*Econometric Theory*, 2009, Page 1 of 30.
doi:10.1017/S0266466609990107

# PREDICTION ERRORS IN NONSTATIONARY AUTOREGRESSIONS OF INFINITE ORDER

Ching-Kang Ing
*Academia Sinica and National Taiwan University*

Chor-yiu Sin
*National Tsing Hua University and Xiamen University*

Shu-Hui Yu
*National University of Kaohsiung*

Assume that observations are generated from nonstationary autoregressive (AR) processes of infinite order. We adopt a finite-order approximation model to predict future observations and obtain an asymptotic expression for the mean-squared prediction error (MSPE) of the least squares predictor. This expression provides the first exact assessment of the impacts of nonstationarity, model complexity, and model misspecification on the corresponding MSPE. It not only provides a deeper understanding of the least squares predictors in nonstationary time series, but also forms the theoretical foundation for a companion paper by the same authors, which obtains asymptotically efficient order selection in nonstationary AR processes of possibly infinite order.

## 1. INTRODUCTION

One of the most popular models for modeling a stationary time series *nonparametrically* is the autoregressive process of infinite order (AR($\infty$)). However, since there are infinitely many unknown coefficients in the model, statistical inferences are usually based on an approximation AR($k$) model for some $1 \le k < \infty$. Berk (1974) showed that when $k$ goes to infinity with the sample size at a suitable rate, the autoregressive spectral estimates are asymptotically normal and uncorrelated at different fixed frequencies. Shibata (1980) considered the problem of predicting the future of an independent copy of the observed time series (referred to as the *independent*-realization prediction) using a class of candidate AR models.

The authors are deeply grateful to a co-editor and two referees for their helpful suggestions and comments. The research of the first and third authors as partially supported by the National Science Council of Taiwan under grants NSC 94-2118-M-001-013 and NSC 94-2416-H-260-019, respectively. Address correspondence to Ching-Kang Ing, Institute of Statistical Science, Academia Sinica, No. 128, Sec. 2, Academia Rd., Taipei 11529, Taiwan, R.O.C.; e-mail: cking@stat.sinica.edu.tw.

© Cambridge University Press, 2009

He showed that Akaike's information criterion (AIC) (Akaike, 1974) and its variants are asymptotically efficient in choosing model orders for independent-realization predictions. In contrast, Gerencsér (1992) focused on the more natural same-realization prediction (the prediction of the future of the observed time series) and gave an asymptotic expression for the mean-squared prediction error (MSPE) of the ridge regression predictor when the AR order $k$ tends to infinity sufficiently slowly. Under a less stringent assumption on $k$ than that of Gerencsér (1992), Ing and Wei (2003, 2005) obtained an asymptotic expression for the MSPE of the least squares predictor and showed that AIC and its variants are still asymptotically efficient for same-realization predictions.

However, since the above results are restricted to the stationary case, they may preclude many economic time series, which often exhibit nonstationary characteristics. To fill this gap, in this paper, we consider a data set, $y_1, \ldots, y_n$, which is generated from the following AR($\infty$) model:

$$\left(1+\sum_{j=1}^{\infty} a_j B^j\right)(1-B)^d y_t = \varepsilon_t, \tag{1}$$

where $A(z) = 1+\sum_{j=1}^{\infty} a_j z^j \neq 0$ for all $|z| \leq 1$, $B$ is the backshift operator, $0 \leq d < \infty$ is a nonnegative integer, $\{\varepsilon_t, t = 0, \pm 1, \pm 2, \ldots\}$ are independent random disturbances, each with mean 0, and variance $\sigma^2 > 0$ and the initial conditions are given by $y_t = 0$, for $t \leq 0$. For a discussion of other initial conditions, see Section 4. Model (1), including the ARIMA$(p,d,q)$ model as a special case, can accommodate many stationary and nonstationary time series encountered in practice. To predict future observations based on observed data, a class of approximation models, AR(1), ..., AR($K_n$), is considered, where $K_n$ is allowed to tend to infinity as $n$ does. When AR($k$) is adopted, following Shibata (1980) and Ing and Wei (2003, 2005), we estimate the associated AR coefficients using the least squares type estimator, $\hat{\mathbf{a}}_n(k)$, where $\hat{\mathbf{a}}_n(k)$ satisfies $-[\sum_{j=K_n}^{n-1} \mathbf{y}_j(k)\mathbf{y}_j'(k)]\hat{\mathbf{a}}_n(k) = \sum_{j=K_n}^{n-1} \mathbf{y}_j(k) y_{j+1}$ with $\mathbf{y}_j(k) = (y_j, \ldots, y_{j-k+1})'$. Let $y_{n+1}$ be predicted by $\hat{y}_{n+1}(k) = -\mathbf{y}_n'(k)\hat{\mathbf{a}}_n(k)$. In Section 3, we give an asymptotic expression for the MSPE of $\hat{y}_{n+1}(k)$, $\mathrm{E}(y_{n+1} - \hat{y}_{n+1}(k))^2$; see Theorems 2 and 3 for details. This expression provides the first exact evaluation of the impacts of nonstationarity, model complexity, and model misspecification on the corresponding MSPE. It not only gives a nontrivial extension of Ing and Wei's (2005) Theorem 3, but also forms the theoretical foundation for a companion paper by Ing, Sin, and Yu (2007), which shows that the asymptotic efficiency (see (32) in Section 3 of the present paper) of AIC and a two-stage information criterion of Ing (2007) in various stationary time series models carries over to nonstationary cases.

This paper is organized as follows. Section 2 develops some moment bounds for the inverse of the normalized Fisher information matrix, which are key tools for proving Theorems 2 and 3 in Section 3. These moment bounds are also of independent interest since the matrix under consideration is of increasing dimension and is formed by highly correlated data. Main results of this paper are given in Section 3, and concluding remarks are in Section 4. All proofs of the results in Sections 2 and 3 are relegated to Appendixes A and B, respectively.

