_{1,n}}^{m_{2,n}} e^2_{t+1,k}$, $\mathbf{a}(j)$ *and* $\mathbf{a}(k)$ *in (A.6) are viewed as infinite-dimensional vectors with undefined entries set to zero, and* $\sigma_k^2 = E(\,e^2_{1,k}\,)$. *Also note that* $\|\,\mathbf{a}(j) - \mathbf{a}(k)\,\|_R^2 = |\|\mathbf{a} - \mathbf{a}(j)\|_R^2 - \|\mathbf{a} - \mathbf{a}(k)\|_R^2|$.

Since the proof of (A.6) is similar to that of Ing and Wei (2005, Lemma 6), the details are omitted. Now, moment bounds for $N_i(k), i = 1, \cdots, 12$ are established in Lemmas A.6-A.9.

Lemma A.6. *Let the assumptions of Proposition 1 hold and $0 < \delta_n < 1$ satisfy (3.6). Then, for $q > 0$, all $1 \le k \le K_n$ and all sufficiently large n,*

$$
\begin{aligned}
& E(|N_1(k)|^q) \\
\le \; & \frac{C}{(L_n^{(\delta_n)}(k))^q(1-\delta_n)^q}\left\{\frac{k^{2q}}{(n\delta_n)^{q/2}n^q} + \frac{(\log \delta_n^{-1})^q k^q \|\mathbf{a}-\mathbf{a}(k)\|_R^{2q}}{n^q}\right\}, \quad \text{(A.7)}
\end{aligned}
$$

and

$$
\begin{aligned}
& E(|N_2(k)|^q) \\
\le \; & \frac{C}{(L_n^{(\delta_n)}(k))^q(1-\delta_n)^q}\left\{\frac{k_{n,\delta_n}^{*2q}}{(n\delta_n)^{q/2}n^q} + \frac{(\log \delta_n^{-1})^q k_{n,\delta_n}^{*q} \|\mathbf{a}-\mathbf{a}(k_{n,\delta_n}^{*})\|_R^{2q}}{n^q}\right\}. \quad \text{(A.8)}
\end{aligned}
$$

PROOF. We only prove (A.7) because the proof of (A.8) is similar. First note that

$$
\begin{aligned}
& L_n^{(\delta_n)}(k) n (1-\delta_n)|N_1(k)| \le \left|\sum_{i=n\delta_n}^{n-1} \frac{\mathbf{x}_i'(k)(\hat{R}_{i+1}^{-1}(k) - R^{-1}(k))\mathbf{x}_i(k)}{i+1-K_n} e_{i+1,k}^2\right| \\
+ \; & \left|\sum_{i=n\delta_n}^{n-1} \frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)}{i+1-K_n}(e_{i+1,k}^2 - e_{i+1}^2)\right| + \left|\sum_{i=n\delta_n}^{n-1} \frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)}{i+1-K_n}(e_{i+1}^2 - \sigma^2)\right| \\
+ \; & \left|\sum_{i=n\delta_n}^{n-1} \frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k) - k}{i+1-K_n}\sigma^2\right| + k\sigma^2\left|\left(\sum_{i=n\delta_n}^{n-1} \frac{1}{i+1-K_n}\right) - \log \delta_n^{-1}\right| \\
\equiv \; & I(k) + II(k) + III(k) + IV(k) + V(k). \quad \text{(A.9)}
\end{aligned}
$$

By (3.6) and Lemma A.1, we have for any $q > 0$, all $n\delta_n \le i \le n-1$, all $1 \le k \le K_n$ and all sufficiently large n,

$$
E\|\hat{R}_{i+1}^{-1}(k) - R^{-1}(k)\|^q \le C\frac{k^q}{(i+1-K_n)^{q/2}}. \quad \text{(A.10)}
$$

In addition, Lemma A.0 and Jensen's inequality yield that for any $r > 0$, all $n\delta_n \le i \le n-1$ and all $1 \le k \le K_n$,

$$
E(\|\mathbf{x}_i(k)\|^r) \le Ck^{r/2}, \quad \text{(A.11)}
$$

and

$$
E|e_{i+1,k}|^r) \le C. \quad \text{(A.12)}
$$

According to (A.10)-(A.12), Minkowski's inequality and Hölder's inequality, we have for $q \geq 1$, all $1 \leq k \leq K_n$ and all sufficiently large n,

$$
\begin{aligned}
E(I(k))^q &\leq \left(\sum_{i=n\delta_n}^{n-1} \left\| \frac{\mathbf{x}_i'(k)(\hat{R}_{i+1}^{-1}(k) - R^{-1}(k))\mathbf{x}_i(k)}{i+1-K_n} e_{i+1,k}^2 \right\|_q \right)^q \\
&\leq \left[\sum_{i=n\delta_n}^{n-1} \frac{\left\{ E(\|\mathbf{x}_i(k)\|^{6q}) E(\|\hat{R}_{i+1}^{-1}(k) - R^{-1}(k)\|^{3q}) E(|e_{i+1,k}^{6q}|) \right\}^{1/3q}}{i+1-K_n} \right]^q \\
&\leq Ck^{2q}(\frac{1}{n\delta_n})^{q/2}, \qquad \text{(A.13)}
\end{aligned}
$$

where for random variable z and positive number s, $\|z\|_s = E(|z|^r)^{1/s}$.

To deal with $II(k)$, notice that the first moment bound theorem of Findley and Wei (1993) and Jensen's inequality yield for any $r > 0$, all $K_n \leq i \leq n-1$ and all $1 \leq k \leq K_n$,

$$
E(|\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k) - k|^r) \leq Ck^{r/2}. \qquad \text{(A.14)}
$$

Reasoning as for (A.12), we have for any $r > 0$, all $K_n \leq i \leq n-1$ and all $1 \leq k \leq K_n$,

$$
E(|e_{i+1,k} - e_{i+1}|^r) \leq C\|\mathbf{a} - \mathbf{a}(k)\|_R^r. \qquad \text{(A.15)}
$$

Also observe that

$$
\left\{ \frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)}{i+1-K_n}(e_{i+1,k} - e_{i+1})e_{i+1}, \mathcal{M}_{i+1} \right\}
$$

is a sequence of martingale differences, where $\mathcal{M}_{i+1}$ is the σ-algebra generated by $\{e_{i+1}, e_i, e_{i-1}, \cdots\}$, and

$$
\begin{aligned}
& E\left| \sum_{i=n\delta_n}^{n-1} \frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)}{i+1-K_n}(e_{i+1,k} - e_{i+1})e_{i+1} \right|^q \\
\leq\ & E \max_{n\delta_n \leq m \leq n-1} \left| \sum_{i=n\delta_n}^{m} \frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)}{i+1-K_n}(e_{i+1,k} - e_{i+1})e_{i+1} \right|^q.
\end{aligned}
$$

These facts, Lemma A.0, (A.14), (A.15) and an argument similar to that used for obtaining (A.13) together imply that for $q \geq 2$ and all $1 \leq k \leq K_n$,

$$
E(II(k))^q
$$

$$
\begin{aligned}
&\le C\left\{E\left|\sum_{i=n\delta_n}^{n-1}\frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)}{i+1-K_n}(e_{i+1,k}-e_{i+1})^2\right|^q\right.\\
&+\left.E\left|\sum_{i=n\delta_n}^{n-1}\frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)}{i+1-K_n}(e_{i+1,k}-e_{i+1})e_{i+1}\right|^q\right\}\\
&\le C\left[\left(\sum_{i=n\delta_n}^{n-1}\left\|\frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)}{i+1-K_n}(e_{i+1,k}-e_{i+1})^2\right\|_q\right)^q\right.\\
&+\left.E\left\{\sum_{i=n\delta_n}^{n-1}\left(\frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)}{i+1-K_n}\right)^2(e_{i+1,k}-e_{i+1})^2\right\}^{q/2}\right]\\
&\le C\left\{(\log\delta_n^{-1})^q k^q\|\mathbf{a}-\mathbf{a}(k)\|_R^{2q}+\left(\frac{k}{n\delta_n}\right)^{q/2}k^{q/2}\|\mathbf{a}-\mathbf{a}(k)\|_R^q\right\}. \qquad \text{(A.16)}
\end{aligned}
$$

Similarly, by Lemma A.0, (A.14) and the Minkowski inequality,

$$
E(III(k))^q \le CE\left\{\sum_{i=n\delta_n}^{n-1}\left(\frac{\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)}{i+1-K_n}\right)^2\right\}^{q/2} \le C(\frac{k^2}{n\delta_n})^{q/2} \qquad \text{(A.17)}
$$

holds for $q \ge 2$ and all $1 \le k \le K_n$.