## 2. MOMENT BOUNDS FOR THE INVERSE OF THE NORMALIZED FISHER INFORMATION MATRIX OF INCREASING DIMENSION

In the sequel, unless otherwise stated, we assume that

$$A(z) = \sum_{j=0}^{\infty} a_j z^j \neq 0 \quad \text{for all } |z| \leq 1 \quad \text{and} \quad \sum_{j=1}^{\infty} | \, j a_j \, | < \infty, \tag{2}$$

where $a_0 = 1$. Note that (2) yields

$$A^{-1}(z) = B(z) = \sum_{j=0}^{\infty} b_j z^j \neq 0 \quad \text{for all } |z| \leq 1 \quad \text{and} \quad \sum_{j=1}^{\infty} |j b_j| < \infty, \tag{3}$$

where $b_0 = 1$. A thorough discussion of condition (2) can be found in Remark 1 of Ing and Wei (2005) and references given therein. Define $z_t = (1-B)^d y_t$. Then, it is not difficult to see that $z_t = \sum_{j=0}^{t-1} b_j \varepsilon_{t-j}$. Let $z_{t,\infty} = \sum_{i=0}^{\infty} b_i \varepsilon_{t-i}$, $\mathbf{z}_t(\upsilon) = (z_t, \ldots, z_{t-\upsilon+1})'$, $\mathbf{z}_{t,\infty}(\upsilon) = (z_{t,\infty}, \ldots, z_{t-\upsilon+1,\infty})'$, and $\mathbf{a}(\upsilon) = (a_1(\upsilon), \ldots, a_\upsilon(\upsilon))' = \arg\min_{\mathbf{c} \in R^\upsilon} \mathrm{E}(z_{t,\infty} + \mathbf{z}'_{t-1,\infty}(\upsilon)\mathbf{c})^2$. To give a more analyzable expression for $y_{n+1} - \hat{y}_{n+1}(k)$, we define $\varepsilon_{j+1,k-d}$, $\imath(k)$ and $G_n(k)$ as follows:

$$\varepsilon_{j+1,k-d} = \begin{cases} z_{j+1}, & k = d, \\ z_{j+1} + \mathbf{a}'(k-d)\mathbf{z}_j(k-d), & k > d; \end{cases}$$

$$\imath(k) = \begin{cases} \left(\mathbf{a}'(k-d), -\mathbf{1}'_d\right)', & k > d \geq 1, \\ -\mathbf{1}_k, & 1 \leq k \leq d, \\ \mathbf{a}(k), & d = 0, \end{cases}$$

with $\mathbf{1}_l$ denoting the $l$-dimensional vector of 1's; and

$$G_n(k) = \begin{cases} \mathrm{diag}(1, \ldots, 1, N^{-d+1/2}, \ldots, N^{-1/2}), & k > d \geq 1, \\ \mathrm{diag}(N^{-d+1/2}, \ldots, N^{-d+k-1/2}), & 1 \leq k \leq d, \\ \mathrm{diag}(1, \ldots, 1), & d = 0, \end{cases}$$

with $N = n - K_n$, where $G_n(k)$ is a $k \times k$ matrix and $\imath(k)$ is a $k$-dimensional vector. In addition, the $k \times k$ matrix $Q(k)$ is implicitly defined by

$$Q(k)\mathbf{y}_j(k) = \begin{cases} \left(\mathbf{z}'_j(k-d), y_j(d), \ldots, y_j(1)\right)', & k > d \geq 1, \\ \left(y_j(d), \ldots, y_j(d-k+1)\right)', & 1 \leq k \leq d, \\ \mathbf{z}_j(k), & d = 0, \end{cases}$$

with $y_j(v) = (1-B)^{d-v} y_j$. Now, for $k \geq \max\{1, d\}$,

$$
\begin{aligned}
& y_{n+1} - \hat{y}_{n+1}(k) \\
&= y_{n+1} - \mathbf{y}_n'(k) \left[ \sum_{j=K_n}^{n-1} \mathbf{y}_j(k)\mathbf{y}_j'(k) \right]^{-1} \sum_{j=K_n}^{n-1} \mathbf{y}_j(k) y_{j+1} \\
&= y_{n+1} \\
&\quad - \mathbf{y}_n'(k) Q'(k) \left[ \sum_{j=K_n}^{n-1} Q(k)\mathbf{y}_j(k)\mathbf{y}_j'(k) Q'(k) \right]^{-1} \\
&\quad \times \sum_{j=K_n}^{n-1} Q(k)\mathbf{y}_j(k)[\varepsilon_{j+1,k-d} - \mathbf{y}_j'(k) Q'(k)\imath(k)] \\
&= \left\{ -N^{-1}\mathbf{s}_{n,n}'(k) \left[ \frac{1}{N} \sum_{j=K_n}^{n-1} \mathbf{s}_{j,n}(k)\mathbf{s}_{j,n}'(k) \right]^{-1} \sum_{j=K_n}^{n-1} \mathbf{s}_{j,n}(k)\varepsilon_{j+1,k-d} \right\} \\
&\quad + \varepsilon_{n+1,k-d},
\end{aligned} \tag{4}
$$

where the existence of the above inverse matrices are discussed in Remark 3 below, $\mathbf{s}_{j,n}(k) = G_n(k)Q(k)\mathbf{y}_j(k)$, and the second and third equalities are ensured by the fact that $\varepsilon_{j+1,k-d} = y_{j+1} + \mathbf{y}_j'(k)Q'(k)\imath(k)$. Let $a_j(l) = 0$ for $j > l \geq 0$. In the rest of this paper, $\mathbf{a}(v), v \geq 0$, will sometimes be viewed as an infinite-dimensional vector with the $i$th component equal to $a_i(v)$, $i = 1, 2, \ldots$. Define $\|\mathbf{d}\|_z^2 = \sum_{1 \leq i,j \leq \infty} d_i d_j \gamma_{i-j}$, where $\gamma_{i-j} = \mathrm{E}(z_{i,\infty} z_{j,\infty})$ and $\mathbf{d} = (d_1, d_2, \ldots)'$ is an infinite-dimensional vector satisfying $\|\mathbf{d}\|^2 = \sum_{j=1}^{\infty} d_j^2 < \infty$. Then, by observing $z_{t+1,\infty} + \sum_{j=1}^{\infty} a_j z_{t+1-j,\infty} = \varepsilon_{t+1}$, one has, for $l \geq 0$,

$$
\begin{aligned}
\|\mathbf{a} - \mathbf{a}(l)\|_z^2 &= \mathrm{E}\left( \sum_{j=1}^{\infty} (a_j - a_j(l)) z_{t+1-j,\infty} \right)^2 \\
&= \mathrm{E}\left( z_{t+1,\infty} + \sum_{j=1}^{l} a_j(l) z_{t+1-j,\infty} \right)^2 - \sigma^2.
\end{aligned} \tag{5}
$$

As will be seen in Section 3, $\|\mathbf{a} - \mathbf{a}(k-d)\|_z^2$ is one of the key components in our asymptotic expression for the MSPE of $\hat{y}_{n+1}(k)$.