To deal with IV(k), we have by some algebraic manipulations that

$$
\begin{aligned}
IV(k) &= \sigma^2\left|\sum_{i=n\delta_n}^{n-1}\frac{T_i(k)-T_{i-1}(k)}{i+1-K_n}\right|\\
&= \sigma^2\left|\frac{T_{n-1}(k)}{N}-\frac{T_{n\delta_n-1}(k)}{n\delta_n-K_n}+\sum_{i=n\delta_n}^{n-1}\frac{T_{i-1}(k)}{(i-K_n)(i+1-K_n)}\right|,
\end{aligned}
$$

where $T_i(k)=\sum_{j=K_n}^{i}\mathbf{x}_i'(k)R^{-1}(k)\mathbf{x}_i(k)-k$. By an argument similar to that given in the proof of Lemma 3 of Ing and Wei (2005) and Jensen's inequality, one has for any $q>0$, all $n\delta_n-1\le i\le n-1$, and all $1\le k\le K_n$,

$$
E\left|\frac{T_i(k)}{i+1-K_n}\right|^q \le C\frac{k^{3q/2}}{(i+1-K_n)^{q/2}}.
$$

This and the Minkowski inequality yield that for $q \ge 1$ and all $1 \le k \le K_n$,

$$
E(IV(k))^q \le C\frac{k^{3q/2}}{(n\delta_n)^{q/2}}. \qquad \text{(A.18)}
$$

Moreover, it is also not difficult to see that for all $1 \le k \le K_n$,

$$
V(k) \le C(\frac{1-\delta_n}{\delta_n})k(\frac{K_n}{n}). \qquad \text{(A.19)}
$$

Consequently, (A.7) follows from (A.9), (A.13), (A.16)-(A.19), Jensen's inequality, and the fact that for any $r > 0$,

$$\lim_{k\to\infty} k^r \|\mathbf{a} - \mathbf{a}(k)\|_R^{2r} = 0, \tag{A.20}$$

which is ensured by (K.1). □

Lemma A.7. *Under the assumptions of Lemma A.6, we have for $q > 0$, all $1 \le k \le K_n$ and all sufficiently large n,*

$$E(|N_3(k)|^q) \le \frac{C}{(1-\delta_n)^q (L_n^{(\delta_n)}(k))^q} \left(\frac{k^{2q}}{n^{3q/2}\delta_n^{q/2}} + \frac{k^{q/2}}{n^q} \right), \tag{A.21}$$

$$E(|N_4(k)|^q) \le \frac{C}{(1-\delta_n)^q (L_n^{(\delta_n)}(k))^q} \left(\frac{k^{*2q}_{n,\delta_n}}{n^{3q/2}\delta_n^{q/2}} + \frac{k^{*q/2}_{n,\delta_n}}{n^q} \right), \tag{A.22}$$

$$E(|N_5(k)|^q) \le \frac{C}{(1-\delta_n)^q (L_n^{(\delta_n)}(k))^q} \left(\frac{k^{2q}}{n^{3q/2}} + \frac{k^{q/2}}{n^q} \right), \tag{A.23}$$

and

$$E(|N_6(k)|^q) \le \frac{C}{(1-\delta_n)^q (L_n^{(\delta_n)}(k))^q} \left(\frac{k^{*2q}_{n,\delta_n}}{n^{3q/2}} + \frac{k^{*q/2}_{n,\delta_n}}{n^q} \right). \tag{A.24}$$

PROOF. We only prove (A.21) because the proofs of (A.22), (A.23) and (A.24) are similar. Some algebraic manipulations give

$$\begin{aligned}
& |Q_{n\delta_n}(k) - k\sigma^2| \\
\le\ & \frac{1}{n\delta_n - K_n} \left| \sum_{j=K_n}^{n\delta_n - 1} \mathbf{x}'_j(k) e_{j+1,k} \left(\hat{R}^{-1}_{n\delta_n}(k) - R^{-1}(k) \right) \sum_{j=K_n}^{n\delta_n - 1} \mathbf{x}_j(k) e_{j+1,k} \right| \\
+\ & \frac{1}{n\delta_n - K_n} \left| \sum_{j=K_n}^{n\delta_n - 1} \mathbf{x}'_j(k)(e_{j+1,k} - e_{j+1}) R^{-1}(k) \sum_{j=K_n}^{n\delta_n - 1} \mathbf{x}_j(k) e_{j+1,k} \right| \\
+\ & \frac{1}{n\delta_n - K_n} \left| \sum_{j=K_n}^{n\delta_n - 1} \mathbf{x}'_j(k) e_{j+1} R^{-1}(k) \sum_{j=K_n}^{n\delta_n - 1} \mathbf{x}_j(k)(e_{j+1,k} - e_{j+1}) \right| \\
+\ & \left| \frac{1}{n\delta_n - K_n} \left(\sum_{j=K_n}^{n\delta_n - 1} \mathbf{x}'_j(k) e_{j+1} R^{-1}(k) \sum_{j=K_n}^{n\delta_n - 1} \mathbf{x}_j(k) e_{j+1} \right) - k\sigma^2 \right|.
\end{aligned}$$

This, Lemmas A.1-A.4, (3.6), (A.20) and Jensen's inequality together imply that for $q > 0$, all $1 \le k \le K_n$ and all sufficiently large n,

$$E|Q_n(k) - k\sigma^2|^q \le C\left(\frac{k^{2q}}{(n\delta_n)^{q/2}} + k^{q/2}\right),$$

and hence (A.21) follows. □

Lemma A.8. *Under the assumptions of Lemma A.6, we have for $q > 0$, all $1 \le k \le K_n$ and all sufficiently large n,*

$$E(|N_7(k)|^q) \le \frac{Ck^{3q}}{(1-\delta_n)^q (L_n^{(\delta_n)}(k))^q n^q (n\delta_n)^q}, \tag{A.25}$$

$$E(|N_8(k)|^q) \le \frac{Ck_{n,\delta_n}^{*3q}}{(1-\delta_n)^q (L_n^{(\delta_n)}(k))^q n^q (n\delta_n)^q}, \tag{A.26}$$

$$E(|N_9(k)|^q) \le \frac{Ck^{2q}}{(1-\delta_n)^q (L_n^{(\delta_n)}(k))^q n^q (n\delta_n)^{q/2}}, \tag{A.27}$$

and

$$E(|N_{10}(k)|^q) \le \frac{Ck_{n,\delta_n}^{*2q}}{(1-\delta_n)^q (L_n^{(\delta_n)}(k))^q n^q (n\delta_n)^{q/2}}. \tag{A.28}$$

PROOF. By an analogy with (A.13), one has for $q \ge 1$,

$$\begin{aligned} &E|n(1-\delta_n)L_n^{(\delta_n)}(k)N_7(k)|^q \\ \le &\left(\sum_{i=n\delta_n}^{n-1} \frac{\left\{E(\|\mathbf{x}_i(k)\|^{6q})E(\|\hat{R}_{i+1}^{-1}(k)\|^{3q})E(|\hat{e}_{i,k}|^{6q})\right\}^{1/3q}}{i-K_n+1}\right)^q. \end{aligned} \tag{A.29}$$

Since for $r > 0$,

$$E(|\hat{e}_{i,k}|^r) \le \left\{E(\|\mathbf{x}_i(k)\|^{2r})E(\|\hat{\mathbf{a}}_i(k) - \mathbf{a}(k)\|^{2r})\right\}^{1/2},$$

and

$$\|\hat{\mathbf{a}}_i(k) - \mathbf{a}(k)\| \le \|\hat{R}_i^{-1}(k)\|\|(i-K_n)^{-1}\sum_{j=K_n}^{i-1}\mathbf{x}_j(k)e_{j+1,k}\|,$$

by (A.11), (3.6) and Lemmas A.1-A.3, one has for all $n\delta_n \le i \le n-1$, all $1 \le k \le K_n$ and all sufficiently large n,

$$E(|\hat{e}_{i,k}|^r) \le C\frac{k^r}{i^{r/2}}. \tag{A.30}$$

Consequently, (A.25) follows from (A.29), (A.30), (A.11), (A.1), and Jensen's inequality. Since the proof of (A.26) is similar to that of (A.25), the details are omitted. In addition, by (A.12) and an argument similar to that used for showing (A.25), (A.27) and (A.28) follow. □

Lemma A.9. *Let the assumptions of Lemma A.5 hold. Then, for some $q \ge 2$, $0 < \delta_n < 1$ and all $1 \le k \le K_n$, with $K_n \le n\delta_n$,*

$$E(|N_{11}(k)|^q) \le \frac{C\|\mathbf{a}(k) - \mathbf{a}(k^*_{n,\delta_n})\|_R^q}{(1-\delta_n)^{q/2}(L_n^{(\delta_n)}(k))^q n^{q/2}}, \tag{A.31}$$

and

$$E(|N_{12}(k)|^q) \le \frac{C\|\mathbf{a}(k) - \mathbf{a}(k^*_{n,\delta_n})\|_R^q}{(1-\delta_n)^{q/2}(L_n^{(\delta_n)}(k))^q n^{q/2}}. \tag{A.32}$$

PROOF. First note that

$$\begin{aligned} E|L_n^{(\delta_n)}(k)N_{11}(k)|^q &\le CE\left|\frac{1}{n(1-\delta_n)}\sum_{i=n\delta_n}^{n-1}\left\{e^2_{i+1,k} - e^2_{i+1,k^*_{n,\delta_n}} - \sigma^2_k + \sigma^2_{k^*_{n,\delta_n}}\right\}\right|^q \\ &+ CE\left|\frac{1}{n(1-\delta_n)}\sum_{i=n\delta_n}^{n-1}(e_{i+1,k} - e_{i+1,k^*_{n,\delta_n}})e_{i+1}\right|^q \\ &\equiv (I) + (II) \end{aligned} \tag{A.33}$$

According to (A.6), one has for all $1 \le k \le K_n$,

$$(I) \le \frac{C\|\mathbf{a}(k) - \mathbf{a}(k^*_{n,\delta_n})\|_R^q}{(1-\delta_n)^{q/2}n^{q/2}}. \tag{A.34}$$

Lemma A.0 and the convexity of $x^{q/2}, x > 0$, yield for all $1 \le k \le K_n$,

$$\begin{aligned} (II) &\le \frac{C}{n^q(1-\delta_n)^q}E\left[\left\{\sum_{i=n\delta_n}^{n-1}(e_{i+1,k} - e_{i+1,k^*_{n,\delta_n}})^2\right\}^{q/2}\right] \\ &\le \frac{C}{n^{q/2}(1-\delta_n)^{q/2}}\frac{1}{n(1-\delta_n)}\sum_{i=n\delta_n}^{n-1}E(|e_{i+1,k} - e_{i+1,k^*_{n,\delta_n}}|^q) \\ &\le \frac{C\|\mathbf{a}(k) - \mathbf{a}(k^*_{n,\delta_n})\|_R^q}{(1-\delta_n)^{q/2}n^{q/2}}. \end{aligned} \tag{A.35}$$

Consequently, (A.31) is ensured by (A.33)-(A.35). The proof is completed by noting that (A.32) is an immediate consequence of (A.35). □

Armed with Lemmas A.6-A.9, we have the following result.

Corollary A.1. *Let (K.1)–(K.5), (3.5) and (3.6) hold. Then, for any* $r > 0$*,*

$$\lim_{n\to\infty} E\left(\frac{L_n^{(\delta_n)}(\hat{k}_{n,\delta_n})}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1\right)^r = 0. \tag{A.36}$$

PROOF. Let $\epsilon > 0$ be arbitrarily given. By (A.3), one has for $r > 0$,

$$\begin{aligned} E\left(\frac{L_n^{(\delta_n)}(\hat{k}_{n,\delta_n})}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1\right)^r &= \sum_{k=1}^{K_n}\left(\frac{L_n^{(\delta_n)}(k)}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1\right)^r P\left(\hat{k}_{n,\delta_n} = k\right) \\ \leq\ \epsilon^r &+ \sum_{k\in A_{\epsilon,n}^{(\delta_n)}}\left(\frac{L_n^{(\delta_n)}(k)}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1\right)^r P\left(\hat{k}_{n,\delta_n} = k\right) \\ \leq\ \epsilon^r &+ \sum_{l=1}^{12}\left\{\sum_{k\in A_{\epsilon,n}^{(\delta_n)}}\left(\frac{L_n^{(\delta_n)}(k)}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1\right)^r P\left(N_l(k) \geq \frac{1}{12}V_{n,\delta_n}(k)\right)\right\}, \end{aligned} \tag{A.37}$$

where

$$A_{\epsilon,n}^{(\delta_n)} = \left\{k:\ 1 \leq k \leq K_n,\quad \frac{L_n^{(\delta_n)}(k)}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1 > \epsilon\right\}.$$

In view of (A.37), (A.36) holds if for $l = 1, \cdots, 12$,

$$\lim_{n\to\infty}\sum_{k\in A_{\epsilon,n}^{(\delta_n)}}\left(\frac{L_n^{(\delta_n)}(k)}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1\right)^r P\left(N_l(k) > \frac{1}{12}V_{n,\delta_n}(k)\right) = 0. \tag{A.38}$$

In the following, we only prove (A.38) for $l = 1, 3$, and 11 because the proofs for $l = 2, 7, 8, 9$, and 10 are similar to that for $l = 1$, the proofs for $l = 4, 5$, and 6 are similar to that for $l = 3$, and the proof for $l = 12$ is similar to that for $l = 11$.

By (A.7), Chebyshev's inequality, (3.5), (3.6) and the facts that

$$L_n^{(\delta_n)}(k) \geq \|\mathbf{a} - \mathbf{a}(k)\|_R^2, nL_n^{(\delta_n)}(k) \geq C\frac{k\log\delta_n^{-1}}{1-\delta_n} \tag{A.39}$$

and $L_n^{(\delta_n)}(k)/L_n^{(\delta_n)}(k^*_{n,\delta_n}) \le C/L_n^{(\delta_n)}(k^*_{n,\delta_n})$ if $1 \le k \le k^*_{n,\delta_n}$ and $L_n^{(\delta_n)}(k)/L_n^{(\delta_n)}(k^*_{n,\delta_n}) \le Ck/k^*_{n,\delta_n}$ if $k^*_{n,\delta_n} < k \le K_n$, we have for sufficiently large q,

$$
\begin{aligned}
& \sum_{k\in A_{\epsilon,n}^{(\delta_n)}} \left(\frac{L_n^{(\delta_n)}(k)}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1 \right)^r P\left(N_1(k) > \frac{1}{12} V_{n,\delta_n}(k) \right) \\
\le\ & C \sum_{k\in A_{\epsilon,n}^{(\delta_n)}} \left(\frac{L_n^{(\delta_n)}(k)}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} \right)^r (V_{n,\delta_n}(k))^{-(q-r)} \left\{ \frac{k^q}{(\log \delta_n^{-1})^q (n\delta_n)^{q/2}} + \frac{(\frac{\log \delta_n^{-1}}{1-\delta_n})^q k^q}{n^q} \right\} \\
\le\ & C\left(\frac{1+\epsilon}{\epsilon}\right)^{q-r} \left[\sum_{k=1}^{k^*_{n,\delta_n}} \left\{ \frac{k^q}{(k^*_{n,\delta_n})^r (\log \delta_n^{-1})^{q+r} (n\delta_n)^{(q/2)-r} (\frac{\delta_n}{1-\delta_n})^r} + \frac{(\frac{\log \delta_n^{-1}}{1-\delta_n})^{q-r} k^q}{(k^*_{n,\delta_n})^r n^{q-r}} \right\} \right. \\
+ & \left. \sum_{k=k^*_{n,\delta_n}+1}^{K_n} \left\{ \frac{k^{q+r}}{(k^*_{n,\delta_n})^r (\log \delta_n^{-1})^q (n\delta_n)^{q/2}} + \frac{(\frac{\log \delta_n^{-1}}{1-\delta_n})^q k^{q+r}}{(k^*_{n,\delta_n})^r n^q} \right\} \right] = o(1). \qquad \text{(A.40)}
\end{aligned}
$$

Therefore, (A.38) holds for $l = 1$.

By (A.21), (A.31), an argument similar to that used for obtaining (A.40) and the fact that $k^*_{n,\delta_n} \to \infty$ as $n \to \infty$, we have for sufficiently large q,

$$
\begin{aligned}
& \sum_{k\in A_{\epsilon,n}^{(\delta_n)}} \left(\frac{L_n^{(\delta_n)}(k)}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1 \right)^r P\left(N_3(k) > \frac{1}{12} V_{n,\delta_n}(k) \right) \\
\le\ & C\left(\frac{1+\epsilon}{\epsilon}\right)^{q-r} \left\{ \sum_{k=1}^{K_n} \left(\frac{L_n^{(\delta_n)}(k)}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} \right)^r \frac{k^q}{(\log \delta_n^{-1})^q (n\delta_n)^{q/2}} \right. \\
+ & \sum_{k=1}^{k^*_{n,\delta_n}} \frac{k^{q/2}}{n^q (L_n^{(\delta_n)}(k^*_{n,\delta_n}))^q (1-\delta_n)^q} \\
+ & \left. \sum_{k=k^*_{n,\delta_n}+1}^{K_n} \frac{k^{q/2}}{n^q (L_n^{(\delta_n)}(k))^{q-r} (L_n^{(\delta_n)}(k^*_{n,\delta_n}))^r (1-\delta_n)^q} \right\} = o(1), \qquad \text{(A.41)}
\end{aligned}
$$

and

$$
\begin{aligned}
& \sum_{k\in A_{\epsilon,n}^{(\delta_n)}} \left(\frac{L_n^{(\delta_n)}(k)}{L_n^{(\delta_n)}(k^*_{n,\delta_n})} - 1 \right)^r P\left(N_{11}(k) > \frac{1}{12} V_{n,\delta_n}(k) \right) \\
\le\ & C\left(\frac{1+\epsilon}{\epsilon}\right)^{q-r} \left\{ \sum_{k=1}^{k^*_{n,\delta_n}} \frac{1}{n^{q/2} (L_n^{(\delta_n)}(k^*_{n,\delta_n}))^{q/2} (1-\delta_n)^{q/2}} \right.
\end{aligned}
$$

$$+ \quad \sum_{k=k^*_{n,\delta_n}+1}^{K_n} \frac{1}{n^{q/2}(L_n^{(\delta_n)}(k))^{(q/2)-r}(L_n^{(\delta_n)}(k^*_{n,\delta_n}))^r(1-\delta_n)^{q/2}} \Bigg\} = o(1). \quad \text{(A.42)}$$

In view of (A.40)-(A.42), the proof is complete. □

Corollary A.2. *Assume that (K.1)–(K.5) and condition (i) of Theorem 1 hold. Also assume that* δ_n *satisfies (3.5) and (3.6). Then, for sufficiently large* q*,*

$$E\left|\frac{\mathcal{S}(\hat{k}_{n,\delta_n}) - \mathcal{S}(k^*_{n,\delta_n})}{(L_n^{(\delta_n)}(\hat{k}_{n,\delta_n}))^{1/2}}\right|^{2q} = O((k^*_{n,\delta_n})^{-(1-\theta)q+\theta}) + o((\log \delta_n^{-1})^{-q}) + O((\log \delta_n^{-1})^{-q/2}(k^*_{n,\delta_n})^{(-q/2)+\theta}), \quad \text{(A.43)}$$

where $\mathcal{S}(k)$ *is defined in Section 2 and* $0 \leq \theta = \theta(\xi) < 1$ *is obtained from condition (i) of Theorem 1 with* ξ *being limited to the open interval* $(0, \min\{1/2, \{(2+\delta_1^*)(1-\theta_3)/2\} - 1\})$.