In view of (4), moment properties of the inverse of the normalized Fisher information matrix,

$$
\hat{S}_n(k) = \frac{1}{N} \sum_{j=K_n}^{n-1} \mathbf{s}_{j,n}(k)\mathbf{s}_{j,n}'(k), \tag{6}
$$

play crucial roles in investigating the MSPE of $\hat{y}_{n+1}(k)$. In particular, it is interesting to ask whether $\hat{S}_n(k)$ is nonsingular in the $q$th moment, $q > 0$, in the sense that

$$\max_{1 \le k \le K_n} \mathrm{E}\{\lambda_{\min}^{-q}(\hat{S}_n(k))\} = O(1), \tag{7}$$

where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of matrix A. For the special case where $K_n = k > d \ge 1$, $a_i = 0$ for all $i \ge k+1$, and $k$ is a constant fixed with $n$, Theorem 3.5.1 of Chan and Wei (1988) and the continuous mapping theorem can be applied to show that

$$\lambda_{\min}^{-1}(\hat{S}_n(k)) \Rightarrow \sigma^{-2}\lambda_{\min}^{-1}\begin{pmatrix} \Gamma(k-d) & \mathbf{0}_{(k-d)\times d} \\ \mathbf{0}_{d\times(k-d)} & F \end{pmatrix} \tag{8}$$

and

$$\lambda_{\min}^{-1}\begin{pmatrix} \Gamma(k-d) & \mathbf{0}_{d\times(k-d)} \\ \mathbf{0}_{(k-d)\times d} & F \end{pmatrix} < \infty \quad \text{a.s.}, \tag{9}$$

where $\Rightarrow$ denotes convergence in distribution, $\Gamma(l) = \mathrm{E}\left(\mathbf{z}_{t,\infty}(l)\mathbf{z}'_{t,\infty}(l)\right), l \ge 1$, $\mathbf{0}_{s\times t}$ denotes an $s \times t$ matrix of 0's, and with $F_0(t) = W(t)$ representing the standard Brownian motion and $F_j(t) = \int_0^t F_{j-1}(s)\,ds$,

$$F = \left[\varsigma_{i,j}\right]_{i,j=1,\ldots,d} = \left[\int_0^1 F_{d-i}(t)F_{d-j}(t)\,dt\right]_{i,j=1,\ldots,d}.$$

While (8) and (9) suggest that (7) is a valid goal to pursue, some extra complexities are worth mentioning. First, since the dimension of $\hat{S}_n(k)$ in (7) is allowed to increase to infinity with $n$ and the larger the dimension is, the smaller the corresponding minimum eigenvalue is, a certain limitation on the rate of divergence of $K_n$ is required in order to prevent the matrix from being ill-conditioned. Second, since convergence in distribution does not necessarily imply convergence in mean, (8) cannot guarantee

$$\lim_{n\to\infty} \mathrm{E}\lambda_{\min}^{-1}(\hat{S}_n(k)) = \sigma^{-2}\mathrm{E}\left\{\lambda_{\min}^{-1}\begin{pmatrix} \Gamma(k-d) & \mathbf{0}_{(k-d)\times d} \\ \mathbf{0}_{d\times(k-d)} & F \end{pmatrix}\right\}.$$

In fact, when the distribution of $\varepsilon_t$ has a mass at zero, it is easy to construct a counterexample showing that (7) does not hold for any $q > 0$, even in stationary and fixed-dimensional cases. Therefore, some smoothness conditions on the distribution of $\varepsilon_t$ are needed. In response to these requirements, we impose the following assumptions:

**Maximal Order (MO).** $K_n \to \infty$ with

(i) $K_n^{\max\{4d-1,2\}} = o(n)$,
(ii) $K_n^{\max\{4d-1,2+\delta_1\}} = o(n)$ for some $\delta_1 > 0$,
(iii) $K_n^{4\max\{1,d\}+2+\delta_1} = o(n)$ for some $\delta_1 > 0$, or
(iv) $K_n^{4\max\{1,d\}+3+\delta_1} = o(n)$ for some $\delta_1 > 0$.

**Nonsingularity (NS).** Let $F_{t,m,\mathbf{v}_m}(.)$ be the distribution function of $\mathbf{v}_m'(\varepsilon_t, \ldots, \varepsilon_{t+1-m})'$, where $\mathbf{v}_m = (v_1, \ldots, v_m)' \in R^m$. There exist positive numbers $\alpha$, $\delta$, and $M$ such that, for all $m \geq 1$, $m \leq t < \infty$, and $\|\mathbf{v}_m\|^2 = \sum_{j=1}^m v_j^2 = 1$,

$$| F_{t,m,\mathbf{v}_m}(x) - F_{t,m,\mathbf{v}_m}(y) | \leq M \, | x - y |^{\alpha}, \quad \text{as } | x - y | \leq \delta.$$

**Remark 1.** MO imposes some limitations on $K_n$, reflecting the fact that a larger order of integratedness yields larger correlations among the observations, and hence (possibly) a smaller minimum eigenvalue of $\hat{S}_n(k)$. There is also a tradeoff between the rate of divergence of $K_n$ and the moment restriction on $\varepsilon_t$. As will be shown in Theorem 1, (7) can be achieved under MO(ii) accompanying a stronger moment condition on $\varepsilon_t$ or MO(iii) accompanying a weaker one. MO(iv) is slightly stronger than MO(iii) and will be used in the next section.

**Remark 2.** Note that NS is fulfilled by most continuous-type distributions. For instance, when $\varepsilon_t$'s are normally distributed, NS is satisfied with $M = (2\pi\sigma^2)^{-1/2}$, $\alpha = 1$, and any $\delta > 0$. In addition, when $\varepsilon_t$'s are i.i.d. with an integrable characteristic function, NS is satisfied with any $\delta > 0$, $\alpha = 1$, and some $M > 0$. For more details, see Lemma 4 of Ing and Sin (2006). For some other similar distributional assumptions used to establish negative moment bounds for the minimum eigenvalue of the Fisher information matrix in stationary time series; see Findley and Wei (2002), Ing and Wei (2003), and references given therein.

Lemma 1 provides an upper bound for $\mathrm{E}\{\lambda_{\min}^{-q}(\hat{S}_n(k_n))\}$ under model (1), where $q > 0$ and $1 \leq k_n \leq K_n$. The proof of Lemma 1 is inspired by Lemma 1 of Ing and Wei (2003). However, as shown in Appendix A, a much more delicate analysis is required to deal with the extra difficulty introduced by nonstationarity.