PROOF. Let $0 < \xi < \min\{1/2, \{(2+\delta_1^*)(1-\theta_3)/2\} - 1\}$. Then, by condition (i) of Theorem 1, there are $0 \leq \theta = \theta(\xi) < 1$ and $M = M(\xi) > 0$ such that (3.7) is satisfied. By Hölder's inequality and the fact that for any $h > 0$,

$$E\left|\mathcal{S}(k) - \mathcal{S}(k^*_{n,\delta_n})\right|^{2h} \leq C\left\|\mathbf{a}(k) - \mathbf{a}(k^*_{n,\delta_n})\right\|_R^{2h}$$
$$\leq C\left(L_n^{(\delta_n)}(k) - L_n^{(\delta_n)}(k^*_{n,\delta_n}) + \frac{\log \delta_n^{-1}}{1-\delta_n}\left|\frac{k - k^*_{n,\delta_n}}{N}\right|\sigma^2\right)^h \quad \text{(A.44)}$$

(which follows from Lemma A.0, (K.3) and the definition of k^*_{n,δ_n}), we have for $q > 0$ and $1 < r < \infty$,

$$E\left|\frac{\mathcal{S}(\hat{k}_{n,\delta_n}) - \mathcal{S}(k^*_{n,\delta_n})}{(L_n^{(\delta_n)}(\hat{k}_{n,\delta_n}))^{1/2}}\right|^{2q} \leq \sum_{k=1}^{K_n}\left(E\left|\frac{\mathcal{S}(k) - \mathcal{S}(k^*_{n,\delta_n})}{(L_n^{(\delta_n)}(k))^{1/2}}\right|^{2qr}\right)^{\frac{1}{r}} P^{\frac{r-1}{r}}\left(\hat{k}_{n,\delta_n} = k\right)$$

$$\leq C\sum_{k=1}^{K_n}\left\{V^q_{n,\delta_n}(k) + \left|\frac{\frac{\log \delta_n^{-1}}{1-\delta_n}(k - k^*_{n,\delta_n})}{NL_n^{(\delta_n)}(k)}\right|^q\right\} P^{\frac{r-1}{r}}\left(\hat{k}_{n,\delta_n} = k\right)$$

$$\leq C\Bigg\{\sum_{k=1}^{K_n} V^q_{n,\delta_n}(k) P^{\frac{r-1}{r}}\left(\hat{k}_{n,\delta_n} = k\right) + \sum_{\substack{k=1 \\ k \notin A^{(\delta_n)}_{n,\theta,M}}}^{K_n}\left|\frac{\frac{\log \delta_n^{-1}}{1-\delta_n}(k - k^*_{n,\delta_n})}{NL_n^{(\delta_n)}(k)}\right|^q$$

$$+ \sum_{\substack{k=1 \\ k \in A^{(\delta_n)}_{n,\theta,M}}}^{K_n}\left|\frac{\frac{\log \delta_n^{-1}}{1-\delta_n}(k - k^*_{n,\delta_n})}{NL_n^{(\delta_n)}(k)}\right|^q P^{\frac{r-1}{r}}\left(\hat{k}_{n,\delta_n} = k\right)\Bigg\}$$

$$\equiv \quad C\{(I)+(II)+(III)\}, \tag{A.45}$$

where $A_{n,\theta,M}^{(\delta_n)}$ is a set of positive integers defined in condition (i) of Theorem 1.

By the definitions of $A_{n,\theta,M}^{(\delta_n)}$, $L_n^{(\delta_n)}(k)$ and $L_n^{(\delta_n)}(k_{n,\delta_n}^*)$, it is easy to see that

$$(II) \leq C(k_{n,\delta_n}^*)^{-(1-\theta)q+\theta}. \tag{A.46}$$

Since for $a, b \geq 0, (a+b)^{(r-1)/r} \leq a^{(r-1)/r} + b^{(r-1)/r}$, we have by (3.4) that

$$(I) \leq \sum_{l=1}^{12} \left\{ \sum_{k=1}^{K_n} V_{n,\delta_n}^q(k) P^{\frac{r-1}{r}} \left(N_l(k) \geq (1/12) V_{n,\delta_n}(k)\right) \right\}. \tag{A.47}$$

In the following, we shall show that when q is sufficiently large,

$$\sum_{k=1}^{K_n} V_{n,\delta_n}^q(k) P^{\frac{r-1}{r}} \left(N_l(k) \geq (1/12) V_{n,\delta_n}(k)\right) = o((\log \delta_n^{-1})^{-q}), \tag{A.48}$$

for $l = 1, \cdots, 10$, and

$$\begin{aligned} &\sum_{k=1}^{K_n} V_{n,\delta_n}^q(k) P^{\frac{r-1}{r}} \left(N_l(k) \geq (1/12) V_{n,\delta_n}(k)\right) \\ = \; & O((\log \delta_n^{-1})^{-q/2} (k_{n,\delta_n}^*)^{(-q/2)+\theta}) + o((\log \delta_n^{-1})^{-q}), \end{aligned} \tag{A.49}$$

for $l = 11$ and 12. As a result, (A.47)-(A.49) yield that for sufficiently large q,

$$(I) = O((\log \delta_n^{-1})^{-q/2} (k_{n,\delta_n}^*)^{(-q/2)+\theta}) + o((\log \delta_n^{-1})^{-q}). \tag{A.50}$$

By (A.7), (3.5), (3.6), (K.4) and (A.39), we have for sufficiently large q,

$$\begin{aligned} &\sum_{k=1}^{K_n} V_{n,\delta_n}^q(k) P^{\frac{r-1}{r}} \left(N_1(k) \geq (1/12) V_{n,\delta_n}(k)\right) \leq C \sum_{k=1}^{K_n} [E\{N_1(k)\}^{qr/(r-1)}]^{(r-1)/r} \\ \leq \; & \frac{C}{(\log \delta_n^{-1})^q} \left(\sum_{k=1}^{K_n} \frac{k^q}{(n\delta_n)^{q/2}} + \frac{(\log \delta_n^{-1})^{2q} k^q}{(1-\delta_n)^q n^q} \right) = o((\log \delta_n^{-1})^{-q}), \end{aligned} \tag{A.51}$$

which guarantees that (A.48) holds for $l = 1$. For $l = 3$, according to (3.6), (A.21), (A.39) and the fact that $k_{n,\delta_n}^* \to \infty$ as $n \to \infty$, we have for sufficiently large q,

$$\begin{aligned} &\sum_{k=1}^{K_n} V_{n,\delta_n}^q(k) P^{\frac{r-1}{r}} \left(N_3(k) \geq (1/12) V_{n,\delta_n}(k)\right) \leq C \sum_{k=1}^{K_n} [E\{N_3(k)\}^{qr/(r-1)}]^{(r-1)/r} \\ \leq \; & \frac{C}{(\log \delta_n^{-1})^q} \left(\sum_{k=1}^{K_n} \frac{k^q}{(n\delta_n)^{q/2}} + \sum_{k=1}^{k_{n,\delta_n}^*} \frac{k^{q/2}}{k_{n,\delta_n}^{*q}} + \sum_{k=k_{n,\delta_n}^*+1}^{K_n} k^{-q/2} \right) \\ = \; & o((\log \delta_n^{-1})^{-q}). \end{aligned} \tag{A.52}$$

The proofs of (A.48) for $l = 2, 7, 8, 9,$ and 10 are similar to that of (A.51) and the proofs of (A.47) for $l = 4, 5$ and 6 are similar to that of (A.52). We skip the details in order to save space. The proof of (A.49) is a bit more complicated. By (3.7), Lemma A.9, (A.44), (3.5), the arguments used in (A.51) and (A.52), and the restriction on ξ, one has for sufficiently large q,

$$\begin{aligned}
&\sum_{k=1}^{K_n} V_{n,\delta_n}^q(k) P^{\frac{r-1}{r}}\left(N_l(k) \geq (1/12)V_{n,\delta_n}(k)\right) \\
\leq\ & \sum_{\substack{k=1 \\ k \notin A_{n,\theta,M}^{(\delta_n)}}}^{K_n} [E\{N_{11}(k)\}^{qr/(r-1)}]^{(r-1)/r} + \sum_{\substack{k=1 \\ k \in A_{n,\theta,M}^{(\delta_n)}}}^{K_n} V_{n,\delta_n}^{-q}(k)[E\{N_{11}(k)\}^{2qr/(r-1)}]^{(r-1)/r} \\
\leq\ & C \left\{ \sum_{\substack{k=1 \\ k \notin A_{n,\theta,M}^{(\delta_n)}}}^{K_n} \frac{\|\mathbf{a}(k) - \mathbf{a}(k_{n,\delta_n}^*)\|_R^q}{(1-\delta_n)^{q/2}(L_n^{(\delta_n)}(k))^q n^{q/2}} \right. \\
+\ & \left. \sum_{\substack{k=1 \\ k \in A_{n,\theta,M}^{(\delta_n)}}}^{K_n} \frac{(L_n^{(\delta_n)}(k) - L_n^{(\delta_n)}(k_{n,\delta_n}^*))^q + |\frac{(\log \delta_n^{-1})(k - k_{n,\delta_n}^*)}{(1-\delta_n)N}|^q}{(1-\delta_n)^q (L_n^{(\delta_n)}(k))^q n^q (L_n^{(\delta_n)}(k) - L_n^{(\delta_n)}(k_{n,\delta_n}^*))^q} \right\} \\
\leq\ & O\left((\log \delta_n^{-1})^{-q/2}(k_{n,\delta_n}^*)^{(-q/2)+\theta}\right) \\
+\ & \frac{C}{(\log \delta_n^{-1})^q}(1 + (k_{n,\delta_n}^*)^{\xi q}) \left\{ \sum_{k=1}^{k_{n,\delta_n}^*} k_{n,\delta_n}^{*\,-q} + \sum_{k=k_{n,\delta_n}^*+1}^{K_n} k^{-q} \right\} \\
=\ & O((\log \delta_n^{-1})^{-q/2}(k_{n,\delta_n}^*)^{(-q/2)+\theta}) + o((\log \delta_n^{-1})^{-q}), \qquad \text{(A.53)}
\end{aligned}$$

where $l = 11$ or 12.