LEMMA 1. *Assume (1), (2), and NS. Then, for* $1 \leq k_n \leq K_n$ *with* $K_n$ *satisfying MO(i), any* $q > 0$*, and any* $0 < \theta < 1$,

$$\mathrm{E}\left\{\lambda_{min}^{-q}\left(\hat{S}_n(k_n)\right)\right\} = O(k_n^{(2\max\{1,d\}+\theta)q}). \tag{10}$$

**Remark 3.** Equality (10) guarantees that the inverse of $\hat{S}_n(k_n)$ almost surely (a.s.) exists for all large $n$. Therefore, we can define $\hat{S}_n^{-1}(k_n)$ as any generalized inverse of $\hat{S}_n(k_n)$ without causing ambiguity asymptotically. This enables us to rewrite (10) as

$$\mathrm{E}\left\|\hat{S}_n^{-1}(k_n)\right\|^q = O\left(k_n^{(2\max\{1,d\}+\theta)q}\right), \tag{11}$$

where for a matrix $A$, $\|A\|^2 = \lambda_{\max}(A'A) = \sup_{\|\mathbf{v}\|=1} \mathbf{v}'A'A\mathbf{v}$. Since

$$\hat{S}_n(k) = \frac{1}{N} G_n(k) Q(k) \left( \sum_{j=K_n}^{n-1} \mathbf{y}_j(k)\mathbf{y}_j'(k) \right) Q'(k) G_n'(k),$$

and $G_n(k)$ and $Q(k)$ are nonsingular, (10) implies that the inverses of $\sum_{j=K_n}^{n-1} \mathbf{y}_j(k)$ $\mathbf{y}_j^{'}(k)$ and $\sum_{j=K_n}^{n-1} Q(k)\mathbf{y}_j(k)\mathbf{y}_j^{'}(k)Q'(k)$ a.s. exist for all large $n$, and hence $(\sum_{j=K_n}^{n-1}$ $\mathbf{y}_j(k)\mathbf{y}_j^{'}(k))^{-1}$ and $(\sum_{j=K_n}^{n-1} Q(k)\mathbf{y}_j(k)\mathbf{y}_j^{'}(k)Q'(k))^{-1}$ in (4) can be similarly defined.

The following example illustrates the usefulness of (10) (or (11)) in situations where $k_n$ is bounded by a finite constant.

**Example 1**

Consider an ARI$(p,d)$ model,

$$(1+a_1 B+\cdots+a_p B^p)(1-B)^d y_t=\varepsilon_t,$$

where $p \geq 0$, $d \geq 1$, and $\varepsilon_t$'s satisfy the assumptions imposed by (1). Assume further that $\sup_{0<t<\infty} \mathrm{E}|\varepsilon_t|^{q_1} < \infty, q_1 \geq 2$, and NS holds. Define $p^* = p+d$ and let $K_n = p^*$. Then, an argument similar to that used in (4) yields

$$\begin{aligned}
&\|\sqrt{N}\{Q'(p^*)G_n'(p^*)\}^{-1}(\hat{\mathbf{a}}_n(p^*)-\bar{\mathbf{a}}(p^*))\| \\
&\quad = \left\|\hat{S}_n^{-1}(p^*)N^{-1/2}\sum_{j=K_n}^{n-1}\mathbf{s}_{j,n}(p^*)\varepsilon_{j+1}\right\| \\
&\quad \leq \|\hat{S}_n^{-1}(p^*)\| \left\|N^{-1/2}\sum_{j=K_n}^{n-1}\mathbf{s}_{j,n}(p^*)\varepsilon_{j+1}\right\|, \qquad \textbf{(12)}
\end{aligned}$$

where

$$\bar{\mathbf{a}}(p^*) = (\bar{a}_1(p^*),\ldots,\bar{a}_{p^*}(p^*))' = Q'(p^*)\imath(p^*). \qquad \textbf{(13)}$$

(Note that $1+\bar{a}_1(p^*)B+\cdots+\bar{a}_{p^*}(p^*)B^{p^*} = (1+a_1B+\cdots+a_pB^p)(1-B)^d$.) By (12), Lemma 1, Hölder's inequality, and Lemmas B.1 and B.3 (see Appendix B), one has, for any $0<q<q_1$,

$$\begin{aligned}
&\mathrm{E}\|\sqrt{N}\{Q'(p^*)G_n'(p^*)\}^{-1}(\hat{\mathbf{a}}_n(p^*)-\bar{\mathbf{a}}(p^*))\|^q \\
&\quad \leq (\mathrm{E}\|\hat{S}_n^{-1}(p^*)\|^{qq_1/(q_1-q)})^{(q_1-q)/q_1}\left(\mathrm{E}\left\|N^{-1/2}\sum_{j=K_n}^{n-1}\mathbf{s}_{j,n}(p^*)\varepsilon_{j+1}\right\|^{q_1}\right)^{q/q_1} \\
&\quad = O(1). \qquad \textbf{(14)}
\end{aligned}$$

It is worth mentioning that while the limiting distribution of $\sqrt{N}\{Q'(p^*)$ $G_n'(p^*)\}^{-1}(\hat{\mathbf{a}}_n(p^*)-\bar{\mathbf{a}}(p^*))$ has been extensively studied in the literature, (14) seems to be the first result that reports its moment properties.

Although the moment bound provided by (10) (or (11)) tends to infinity as $k_n$ does, it serves as a vehicle for pursuing sharper results (e.g., (7)) at the price of imposing stronger moment conditions on $\varepsilon_t$, as shown in Theorem 1. To state the result, define

$$\hat{S}_{d,n}(k) = \begin{cases} \hat{S}_n(k), & 1 \le k \le d, \\ \begin{pmatrix} \Gamma(k-d) & \mathbf{0}_{(k-d)\times d} \\ \mathbf{0}_{d\times(k-d)} & \hat{S}_n(d) \end{pmatrix}, & k > d \ge 1, \\ \Gamma(k), & d = 0. \end{cases}$$

THEOREM 1.

*(i) Assume (1), (2), NS, and, for some* $q_1 \ge 2$,

$$\sup_{0<t<\infty} \mathrm{E}(|\varepsilon_t|^{2q_1}) < \infty. \tag{15}$$

*Then, for* $K_n$ *that satisfies MO(iii) and any* $0 < q < q_1$,

$$\max_{1\le k\le K_n} \mathrm{E}\|\hat{S}_n^{-1}(k) - \hat{S}_{d,n}^{-1}(k)\|^q = o(1), \tag{16}$$

$$\max_{1\le k\le K_n} \mathrm{E}\|\hat{S}_n^{-1}(k)\|^q = O(1), \tag{17}$$

*and*

$$\max_{1\le k\le K_n} \frac{\mathrm{E}\|\hat{S}_n^{-1}(k) - \hat{S}_{d,n}^{-1}(k)\|^{q/2}}{\left(\frac{k^2}{N}\right)^{q/4}} = O(1). \tag{18}$$

*(ii) Assume the same assumptions as in (i), but with MO(iii) weakened to MO(ii) and (15) strengthened to*

$$\sup_{0<t<\infty} \mathrm{E}(|\varepsilon_t|^s) < \infty, \qquad s = 1, 2, \ldots. \tag{19}$$

*Then, (17) and (18) hold for any* $q > 0$. ■

Before leaving this section, we note that Theorem 1 and Lemma 1 play important roles in decomposing the prediction error due to estimation uncertainty into one (asymptotically) stationary part and one nonstationary part; see Section 3. These results are in line with those developed in Chan and Wei (1988), in which limiting distributions of the least squares estimator were considered. On the other hand, it is worth mentioning that the (normalized) regressors and the (normalized) estimators used for prediction are not asymptotically independent in nonstationary autoregressions; see Ing and Sin (2006) for simple random walk models. Therefore, their joint effects need to be considered. In Section 3, a novel approach is taken to alleviate this difficulty.