Reasoning as for (A.47), we have

$$(III) \leq \sum_{l=1}^{12} \left\{ \sum_{\substack{k=1 \\ k \in A_{n,\theta,M}^{(\delta_n)}}}^{K_n} \left| \frac{\frac{\log \delta_n^{-1}}{1-\delta_n}(k - k_{n,\delta_n}^*)}{N L_n^{(\delta_n)}(k)} \right|^q P^{\frac{r-1}{r}}\left(N_l(k) \geq (1/12)V_{n,\delta_n}(k)\right) \right\}.$$

Since $0 < \xi < \min\{1/2, \{(2 + \delta_1^*)(1 - \theta_3)/2\} - 1\}$, by arguments similar to those used to prove (A.51) and (A.52), one obtains for sufficiently large q,

$$\sum_{\substack{k=1 \\ k \in A_{n,\theta,M}^{(\delta_n)}}}^{K_n} \left| \frac{\frac{\log \delta_n^{-1}}{1-\delta_n}(k - k_{n,\delta_n}^*)}{N L_n^{(\delta_n)}(k)} \right|^q P^{\frac{r-1}{r}}\left(N_1(k) \geq (1/12)V_{n,\delta_n}(k)\right)$$

$$\leq \quad \frac{Ck^{*\xi q}_{n,\delta_n}}{(\log\delta_n^{-1})^q}\left(\sum_{k=1}^{K_n}\frac{k^q}{(n\delta_n)^{q/2}}+\frac{(\log\delta_n^{-1})^{2q}k^q}{(1-\delta_n)^q n^q}\right)=o((\log\delta_n^{-1})^{-q}),$$

and

$$\sum_{\substack{k=1\\ k\in A_{n,\theta,M}^{(\delta_n)}}}^{K_n}\left|\frac{\frac{\log\delta_n^{-1}}{1-\delta_n}(k-k^*_{n,\delta_n})}{NL_n^{(\delta_n)}(k)}\right|^q P^{\frac{r-1}{r}}\left(N_3(k)\geq(1/12)V_{n,\delta_n}(k)\right)$$

$$\leq \quad \frac{Ck^{\xi q}_{n,\delta_n}}{(\log\delta_n^{-1})^q}\left(\sum_{k=1}^{K_n}\frac{k^q}{(n\delta_n)^{q/2}}+\sum_{k=1}^{k^*_{n,\delta_n}}\frac{k^{q/2}}{k^{*q}_{n,\delta_n}}+\sum_{k=k^*_{n,\delta_n}+1}^{K_n}k^{-q/2}\right)=o((\log\delta_n^{-1})^{-q}),$$

respectively. Similarly, it can be shown that for $l=2,4,5,6,7,8,9$ and 10,

$$\sum_{\substack{k=1\\ k\in A_{n,\theta,M}^{(\delta_n)}}}^{K_n}\left|\frac{\frac{\log\delta_n^{-1}}{1-\delta_n}(k-k^*_{n,\delta_n})}{NL_n^{(\delta_n)}(k)}\right|^q P^{\frac{r-1}{r}}\left(N_l(k)\geq(1/12)V_{n,\delta_n}(k)\right)=o((\log\delta_n^{-1})^{-q}).$$

By analogy with (A.53), we have

$$\sum_{\substack{k=1\\ k\in A_{n,\theta,M}^{(\delta_n)}}}^{K_n}\left|\frac{\frac{\log\delta_n^{-1}}{1-\delta_n}(k-k^*_{n,\delta_n})}{NL_n^{(\delta_n)}(k)}\right|^q P^{\frac{r-1}{r}}\left(N_l(k)\geq(1/12)V_{n,\delta_n}(k)\right)$$

$$\leq \quad \frac{C}{(\log\delta_n^{-1})^q}((k^*_{n,\delta_n})^{\xi q}+(k^*_{n,\delta_n})^{2\xi q})\left\{\sum_{k=1}^{k^*_{n,\delta_n}}k^{*-q}_{n,\delta_n}+\sum_{k=k^*_{n,\delta_n}+1}^{K_n}k^{-q}\right\}=o((\log\delta_n^{-1})^{-q}),$$

where $l=11$ and 12. Hence,

$$(III)=o((\log\delta_n^{-1})^{-q}) \tag{A.54}$$

holds for sufficiently large q. Consequently, (A.43) follows from (A.45), (A.46), (A.50) and (A.54). □

Corollary A.3. *Assume that the assumptions of Corollary A.2 hold. Then, for sufficiently large* q*,*

$$\lim_{n\to\infty}E\left|\frac{\mathbf{f}(\hat{k}_{n,\delta_n})-\mathbf{f}(k^*_{n,\delta_n})}{(L_n^{(\delta_n)}(\hat{k}_{n,\delta_n}))^{1/2}}\right|^{2q}=o((\log\delta_n^{-1})^q). \tag{A.55}$$

PROOF. Define

$$\mathbf{f}_1(k) = \mathbf{x}_n^{*'}(k) R^{-1}(k) \frac{1}{N} \sum_{j=K_n}^{n-\sqrt{n}-1} \mathbf{x}_j(k) e_{j+1},$$

where $\mathbf{x}_j^{*'}(k) = (x_j^*, \cdots, x_{j-k+1}^*) = (\sum_{r=0}^{\sqrt{n}/2-K_n} b_r e_{j-r}, \cdots, \sum_{r=0}^{\sqrt{n}/2-K_n} b_r e_{j-k+1-r})$. (Note that $\mathbf{f}_1(k)$ can be used to approximate $\mathbf{f}(k)$ as defined after (2.1).) Since Lemma 8 of Ing and Wei (2005) shows that for $q > 0$,

$$\max_{\substack{1 \leq i, l \leq K_n \\ i \neq l}} \frac{E \mid \mathbf{f}_1(i) - \mathbf{f}_1(l) \mid^{2q}}{\left| \dfrac{i-l}{N} \right|^q} \leq C, \tag{A.56}$$

by Hölder's inequality and (A.56), one obtains for $q > 0$ and $1 < r < \infty$,

$$\begin{aligned} E \left| \frac{\mathbf{f}_1(\hat{k}_{n,\delta_n}) - \mathbf{f}_1(k^*_{n,\delta_n})}{(L_n^{(\delta_n)}(\hat{k}_{n,\delta_n}))^{1/2}} \right|^{2q} &\leq \sum_{k=1}^{K_n} \left(E \left| \frac{\mathbf{f}_1(k) - \mathbf{f}_1(k^*_{n,\delta_n})}{(L_n^{(\delta_n)}(k))^{1/2}} \right|^{2qr} \right)^{\frac{1}{r}} P^{\frac{r-1}{r}} \Big(\hat{k}_{n,\delta_n} = k \Big) \\ &\leq C \sum_{k=1}^{K_n} \left| \frac{k - k^*_{n,\delta_n}}{N L_n^{(\delta_n)}(k)} \right|^q P^{\frac{r-1}{r}} \left(\hat{k}_{n,\delta_n} = k \right). \end{aligned} \tag{A.57}$$

Let $0 < \xi < \min\{1/2, \{(2+\delta_1^*)(1-\theta_3)/2\} - 1\}$. Then, condition (i) of Theorem 1 guarantees that there are $0 \leq \theta = \theta(\xi) < 1$ and $M = M(\xi) > 0$ such that (3.7) holds. By arguments similar to those used to verify (A.45), (A.46), (A.50) and (A.54), we have for sufficiently large q,

$$\begin{aligned} & \sum_{k=1}^{K_n} \left| \frac{k - k^*_{n,\delta_n}}{N L_n^{(\delta_n)}(k)} \right|^q P^{\frac{r-1}{r}} \left(\hat{k}_{n,\delta_n} = k \right) \\ \leq\ & C \left\{ \sum_{\substack{k=1 \\ k \notin A^{(\delta_n)}_{n,\theta,M}}}^{K_n} \left| \frac{k - k^*_{n,\delta_n}}{N L_n^{(\delta_n)}(k)} \right|^q + \sum_{\substack{k=1 \\ k \in A^{(\delta_n)}_{n,\theta,M}}}^{K_n} \left| \frac{k - k^*_{n,\delta_n}}{N L_n^{(\delta_n)}(k)} \right|^q P^{\frac{r-1}{r}} \left(\hat{k}_{n,\delta_n} = k \right) \right\} \\ \leq\ & C \left\{ \left(\frac{\log \delta_n^{-1}}{1 - \delta_n} \right)^{-q} k^{*(\theta-1)q}_{n,\delta_n} k^{*\theta}_{n,\delta_n} + \sum_{\substack{k=1 \\ k \in A^{(\delta_n)}_{n,\theta,M}}}^{K_n} \left| \frac{k - k^*_{n,\delta_n}}{N L_n^{(\delta_n)}(k)} \right|^q P^{\frac{r-1}{r}} \left(\hat{k}_{n,\delta_n} = k \right) \right\} \\ \leq\ & o((\log \delta_n^{-1})^{-q}) + C \sum_{l=1}^{12} \left\{ \sum_{\substack{k=1 \\ k \in A^{(\delta_n)}_{n,\theta,M}}}^{K_n} \left| \frac{k - k^*_{n,\delta_n}}{N L_n^{(\delta_n)}(k)} \right|^q P^{\frac{r-1}{r}} \left(N_l(k) \geq (1/12) V_{n,\delta_n}(k) \right) \right\} \\ =\ & o((\log \delta_n^{-1})^{-q}) + o((\log \delta_n^{-1})^{-2q}), \end{aligned}$$

where $A_{n,\theta,M}^{(\delta_n)}$ is a set of positive integers defined in condition (i) of Theorem 1. This, together with (A.57), yields

$$E\left|\frac{\mathbf{f}_1(\hat{k}_{n,\delta_n})-\mathbf{f}_1(k^*_{n,\delta_n})}{(L_n^{(\delta_n)}(\hat{k}_{n,\delta_n}))^{1/2}}\right|^{2q}=o((\log\delta_n^{-1})^{-q}). \tag{A.58}$$

Moreover, by the same argument as in the proof of Ing and Wei (2005, Lemma 7), we have for $q>0$,

$$E\max_{1\le k\le K_n}\left|\frac{\mathbf{f}(k)-\mathbf{f}_1(k)}{(L_n^{(\delta_n)}(k))^{1/2}}\right|^{2q}=o\left((\log\delta_n^{-1})^{q}\right). \tag{A.59}$$

Consequently, (A.55) follows from (A.58) and (A.59). □

We are now ready to prove Theorem 1.