## 3. ASYMPTOTIC EXPRESSIONS FOR THE MSPE

In this section, we give an asymptotic expression for the MSPE of $\hat{y}_{n+1}(k)$ with $\max\{1,d\} \le k \le K_n$. In view of (4), for $k \ge \max\{1,d\}$,

$$\mathrm{E}(y_{n+1}-\hat{y}_{n+1}(k))^2 = \sigma^2 + \mathrm{E}(\mathbf{f}_n(k)+\mathcal{S}_n(k-d))^2, \tag{20}$$

where

$$\mathbf{f}_n(k) = \frac{1}{\sqrt{N}}\mathbf{s}'_{n,n}(k)\hat{S}_n^{-1}(k)\left(\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1}\mathbf{s}_{j,n}(k)\varepsilon_{j+1,k-d}\right) \tag{21}$$

and

$$\mathcal{S}_n(k-d) = -(\varepsilon_{n+1,k-d}-\varepsilon_{n+1}) = \sum_{i=1}^{n}(a_i - a_i(k-d))z_{n+1-i}. \tag{22}$$

For $k \ge \max\{1,d\}$, it can be shown that

$$\begin{aligned}\mathbf{f}_n(k) &\approx \left\{\frac{U'_{n,n}(d)}{\sqrt{N}}\hat{S}_n^{-1}(d)\left(\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1}U_{j,n}(d)\varepsilon_{j+1,k-d}\right)\right\}I_{\{d\ge 1\}}\\ &\quad+\left\{\frac{\mathbf{z}'_n(k-d)}{\sqrt{N}}\Gamma^{-1}(k-d)\left(\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1}\mathbf{z}_j(k-d)\varepsilon_{j+1,k-d}\right)\right\}I_{\{k>d\}}\\ &\equiv B_{1n}(k,d)+B_{2n}(k-d),\end{aligned} \tag{23}$$

where $U_{j,n}(v) = (y_j(d)/N^{d-(1/2)},\ldots,y_j(d-v+1)/N^{d-v+(1/2)})'$, $I_{\{\cdot\}}$ denotes the indicator function, and the meaning of "≈" is clarified in (B.42). Note that $B_{1n}(k,d)$ and $B_{2n}(k-d)$ can be further approximated by

$$\mathbf{f}_{1,n}(d) = \frac{U'_{n,n}(d)}{\sqrt{N}}\hat{S}_n^{-1}(d)\left(\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1}U_{j,n}(d)\varepsilon_{j+1}\right)I_{\{d\ge 1\}} \tag{24}$$

and

$$\mathbf{f}_{2,n}(k-d) = \frac{\mathbf{z}'_n(k-d)}{\sqrt{N}}\Gamma^{-1}(k-d)\left(\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1}\mathbf{z}_j(k-d)\varepsilon_{j+1}\right)I_{\{k>d\}}, \tag{25}$$

respectively; see (B.44) and (B.46) for details. When $d \ge 1$ and $\sup_{0<t<\infty}\mathrm{E}|\varepsilon_t|^q < \infty, q > 2$, following the arguments used in Phillips (1987) and Chan and Wei (1988), it can be shown that

$$\sqrt{n-d}\,\bar{\mathbf{f}}_{1,n}(d) \Rightarrow \sigma\iota' F^{-1}\zeta, \tag{26}$$

where $\bar{\mathbf{f}}_{1,n}(d)$ is $\mathbf{f}_{1,n}(d)$ with $U_{j,n}(d)$ replaced with $\bar{U}_{j,n}(d) = (y_j(d)/(n-d)^{d-(1/2)},\ldots,y_j(1)/(n-d)^{1/2})'$, $N$ replaced with $n-d$, and $K_n$ replaced with $d$, $\zeta = (\int_0^1 F_{d-1}(t)\,dW(t),\ldots,\int_0^1 F_0(t)\,dW(t))'$, and $\iota = (F_{d-1}(1),\ldots,F_0(1))'$. However, the limiting value of $\mathrm{E}(N\mathbf{f}_{1,n}^2(d))$ remains unclear. As will be clarified later, this limiting value, measuring the contribution of nonstationarity to the MSPE, is one of the key elements in our asymptotic expression. The following lemma provides a solution to this problem.

LEMMA 2. *Assume (1) with* $d \geq 1$*, (2), NS,* $K_n = o(n^{1/2})$*, and* $\sup_{0<t<\infty} \mathrm{E}(|\varepsilon_t|^q) < \infty, q > 4$*. Then,*

$$\lim_{n\to\infty} \mathrm{E}(N\mathbf{f}_{1,n}^2(d)) = d(d+1)\sigma^2. \tag{27}$$

■

Armed with Lemmas 1 and 2 and Theorem 1, the main results of this paper are given in Theorems 2 and 3 below.

THEOREM 2. *Assume (1), (2), NS, MO(iv), and*

$$\sup_{0<t<\infty} \mathrm{E}|\varepsilon_t|^{10+\delta_1} < \infty, \tag{28}$$

*where* $\delta_1$ *is defined in MO(iv). Then,*

$$\lim_{n\to\infty} \max_{\max\{1,d\}\leq k\leq K_n} \left| \frac{\mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(k)\}^2 - \sigma^2}{L_n^d(k)} - 1 \right| = 0, \tag{29}$$

*where*

$$L_n^d(k) = \frac{k+d^2}{N}\sigma^2 + \|\mathbf{a} - \mathbf{a}(k-d)\|_z^2, \tag{30}$$

*with* $\|\mathbf{a} - \mathbf{a}(k-d)\|_z^2$ *defined in (5).* ■

**Remark 4.** It is clear from (27), (29), and (30) that when $k \geq \max\{1, d\}$, the MSPE of $\hat{y}_{n+1}(k)$ (after $\sigma^2$ is subtracted) can be *uniformly* and *asymptotically* decomposed into three terms. The first term, $(k-d)\sigma^2/N$, arising from estimating the stationary component in (1), is mainly contributed by $\mathbf{f}_{2,n}(k-d)$; the second term, $d(d+1)\sigma^2/N$, arising from estimating the nonstationary component in (1), is mainly contributed by $\mathbf{f}_{1,n}(d)$; whereas the last term, $\|\mathbf{a} - \mathbf{a}(k-d)\|_z^2$, due to model misspecification, is the contribution of $\mathcal{S}_n(k-d)$. In fact, it is shown in Lemma B.7 that $\mathbf{f}_{1,n}(d)$, $\mathbf{f}_{2,n}(k-d)$, and $\mathcal{S}_n(k-d)$ are asymptotically pairwise uncorrelated. To see this, note that $\mathcal{S}_n(k-d)$ is short-memory and can be approximated by $\mathcal{S}_n^*(k-d)$, which depends only on the latest $\sqrt{n}$ random noises $\{\varepsilon_n, \ldots, \varepsilon_{n-\sqrt{n}+1}\}$. In contrast to $\mathcal{S}_n(k-d)$, $\mathbf{f}_{1,n}(d)$ can be approximated by $\mathbf{f}_{1,n}^*(d)$, which is completely determined by $\{\varepsilon_1, \ldots, \varepsilon_{n-\sqrt{n}}\}$. On the other hand, $\mathbf{f}_{2,n}(k-d)$ can be approximated by $\mathbf{f}_{2,n}^*(k-d)$, which is the inner product of two random vectors, one of which is a function of $\{\varepsilon_n, \ldots, \varepsilon_{n-\sqrt{n}+1}\}$ and the other is a function of $\{\varepsilon_1, \ldots, \varepsilon_{n-\sqrt{n}}\}$. For the precise definitions of $\mathbf{f}_{1,n}^*(d)$, $\mathbf{f}_{2,n}^*(k-d)$, and $\mathcal{S}_n^*(k-d)$, see the proof of Lemma B.7. When $d = 1$, $a_i = 0$ for all $i \geq p-1 \geq 1$, and the model is correctly specified, the argument used in Fuller and Hasza's (1981) Theorem 3.1 can be applied to show that

$$y_{n+1} - \hat{y}_{n+1}(p) - \varepsilon_{n+1} = O_p(n^{-1/2}). \tag{31}$$

Although (31) guarantees that the prediction error due to estimation uncertainty will vanish as $n$ tends to infinity, it cannot distinguish between different error sources from the stationary and nonstationary components of the model.

**Remark 5.** When $1 \leq k < d$, (29) does not hold. For some asymptotic analyses of $\hat{y}_{n+1}(k)$ in this case, see Ing et al. (2007).

**Remark 6.** When $d = 0$, (29) is the same as (3.9) of Ing and Wei (2003), and hence Theorem 2 can be viewed as an extension of Ing and Wei's (2003) Theorem 3. On the other hand, (29) and Remark 4 indicate that stationary and nonstationary components have substantially different "marginal contributions" to the MSPE. Therefore, it is indeed difficult to foresee (29) through Ing and Wei's (2003) result.

**Remark 7.** Although (28) requires the boundedness of the $10+\delta_1$ moment of $\varepsilon_t$, it does not seem very stringent compared to the moment conditions used in the related literature. For example, to give an asymptotic expression for the MSPE of the least squares predictor in situations where $d = 0$ and the order of the predictor is fixed with $n$, Fuller and Hasza (1981) assume that $\varepsilon_t$'s are independently and identically distributed normal random variables, Kunitomo and Yamamoto (1985) require that $\varepsilon_t$'s are independently, identically, and symmetrically distributed around zero with $\mathrm{E}|\varepsilon_1|^{32} < \infty$, and Ing (2003) assumes that $\varepsilon_t$'s are independently distributed with $\sup_{-\infty<t<\infty} \mathrm{E}|\varepsilon_t|^q < \infty$ for some $q > 8$.

Theorem 3 shows that when the moment condition (28) is strengthened to (19), the asymptotic expression (29) is valid for more candidate predictors, as characterized by the assumption on the maximal order.

THEOREM 3. *Assume (1), (2), NS, MO(ii), and (19). Then (29) follows.* ∎

Define the relative prediction efficiency of $\hat{y}_{n+1}(k_1)$ to $\hat{y}_{n+1}(k_2)$ by

$$\frac{\mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(k_2)\}^2 - \sigma^2}{\mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(k_1)\}^2 - \sigma^2}.$$

Let $\hat{k}_n$ be the order selected by an order selection criterion. This criterion is said to be asymptotically efficient if $\hat{k}_n$ satisfies

$$\limsup_{n\to\infty} \frac{\mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(\hat{k}_n)\}^2 - \sigma^2}{\min_{\min\{d,1\}\leq k\leq K_n} \mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(k)\}^2 - \sigma^2} \leq 1, \tag{32}$$

which means that the relative prediction efficiency of the best predictor among $\{\hat{y}_{n+1}(\max\{1,d\}), \ldots, \hat{y}_{n+1}(K_n)\}$ to $\hat{y}_{n+1}(\hat{k}_n)$ will ultimately not exceed 1. Note that (32) was first proposed by Ing and Wei (2005) for the case of $d = 0$. Since Theorem 2 (or Theorem 3) yields

$$\lim_{n\to\infty} \frac{\min_{\min\{d,1\}\leq k\leq K_n} \mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(k)\}^2 - \sigma^2}{L_n^d(k_n^*(d))} = 1,$$

where $k_n^*(d) = \arg\min_{\min\{d,1\} \le k \le K_n} L_n^d(k)$, we can rewrite (32) as

$$\limsup_{n\to\infty} \frac{\mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(\hat{k}_n)\}^2 - \sigma^2}{L_n^d(k_n^*(d))} \le 1. \tag{33}$$

With the help of (33), we are able to make the first step toward the asymptotic efficiency.

**Example 2**

Assume that (1) holds and the $a_i$'s in (1) satisfy, for some $0 < C_1 \le C_2 < \infty$ and $\beta > 0$,

$$C_1 e^{-\beta l} \le \|\mathbf{a} - \mathbf{a}(l)\|_z^2 \le C_2 e^{-\beta l}. \tag{34}$$

Since (34), equivalent to $C_1' e^{-\beta l} \le \sum_{i>l} a_i^2 \le C_2' e^{-\beta l}$ for some $0 < C_1' \le C_2' < \infty$, is fulfilled by any causal and invertible ARMA$(p,q)$ model, with $q > 0$, the model considered in this example includes the ARIMA$(p,d,q)$ model, with $q > 0$, as a special case. By algebraic manipulations, it can be shown that for some $C_3 > 0$,

$$\frac{1}{\beta}\log n - C_3 \le k_n^*(d) \le \frac{1}{\beta}\log n + C_3. \tag{35}$$

Therefore, the divergence rate of the optimal prediction order is $\log n$. However, asymptotic efficiency cannot be attained if a wrong constant is chosen. According to Theorem 2 (or Theorem 3), (35), and a straightforward calculation,

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(\beta_1^{-1}\log n)\}^2 - \sigma^2}{L_n^d(k_n^*(d))} \\
&= \lim_{n\to\infty} \frac{\mathrm{E}\{y_{n+1} - \hat{y}_{n+1}(\beta_1^{-1}\log n)\}^2 - \sigma^2}{L_n^d(\beta_1^{-1}\log n)} \lim_{n\to\infty} \frac{L_n^d(\beta_1^{-1}\log n)}{L_n^d(k_n^*(d))} \\
&= \lim_{n\to\infty} \frac{L_n^d(\beta_1^{-1}\log n)}{L_n^d(k_n^*(d))} = \begin{cases} \frac{\beta}{\beta_1}, & \text{if } 0 < \beta_1 < \beta, \\ \infty, & \text{if } \beta_1 > \beta. \end{cases}
\end{aligned} \tag{36}$$

Equations (35) and (36) point out the difficulty in achieving asymptotic efficiency: It involves the search not only for the best rate, but also for the best constant $\beta^{-1}$, which is usually unknown in practice. In fact, when the AR coefficients decay algebraically, even the best rate may involve unknown parameters, and hence is unknown; see Ing et al. (2007).