PROOF OF THEOREM 1. First observe that

$$\begin{aligned}\frac{E(x_{n+1}-\hat{x}_{n+1}(\hat{k}_{n,\delta_n}))^2-\sigma^2}{L_n(k^*_{n,\delta_n})}&=E\left[\frac{\{\mathbf{f}(\hat{k}_{n,\delta_n})+\mathcal{S}(\hat{k}_{n,\delta_n})\}^2}{L_n(k^*_{n,\delta_n})}\right]\\ =\ E&\left[\frac{\{\mathbf{f}(\hat{k}_{n,\delta_n})-\mathbf{f}(k^*_{n,\delta_n})+\mathcal{S}(\hat{k}_{n,\delta_n})-\mathcal{S}(k^*_{n,\delta_n})+\mathbf{f}(k^*_{n,\delta_n})+\mathcal{S}(k^*_{n,\delta_n})\}^2}{L_n(k^*_{n,\delta_n})}\right].\end{aligned} \tag{A.60}$$

By Corollary A.2, (A.43) follows. Let $q>1$ if the θ on the right-hand side of (A.43) equals 0, and $q>\max\{\theta/(1-\theta-\eta),2\theta/(1-\eta),1\}$ otherwise, where η is defined in (3.8). Then,

$$\begin{aligned}&E\left[\frac{\{\mathcal{S}(\hat{k}_{n,\delta_n})-\mathcal{S}(k^*_{n,\delta_n})\}^2}{L_n(k^*_{n,\delta_n})}\right]\\ =\ &E\left[\frac{\{\mathcal{S}(\hat{k}_{n,\delta_n})-\mathcal{S}(k^*_{n,\delta_n})\}^2}{L_n^{(\delta_n)}(\hat{k}_{n,\delta_n})}\frac{L_n^{(\delta_n)}(\hat{k}_{n,\delta_n})}{L_n^{(\delta_n)}(k^*_{n,\delta_n})}\frac{L_n^{(\delta_n)}(k^*_{n,\delta_n})}{L_n(k^*_{n,\delta_n})}\right]\\ \le\ &C\frac{\log\delta_n^{-1}}{1-\delta_n}E\left[\frac{\{\mathcal{S}(\hat{k}_{n,\delta_n})-\mathcal{S}(k^*_{n,\delta_n})\}^{2q}}{(L_n^{(\delta_n)}(\hat{k}_{n,\delta_n}))^q}\right]^{1/q}E\left[\left\{\frac{L_n^{(\delta_n)}(\hat{k}_{n,\delta_n})}{L_n^{(\delta_n)}(k^*_{n,\delta_n})}\right\}^{q/(q-1)}\right]^{(q-1)/q}\\ =\ &O\left(\frac{\log\delta_n^{-1}}{(k^*_{n,\delta_n})^{1-\theta(1+\frac{1}{q})}}\right)+o(1)+O\left(\left\{\frac{\log\delta_n^{-1}}{(k^*_{n,\delta_n})^{1-\frac{2\theta}{q}}}\right\}^{1/2}\right)=o(1)\end{aligned} \tag{A.61}$$

where the inequality follows from the fact that $L_n^{(\delta_n)}(k^*_{n,\delta_n})\le L_n(k^*_{n,\delta_n})\log\delta_n^{-1}(1-\delta_n)^{-1}$ and Hölder's inequality, the second equality follows from (3.5), Corollaries

A.1 and A.2 and Jensen's inequality and the last equality is ensured by (3.8). By Corollaries A.1 and A.3 and an argument similar to that used to prove (A.61),

$$E\left[\frac{\{\mathbf{f}(\hat{k}_{n,\delta_n}) - \mathbf{f}(k^*_{n,\delta_n})\}^2}{L_n(k^*_{n,\delta_n})}\right] = o(1). \tag{A.62}$$

Consequently, the desired result is ensured by (A.60)-(A.62), Proposition 1 and the Cauchy-Schwarz inequality. □

PROOF OF THEOREM 2. First note that when $\lim_{n\to\infty}\delta_n = 1$ and condition (i) (or (ii)) of Theorem 2 are assumed instead of (3.5) and (3.6), the left-hand sides of (A.40), (A.41) and (A.42) still converge to 0. Therefore, (A.36) follows. Let $0 < \xi < (1/2) - \xi_2$ if condition (i) of Theorem 2 holds, and $0 < \xi < \min\{(1/2)-\xi_2, (\delta_1^*/2)-\xi_2\}$ if condition (ii) of Theorem 2 holds. Since condition (i) of Theorem 1 is given, there are $0 \le \theta = \theta(\xi) < 1$ and $M = M(\xi) > 0$ such that (3.7) holds. By the same reasoning used in the proof of Corollary A.2 and Jensen's inequality, we have for any $q > 0$,

$$\lim_{n\to\infty} E\left|\frac{\mathcal{S}(\hat{k}_{n,\delta_n}) - \mathcal{S}(k^*_{n,\delta_n})}{(L_n^{(\delta_n)}(\hat{k}_{n,\delta_n}))^{1/2}}\right|^{2q} = 0, \tag{A.63}$$

and

$$\lim_{n\to\infty} E\left|\frac{\mathbf{f}(\hat{k}_{n,\delta_n}) - \mathbf{f}(k^*_{n,\delta_n})}{(L_n^{(\delta_n)}(\hat{k}_{n,\delta_n}))^{1/2}}\right|^{2q} = 0, \tag{A.64}$$

respectively. Consequently, the claimed result follows from (A.36), (A.63), (A.64), $\lim_{n\to\infty}\delta_n = 1$ and an argument similar to the one given in the proof of Theorem 1. □

Appendix B: Proof of Theorem 3

Instead of verifying (4.7) directly, we will first show that (4.7) holds with $\hat{k}_{n,P_n}$ replaced by $\hat{k}^S_{n,P_n}$, where $\hat{k}^S_{n,P_n} = \arg\min_{1\le k\le K_n} S_n^{(P_n)}(k)$ and $S_n^{(P_n)}(k) = (N + P_n k)\hat{\sigma}_n^2(k)$. By an analogy with (4.1) of Shibata (1980),

$$\begin{aligned} S_n^{(P_n)}(k) &= N L_n(k) + P_n k\left(\hat{\sigma}_n^2(k) - \sigma^2\right) + \left(k\sigma^2 - N\,\|\hat{a}(k) - a(k)\,\|^2_{\hat{R}_n(k)}\right) \\ &\quad + N\sigma^2 + N\left(S^2_{K_n,n-1}(k) - \sigma_k^2\right), \end{aligned} \tag{B.1}$$

where the definition of $S^2_{K_n,n-1}(k)$ can be found in Lemma A.5. Based on (B.1) and an argument similar to that used in (5.34) of Ing and Wei (2005), we have

$$P(\hat{k}^S_{n,P_n} = k) \leq \sum_{i=1}^{5} P\ (U_{in}(k) \geq (\,1/5\,)\,U_n(k)\,)\,, \tag{B.2}$$

where

$$\begin{aligned} U_{1,n}(k) &= \left| \frac{P_n k\,(\,\hat{\sigma}^2_n(k) - \sigma^2\,)}{N L_{n,P_n}(k)} \right|, \quad U_{2,n}(k) = \left| \frac{P_n k^*_{n,P_n}\left(\,\hat{\sigma}^2_n(k^*_{n,P_n}) - \sigma^2\,\right)}{N L_{n,P_n}(k)} \right|, \\ U_{3,n}(k) &= \left| \frac{k\sigma^2 - N\,\|\,\hat{\mathbf{a}}_n(k) - \mathbf{a}(k)\,\|^2_{\hat{R}_n(k)}}{N L_{n,P_n}(k)} \right|, \\ U_{4,n}(k) &= \left| \frac{k^*_{n,P_n}\sigma^2 - N\left\|\,\hat{\mathbf{a}}_n(k^*_{n,P_n}) - \mathbf{a}(k^*_{n,P_n})\,\right\|^2_{\hat{R}_n(k^*_{n,P_n})}}{N L_{n,P_n}(k)} \right|, \\ U_{5,n}(k) &= \left| \frac{S^2_{K_n,n-1}(k) - \sigma^2_k - S^2_{K_n,n-1}(k^*_{n,P_n}) - \sigma^2_{k^*_{n,P_n}}}{L_{n,P_n}(k)} \right|, \end{aligned}$$

and

$$U_n(k) = \frac{L_{n,P_n}(k) - L_{n,P_n}(k^*_{n,P_n})}{L_{n,P_n}(k)}.$$

(Note that k^*_{n,P_n} and $L_{n,P_n}(k)$ are defined in (4.1) and (4.2), respectively.)

Theorem B.1. *Let the assumptions of Theorem 3 hold. Then,*

$$\lim_{n\to\infty} \frac{q_n(\hat{k}^S_{n,P_n}) - \sigma^2}{L_n(k^*_{n,P_n})} = 1.$$

PROOF. By an analogy with (5.43) of Ing and Wei (2005), we have for $q > 0$, all $1 \leq k \leq K_n$ and all sufficiently large n,

$$EU^q_{1,n}(k) \leq C\left(\frac{P^q_n k^q}{N^q} + N^{-q/2}\right), \tag{B.3}$$

and

$$EU^q_{2,n}(k) \leq C\left(\frac{P^q_n k^{*q}_{n,P_n}}{N^q} + N^{-q/2}\right). \tag{B.4}$$

The same reasoning as in (5.47) of Ing and Wei (2005) yields for $q > 0$, all $1 \leq k \leq K_n$ and all sufficiently large n,

$$EU_{3,n}^q(k) \leq C\left(\frac{k^q}{N^{q/2}} + \frac{k^{q/2}}{N^q L_{n,P_n}^q(k)}\right), \tag{B.5}$$

and

$$EU_{4,n}^q(k) \leq C\left(\frac{k_{n,P_n}^{*q}}{N^{q/2}} + \frac{k_{n,P_n}^{*q/2}}{N^q L_{n,P_n}^q(k)}\right). \tag{B.6}$$

As an immediate consequence of Lemma A.5 and Jensen's inequality,

$$EU_{5,n}^q(k) \leq C\frac{\|\mathbf{a}(k) - \mathbf{a}(k_{n,P_n}^*)\|_R^q}{N^{q/2}L_{n,P_n}^q(k)} \tag{B.7}$$

holds for $q > 0$ and all $1 \leq k \leq K_n$. Using (B.3)-(B.7) and an argument similar to that used to verify Corollary A.1, we have for $q > 0$,

$$\lim_{n\to\infty} E\left(\frac{L_{n,P_n}(\hat{k}_{n,P_n}^S)}{L_{n,P_n}(k_{n,P_n}^*)} - 1\right)^q = 0. \tag{B.8}$$