## 4. CONCLUDING REMARKS

In analyzing the MSPEs of the least squares predictors of high-dimensional and nonstationary autoregressions, there are two fundamental difficulties. One is that the moment properties of the least squares estimators are difficult to explore because the associated high-dimensional Fisher information matrix involves highly correlated data. Another is, as pointed out by Ing and Sin (2006), the (normalized) regressors and the (normalized) estimators are not asymptotically independent. Hence, unlike the stationary case, their joint effects need to be considered. In Section 2 of this paper, we establish the moment bounds of the inverse of the normalized Fisher information matrix of increasing dimension. To tackle the second difficulty, Section 3 of this paper adopts an indirect approach, which is elaborately developed. In sum, the results in Ing and Wei (2003) are extended from stationary cases to nonstationary cases. An asymptotic expression for the MSPE of the least squares predictors, which can be decomposed into three parts (a stationary part, a nonstationary part, and a model-misspecification part) is obtained at the end of Section 3. The contribution of this paper is two-fold: (1) It provides a deeper understanding of the least squares predictors in nonstationary time series; and (2) it forms the basis for establishing asymptotically efficient order selection in nonstationary AR($\infty$) processes, as detailed in Ing et al. (2007).

Before leaving this section, we remark that when the initial conditions, $y_t = 0$ for all $t \leq 0$, are replaced by

$$\sup_{-\infty < t \leq 0} \mathrm{E}|y_t|^{\nu} < \infty, \quad \text{for some sufficiently large } \nu, \tag{37}$$

and that

$$\{y_t, t \leq 0\} \quad \text{are independent of} \quad \{\varepsilon_t, t \geq 1\}, \tag{38}$$

all theorems and lemmas in the previous sections still hold. It is also possible to extend the analysis in this paper to AR($\infty$) models with deterministic terms or with unit roots located at other frequencies different from zero. However, filling in the details for these extensions is beyond the scope of this paper.

## *REFERENCES*


Akaike, H. (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.

Barria, J. & P.R. Halmos (1982) Asymptotic toeplitz operators. *Transactions of the American Mathematical Society* 273, 621–630.

Berk, K.N. (1974) Consistent autoregressive spectral estimates. *Annals of Statistics* 2, 489–502.

Billingsley, P. (1968) *Convergence of Probability Measure*. Wiley.

Chan, N.H. & C.Z. Wei (1988) Limiting distribution of least squares estimates of unstable autoregressive processes. *Annals of Statistics* 19, 367–401.

Choi, M.-D. (1983) Tricks or treats with the Hilbert matrix. *American Mathematical Monthly* 90, 301–312.

Chow, Y.S. & H. Teicher (1997) *Probability Theory: Independence, Interchangeability, Martingale*, 3rd ed. Springer-Verlag.

Findley, D.F. & C.Z. Wei (2002) AIC, overfitting principles, and the boundedness of moments of inverse matrices for vector autoregressions and related models. *Journal of Multivariate Analysis* 83, 415–450.

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

Gerencsér, L. (1992) AR ($\infty$) estimation and nonparametric stochastic complexity. *IEEE Transactions on Information Theory* 38, 1768–1778.

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

Ing, C.-K. (2007) Accumulated prediction errors, information criteria and optimal forecasting for autoregressive time series. *Annals of Statistics* 35, 1238–1277.

Ing, C.-K. & C.-y. Sin (2006) On prediction errors in regression models with nonstationary regressors. In H.C. Ho, C.-K. Ing, and T.L. Lai (eds.), *Time Series and Related Topics: In Memory of Ching-Zong Wei*, pp. 60–71. IMS Lecture Notes–Monograph Series 52. Institute of Mathematical Statistics.

Ing, C.-K., C.-y. Sin, & S.-H. Yu (2007) Using Information Criteria to Select an Autoregressive Model: A Unified Approach without Knowing the Order of Integratedness. Technical report, Academia Sinica.

Ing, C.-K. & C.Z. Wei (2003) On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis* 85, 130–155.

Ing, C.-K. & C.Z. Wei (2005) Order selection for same-realization predictions in autoregressive processes. *Annals of Statistics* 33, 2423–2474.

Kunitomo, N. & T. Yamamoto (1985) Properties of predictors in misspecified autoregressive time series. *Journal of the American Statistical Association* 80, 941–950.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.

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

Wei, C.Z. (1987) Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Annals of Statistics* 15, 1667–1682.

# APPENDIX A

**Proof of Lemma 1.** We only prove the case of $d \geq 1$ since the case of $d = 0$ can be shown by an argument similar to that used in the proof of Lemma 1 in Ing and Wei (2003). Without loss of generality, we may assume $k_n \geq d+1$ since, for $1 \leq k_{1,n} \leq K_n$ and $k_{2,n} = \max\{d+1, k_{1,n}\}$,

$$\lambda_{\min}^{-q}\left(\frac{1}{N}\sum_{j=K_n}^{n-1} \mathbf{s}_{j,n}(k_{1,n})\mathbf{s}'_{j,n}(k_{1,n})\right) \leq \lambda_{\min}^{-q}\left(\frac{1}{N}\sum_{j=K_n}^{n-1} \mathbf{s}_{j,n}(k_{2,n})\mathbf{s}'_{j,n}(k_{2,n})\right),$$

and $k_{2,n}^{(2+\theta)q} = O\{(k_{1,n})^{(2+\theta)q}\}$. First note that

$$\mathbf{s}_{j,n}(k_n) = \left(\sum_{s=0}^{j-1} b_s\varepsilon_{j-s}, \ldots, \sum_{s=0}^{j-k_n+d} b_s\varepsilon_{j-k_n+d+1-s}, \frac{1}{N^{d-(1/2)}}\sum_{s=0}^{j-1}\kappa_s(d)\varepsilon_{j-s}, \ldots, \right.$$
$$\left. \frac{1}{N^{1/2}}\sum_{s=0}^{j-1}\kappa_s(1)\varepsilon_{j-s}\right)',$$