Let $0 < \xi < \{\delta_1^*/(4+2\delta_1^*)\} - \theta_3\}$. (Recall that θ_3 is some positive number less than $\delta_1^*/(4+2\delta_1^*)$; see (4.4).) By condition (i) of Theorem 3, there are $0 \leq \theta = \theta(\xi) < 1$ and $M = M(\xi) > 0$ such that (4.5) holds. With helps of (4.3)-(4.5), (B.3)-(B.7) and the restriction on ξ, we can follow the ideas of the proofs given in Corollaries A.2 and A.3 and obtain for $q > 0$,

$$\begin{aligned} E\left|\frac{\mathcal{S}(\hat{k}_{n,P_n}^S) - \mathcal{S}(k_{n,P_n}^*)}{(L_{n,P_n}(\hat{k}_{n,P_n}^S))^{1/2}}\right|^{2q} &= O((k_{n,P_n}^*)^{-(1-\theta)q+\theta}) + o((P_n-1)^{-q}) \\ &+ O((P_n-1)^{-q/2}(k_{n,P_n}^*)^{(-q/2)+\theta}), \end{aligned} \tag{B.9}$$

and

$$\lim_{n\to\infty} E\left|\frac{\mathbf{f}(\hat{k}_{n,P_n}^S) - \mathbf{f}(k_{n,P_n}^*)}{(L_{n,P_n}(\hat{k}_{n,P_n}^S))^{1/2}}\right|^{2q} = o((P_n-1)^{-q}). \tag{B.10}$$

Consequently, Theorem B.1 is guaranteed by (4.6), (B.8)-(B.10) and the same argument that we used in the proof of Theorem 1. □

PROOF OF THEOREM 3. In view of the proof of Theorem B.1, (4.7) is ensured by showing that (B.8)-(B.10) hold with $\hat{k}_{n,P_n}^S$ replaced by $\hat{k}_{n,P_n}$. Define

$$G_n(k) = N\exp\{IC_{P_n}(k)\} - S_n^{(P_n)}(k)$$

and

$$U_{6,n}(k) = \frac{|G_n(k) - G_n(k^*_{n,P_n})|}{NL_{n,P_n}(k)}.$$

First note that

$$\begin{aligned} P(\hat{k}_{n,P_n} = k) &\leq P(N\exp\{IC_{P_n}(k)\} \leq N\exp\{IC_{P_n}(k^*_{n,P_n})\}) \\ &\leq \sum_{i=1}^{6} P\ (U_{in}(k) \geq (\,1/6\,)\,U_n(k)\,)\,. \end{aligned} \tag{B.11}$$

In addition, Taylor's expansion and (5.42) of Ing and Wei (2005) give for $q > 0$,

$$EU^q_{6,n}(k) \leq C\left(\frac{P_n^2 K_n^2}{N^2 L_{n,P_n}(k)}\right)^q. \tag{B.12}$$

By (B.3)-(B.7), (B.11), (B.12) and the same reasoning used in the proof of Theorem B.1, the desired results follow. □

Appendix C: Proofs of Theorems 4-6

PROOF OF THEOREM 4. First note for sufficiently large n, we have $p_0 \leq K_n$ and

$$L_n(k^*_n) = \frac{p_0\sigma^2}{N}. \tag{C.1}$$

Thus, for $r > 1$,

$$\begin{aligned} &E\frac{\left\{\mathbf{f}(\hat{k}_{n,\delta_n}) + \mathcal{S}(\hat{k}_{n,\delta_n})\right\}^2}{L_n(k^*_n)} \leq \sum_{k=1}^{p_0-1}\left(E\left|\frac{\mathbf{f}(k)+\mathcal{S}(k)}{(\frac{p_0\sigma^2}{N})^{1/2}}\right|^{2r}\right)^{1/r} P^{\frac{r-1}{r}}(\hat{k}_{n,\delta_n} = k) \\ +\ &\sum_{k=p_0+1}^{K_n}\left(E\left|\frac{\mathbf{f}(k)}{(\frac{p_0\sigma^2}{N})^{1/2}}\right|^{2r}\right)^{1/r} P^{\frac{r-1}{r}}(\hat{k}_{n,\delta_n} = k) + E\left(\frac{N\mathbf{f}^2(k)}{p_0\sigma^2}\right) \\ \equiv\ &(I) + (II) + (III). \end{aligned} \tag{C.2}$$

By Lemmas A.1-A.3, (A.11) and the fact that $E(\mathcal{S}(k))^{2r} \leq C\|\mathbf{a} - \mathbf{a}(k)\|^{2r}_R$ (see (A.15)), we have for $1 \leq k \leq p_0 - 1$,

$$E\left|\frac{\mathbf{f}(k)+\mathcal{S}(k)}{(\frac{p_0\sigma^2}{N})^{1/2}}\right|^{2r} \leq CN^r, \tag{C.3}$$

and for $p_0 + 1 \leq k \leq K_n$

$$E\left|\frac{\mathbf{f}(k)}{(\frac{p_0\sigma^2}{N})^{1/2}}\right|^{2r} \leq Ck^r. \tag{C.4}$$

According to (C.3), (C.4), Lemmas A.6-A.9, the fact that for $1 \leq k \leq K_n$ and $k \neq p_0$, $V_{n,\delta_n}^{-1}(k) \leq C$ and the conditions imposed on δ_n, we can modify the argument given in the proof of Corollary A.1 to obtain that for any $s > 0$,

$$(I) = O(n^{-s}) \text{ and } (II) = o((\log \delta_n^{-1})^{-s}). \tag{C.5}$$

In addition, since by Proposition 1, $(III) = 1 + o(1)$, this, together with (C.5) and (C.2), yields that $\hat{k}_{n,\delta_n}$ satisfies (2.3). In addition, using arguments given above and in the proof of Theorem 3, it can be shown that (2.3) holds for $\hat{k}_n = \hat{k}_{n,P_n}$ with P_n satisfying $P_n \to \infty$ and $P_n/n \to 0$. The details are omitted in order to save space. □

PROOF OF THEOREM 5. When (ii) of (K.5′) holds, we have showed in Examples 1 and 6 that APE_{δ_n} and IC_{P_n}, with δ_n and P_n satisfying the conditions imposed in this theorem, are asymptotically efficient. This and Theorem 4 together yield the claimed result. □

PROOF OF THEOREM 6. Our goal is to show that

$$\limsup_{n\to\infty} \frac{q_n(\hat{k}_n^{(\iota)}) - \sigma^2}{L_n(k_n^*)} \leq 1. \tag{C.6}$$

To verify (C.6), first assume that condition (ii) holds. Choose ξ in condition (ii) to satisfy

$$0 < \xi < \min\{\delta_1^*/2, 1/2\}. \tag{C.7}$$

Then, there are $0 \leq \theta = \theta(\xi) < 1$ and $M = M(\xi) > 0$ such that (5.2) is fulfilled. Define

$$B_{n,M^*} = A_{P_n,\theta,M}^C \cap \{k : 1 \leq k \leq K_n, \frac{L_{n,P_n}(k) - L_{n,P_n}(k_{n,P_n}^*)}{L_{n,P_n}(k_{n,P_n}^*)} < M^*\},$$

where M^* is some positive constant. Then,

$$\begin{aligned} \frac{q_n(\hat{k}_n^{(\iota)}) - \sigma^2}{L_n(k_n^*)} &= E\left\{\frac{(\mathbf{f}(\hat{k}_{n,2}) + \mathcal{S}(\hat{k}_{n,2}))^2}{L_n(k_n^*)} I_{\{\hat{k}_{n,P_n} \neq \hat{k}_{n^\iota,P_{n^\iota}}\}}\right\} \\ + \quad & E\left\{\frac{(\mathbf{f}(\hat{k}_{n,P_n}) + \mathcal{S}(\hat{k}_{n,P_n}))^2}{L_n(k_n^*)} I_{\{\hat{k}_{n,P_n} = \hat{k}_{n^\iota,P_{n^\iota}}\}} \left(I_{\{\hat{k}_{n,P_n} \in B_{n,M^*}\}} + I_{\{\hat{k}_{n,P_n} \notin B_{n,M^*}\}}\right)\right\} \\ \equiv \quad & (I) + (II). \end{aligned} \tag{C.8}$$

Observe that for $r > 1$,

$$
\begin{aligned}
(II) &\leq E\left\{\frac{(\mathbf{f}(\hat{k}_{n^\iota,P_{n^\iota}})+\mathcal{S}(\hat{k}_{n^\iota,P_{n^\iota}}))^2}{L_n(k_n^*)} I_{\{\hat{k}_{n^\iota,P_{n^\iota}}\in B_{n,M^*}\}}\right\} \\
&+ E\left\{\frac{(\mathbf{f}(\hat{k}_{n,P_n})+\mathcal{S}(\hat{k}_{n,P_n}))^2}{L_n(k_n^*)} I_{\{\hat{k}_{n,P_n}\notin B_{n,M^*}\}}\right\} \\
&= \sum_{\substack{k=1\\ k\in B_{n,M^*}}}^{K_{n^\iota}} E\left\{\frac{(\mathbf{f}(k)+\mathcal{S}(k))^2}{L_n(k_n^*)} I_{\{\hat{k}_{n^\iota,P_{n^\iota}}=k\}}\right\} \\
&+ \sum_{\substack{k=1\\ k\notin B_{n,M^*}}}^{K_n} E\left\{\frac{(\mathbf{f}(k)+\mathcal{S}(k))^2}{L_n(k_n^*)} I_{\{\hat{k}_{n,P_n}=k\}}\right\} \\
&\leq C\left\{\sum_{\substack{k=1\\ k\in B_{n,M^*}}}^{K_{n^\iota}} \frac{L_n(k)}{L_n(k_n^*)} P^{(r-1)/r}(\hat{k}_{n^\iota,P_{n^\iota}}=k)\right. \\
&+ \left.\sum_{\substack{k=1\\ k\notin B_{n,M^*}}}^{K_n} \frac{L_n(k)}{L_n(k_n^*)} P^{(r-1)/r}(\hat{k}_{n,P_n}=k)\right\} \equiv C\{(III)+(IV)\}, \quad \text{(C.9)}
\end{aligned}
$$

where the second inequality follows from Hölder's inequality and the fact that for all $1 \leq k \leq K_n$, $E|\mathbf{f}(k)+\mathcal{S}(k)|^{2r} \leq CL_n^r(k)$, which is ensured by Lemmas A.0-A.3 and (A.20).

In the following, we shall show that both (III) and (IV) converge to 0. To deal with (III), notice that by (5.3) the definition of B_{n,M^*},

$$B_{n,M^*} \cap \{1,2,\cdots,K_{n^\iota}\} \subseteq A_{P_{n^\iota},\theta,M}$$

holds eventually in n. Hence, when $B_{n,M^*} \cap \{1,2,\cdots,K_{n^\iota}\}$ is nonempty and n is sufficiently large, we have for all $k \in B_{n,M^*} \cap \{1,2,\cdots,K_{n^\iota}\}$,

$$\frac{L_{n^\iota,P_{n^\iota}}(k)}{L_{n^\iota,P_{n^\iota}}(k) - L_{n^\iota,P_{n^\iota}}(k^*_{n^\iota,P_{n^\iota}})} \leq C(k^*_{n^\iota,P_{n^\iota}})^{\xi}. \quad \text{(C.10)}$$

The definition of B_{n,M^*} also yields for all $k \in B_{n,M^*}$ and $P_n \geq 1$,

$$\frac{L_n(k)}{L_n(k_n^*)} \leq \frac{P_n L_{n,P_n}(k)}{L_{n,P_n}(k^*_{n,P_n})} \leq CP_n. \quad \text{(C.11)}$$

According to (B.3)-(B.7), (B.11), (B.12), (C.10) and (C.11), we have for $q > 0$ and all sufficiently large n,

$$(III) \leq CP_n(k^*_{n^\iota,P_{n^\iota}})^{\xi q} \left\{ \sum_{\substack{k=1 \\ k \in B_{n,M^*}}}^{K_{n^\iota}} \frac{P^q_{n^\iota}(k^q + k^{*q}_{n^\iota,P_{n^\iota}})}{N^q_\iota} + \frac{k^q + k^{*q}_{n^\iota,P_{n^\iota}}}{N^{q/2}_\iota} \right.$$
$$\left. + \frac{k^{q/2} + k^{*q/2}_{n^\iota,P_{n^\iota}}}{N^q_\iota L^q_{n^\iota,P_{n^\iota}}(k)} + \frac{\|\mathbf{a}(k) - \mathbf{a}(k^*_{n^\iota,P_{n^\iota}})\|^q_R}{N^{q/2}_\iota L^q_{n^\iota,P_{n^\iota}}(k)} + \frac{K^{2q}_{n^\iota} P^{2q}_{n^\iota}}{N^{2q}_\iota L^q_{n^\iota,P_{n^\iota}}(k)} \right\}. \quad (C.12)$$

In view of (C.7), (C.12) and the conditions imposed on ι and P_n, we have

$$(III) = o(1) \quad (C.13)$$

by using a sufficiently large q in (C.12). Similarly, (5.2) and the definition of B_{n,M^*} yield that for all $k \in \{k : 1 \leq k \leq K_n, k \notin B_{n,M^*}\}$,

$$\frac{L_{n,P_n}(k)}{L_{n,P_n}(k) - L_{n,P_n}(k^*_{n,P_n})} \leq C(k^*_{n,P_n})^\xi \quad (C.14)$$

holds eventually in n. By (B.3)-(B.7), (B.11), (B.12), (C.14), the fact that for $P_n \geq 1$, $L_n(k)/L_n(k^*_n) \leq P_n L_{n,P_n}(k)/L_{n,P_n}(k^*_{n,P_n})$, and an argument similar to the one used to verify (C.13), we have

$$(IV) = o(1). \quad (C.15)$$

Consequently, (C.6) follows from (C.8), (C.9), (C.13), (C.15) and Proposition 2.

Next, assume that condition (i) holds. In this case, Theorem 4 guarantees that

$$\limsup_{n\to\infty} E\left\{ \frac{\left(\mathbf{f}(\hat{k}_{n,P_n}) + \mathcal{S}(\hat{k}_{n,P_n})\right)^2}{\frac{p_0\sigma^2}{N}} \right\} \leq 1. \quad (C.16)$$

By (C.1), (B.3)-(B.7), (B.11), (B.12) and the same reasoning as that of Theorem 4, we also have

$$\lim_{n\to\infty} P(\hat{k}_{n,P_n} \neq p_0) = 0, \quad (C.17)$$

and for $q > 0$,

$$E\left| \frac{\mathbf{f}(\hat{k}_{n,2}) + \mathcal{S}(\hat{k}_{n,2})}{(\frac{p_0\sigma^2}{N})^{1/2}} \right|^q = O(1). \quad (C.18)$$

Since for sufficiently large n,

$$\frac{q_n(\hat{k}_n^{(\iota)}) - \sigma^2}{L_n(k_n^*)} \leq E\left\{\frac{(\mathbf{f}(\hat{k}_{n,2}) + \mathcal{S}(\hat{k}_{n,2}))^2}{\frac{p_0\sigma^2}{N}}\left(I_{\{\hat{k}_{n,P_n} \neq p_0\}} + I_{\{\hat{k}_{n^\iota,P_{n^\iota}} \neq p_0\}}\right)\right\}$$

$$+ \quad E\left\{\frac{(\mathbf{f}(\hat{k}_{n,P_n}) + \mathcal{S}(\hat{k}_{n,P_n}))^2}{\frac{p_0\sigma^2}{N}}\right\}, \qquad \text{(C.19)}$$

(C.6) follows from (C.16)-(C.19) and Hölder's inequality. This completes the proof of the theorem. □

REFERENCES

Akaike, H. (1969). Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.* **21** 243-247.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Automat. Control* **19** 716-723.

Bhansali, R. J. (1986). Asymptotically efficient selection of the order by the criterion autoregressive transfer function. *Ann. Statist.* **14** 315-325.

Findley, D. F. (2005). Asymptotic second moment properties of out-of-sample forecast errors of misspecified regARIMA models and the optimality of GLS. *Statist. Sinica*, forthcoming.

Findley, D. F. and Wei, C. Z. (1993). Moment bounds for deriving time series CLT's and model selection procedures. *Statist. Sinica* **3** 453-470.

Goldenshluger, A. and Zeevi, A. (2001). Nonasymptotic bounds for autoregressive time series modeling. *Ann. Statist.* **29** 417-444.

Hannan, E. J., McDougall, A. J. and Poskitt, D. S. (1989). Recursive estimation of autoregressions. *J. R. Statist. Soc. B* **51** 217-233.

Hannan, E. J. and Quinn, B. C. (1979). The determination of the order of an autoregression. *J. R. Statist. Soc. B* **41** 190-195.

Hasen, M. H. and Yu, B. (2001). Model selection and the principle of minimum description length. *J. Am. Statist. Assoc.* **96** 746-774.

Hemerly, E. M. and Davis, M. H. A. (1989). Strong consistency of the PLS criterion for order determination of autoregressive processes. *Ann. Statist.* **17** 941-946.

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics* **32** 1-49.

Ing, C. K. (2001). A note on mean-squared prediction errors of the least squares predictors in random walk models. *J. Time Ser. Anal.* **22** 711-724.

Ing, C. K. (2003). Multistep prediction in autoregressive processes. *Economet. Theory* **19** 254-279.

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

Ing, C. K. and Wei, C. Z. (2005). Order selection for same-realization prediction in autoregressive processes. *Ann. Statist.*, forthcoming.
Available at http://www.stat.sinica.edu.tw/Ing/T067.pdf

Inoue, A. and Kilian, L. (2005). On the selection of forecasting models. *J. Economet.*, forthcoming.

Karagrigoriou, A. (1997). Asymptotic efficiency of the order selection of a nongaussian AR process. *Statist. Sinica* **7** 407-423.

Kavalieris, L. (1989). The estimation of the order of an autoregression using recursive residuals and cross-validation. *J. Time Ser. Anal.* **10** 271-281.

Kunitomo, N. and Yamamoto, T. (1985). Properties of predictors in misspecified autoregressive time series. *J. Am. Statist. Assoc.* **80** 941-950.

Lai, T. L. and Wei, C. Z. (1982). Least squares estimates in stochastic regression models with application to identification and control systems. *Ann. Statist.* **10** 154-166.

McCracken, M. W. (2000). Robust out-of-sample inference. *J. Economet.* **99** 195-223.

Mallows, C. L. (1973). Some comments on C_p. *Technometrics* **12** 591-612.

Rissanen, J. (1986). Order estimation by accumulated prediction errors. In *Essays in Time Series and Allied Processes* (J. Gani and M. P. Priestley, eds.) *J. Appl. Probab.* **23A** 55-61.

Schorfheide, F. (2005). VAR forecasting under misspecification. *J. Economet.*, forthcoming.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461-464.

Shibata, R. (1976). Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63** 117-126.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 147-164.

Speed, T. P. and Yu, B. (1993). Model selection and prediction : normal regression. *Ann. Inst. Statist. Math.* **45** 35-54.

Wax, M. (1988). Order selection for AR models by predictive least squares," *IEEE Trans. Acoust. Speech Signal Process.* **36** 581-588.

Wei, C. Z. (1987). Adaptive prediction by least squares predictors in stochastic regression models with application to time series. *Ann. Statist.* **15**, 1667-1682.

Wei, C. Z. (1992). On predictive least squares principles. *Ann. Statist.* **20**, 1-42.

West, K. D. (1996). Asymptotic inference about predictive ability. *Econometrica* **64** 1067-1084.

Zheng, X. and Loh, W. Y. (1997). A consistent variable selection criterion for linear models with high-dimensional covariates. *Statist. Sinica* **7** 311-325.