where $\kappa_t(1) = \sum_{s=0}^{t} b_s$ and for $l_1 \geq 2$, $\kappa_t(l_1) = \sum_{s=0}^{t}\kappa_s(l_1 - 1)$. For notational simplicity, write $\mathbf{s}_j = \mathbf{s}_{j,n}(k_n)$. Let $q > 0$, $0 < \theta < 1$, and $1/2 < \delta_2 < 1$ be arbitrarily chosen. Define $g_n = \lfloor (n - \lfloor \delta_2 n \rfloor)/(\nu k_n) \rfloor$, where $\lfloor a \rfloor$ denotes the largest integer $\leq a$ and $\nu$, a positive integer depending only on $d$, $q$, $\theta$ and $\alpha$, will be specified later. Since MO(i) is imposed, we may assume without loss of generality that $\lfloor n\delta_2 \rfloor > K_n$ and $g_n > K_n$. For $j \geq \lfloor n\delta_2 \rfloor$, define a truncated version of $\mathbf{s}_j$,

$$M_j = \left( \sum_{s=0}^{g_n-1} b_s \varepsilon_{j-s}, \ldots, \sum_{s=0}^{g_n-k_n+d} b_s \varepsilon_{j-k_n+d+1-s}, \frac{1}{N^{d-(1/2)}} \sum_{s=0}^{g_n-1} \kappa_s(d) \varepsilon_{j-s}, \ldots, \right.$$
$$\left. \frac{1}{N^{1/2}} \sum_{s=0}^{g_n-1} \kappa_s(1) \varepsilon_{j-s} \right)',$$

and let $R_j = \mathbf{s}_j - M_j$. It is clear that

$$\lambda_{\min}\left( \sum_{j=\lfloor n\delta_2 \rfloor}^{n-1} \mathbf{s}_j \mathbf{s}_j' \right) \geq \sum_{j=0}^{g_n-1} \lambda_{\min}\left( \sum_{i=0}^{\nu k_n-1} \mathbf{s}_{\lfloor n\delta_2 \rfloor + g_n i + j} \mathbf{s}'_{\lfloor n\delta_2 \rfloor + g_n i + j} \right). \tag{A.1}$$

In addition,

$$\lambda_{\min}^{-q}\left( \frac{1}{N} \sum_{j=K_n}^{n-1} \mathbf{s}_j \mathbf{s}_j' \right) \leq N^q \lambda_{\min}^{-q}\left( \sum_{j=\lfloor n\delta_2 \rfloor}^{n-1} \mathbf{s}_j \mathbf{s}_j' \right), \tag{A.2}$$

By (A.1), (A.2), and the convexity of $x^{-q}, x > 0$,

$$\lambda_{\min}^{-q}\left( \frac{1}{N} \sum_{j=K_n}^{n-1} \mathbf{s}_j \mathbf{s}_j' \right)$$
$$\leq C \left( \frac{\nu k_n}{1-\delta_2} \right)^q \frac{1}{g_n} \sum_{j=0}^{g_n-1} \lambda_{\min}^{-q}\left( \sum_{i=0}^{\nu k_n-1} \mathbf{s}_{\lfloor n\delta_2 \rfloor + g_n i + j} \mathbf{s}'_{\lfloor n\delta_2 \rfloor + g_n i + j} \right), \tag{A.3}$$

where $C$ in (A.3) and the rest of this paper denotes a generic positive constant independent of $n$ and of any index with an upper (or lower) limit depending on $n$ (but it may represent different values in different places). In view of (A.3), (10) is guaranteed by showing that for all $j = 0, \ldots, g_n - 1$, there is a constant $C$ independent of $n$ and $j$ such that, for all sufficiently large $n$,

$$\mathrm{E}\lambda_{\min}^{-q}\left( \sum_{i=0}^{\nu k_n-1} \mathbf{s}_{\lfloor n\delta_2 \rfloor + g_n i + j} \mathbf{s}'_{\lfloor n\delta_2 \rfloor + g_n i + j} \right) \leq C k_n^{(2d-1+\theta)q}. \tag{A.4}$$

In the rest of this proof, we only verify (A.4) for the case of $j = 0$, since the other cases can be similarly verified.

Write $\phi_i = \mathbf{s}_{\lfloor n\delta_2 \rfloor + g_n i}$. Then, by reasoning analogous to (2.10) of Ing and Wei (2003),

$$\mathrm{E}\lambda_{\min}^{-q}\left( \sum_{i=0}^{\nu k_n-1} \phi_i \phi_i' \right)$$
$$\leq [\bar{C} k_n^{(2d-1+\theta)}]^q + \int_{[\bar{C} k_n^{(2d-1+\theta)}]^q}^{\infty} P\left( \inf_{\|\mathbf{h}\|=1} \sum_{i=0}^{\nu k_n-1} (\mathbf{h}'\phi_i)^2 < u^{-1/q}, \right.$$

$$\sum_{i=0}^{\nu k_n-1} \|\phi_i\|^2 \le u^{2lq^{-1}} k_n^{-1}\Bigg) du$$

$$+\int_{[\bar{C}k_n^{(2d-1+\theta)}]^q}^{\infty} P\left(\sum_{i=0}^{\nu k_n-1} \|\phi_i\|^2 > u^{2lq^{-1}} k_n^{-1}\right) du$$

$$\equiv \left[\bar{C}k_n^{(2d-1+\theta)}\right]^q + (I) + (II), \tag{A.5}$$

where $l \ge (3+q)/2$ and $\bar{C} > \max\{1, 36\sigma^2\delta^{-2}C^{*^{-1}}\}$, with $C^* > 0$ defined in (A.10) below. Since (3) implies for all $t \ge 0$ and $j = 1, \dots, d$,

$$|\kappa_t(j)| \le C(t+1)^{j-1}, \tag{A.6}$$

a straightforward calculation gives, for all $i = 0, 1, \dots, \nu k_n - 1$, $\mathrm{E}\|\phi_i\|^2 \le Ck_n$, which, together with Chebyshev's inequality, yields

$$(II) \le C. \tag{A.7}$$

To deal with $(I)$, consider the hypersphere $S_n = \{\varphi : \|\varphi\| = 1\} \subset R^{k_n}$ and the hypercube

$$H^{k_n} = [1 - 2u^{-\{l+(1/2)\}q^{-1}}(\lfloor u^{\{l+(1/2)\}q^{-1}} \rfloor + 1),\ 1]^{k_n},$$

with $u \ge [\bar{C}k_n^{(2d-1+\theta)}]^q$. Divide $H^{k_n}$ into subhypercubes, each of which has an edge of length $2u^{-\{l+(1/2)\}q^{-1}}$ and a circumscribed hypersphere of radius $\sqrt{k_n}u^{-\{l+(1/2)\}q^{-1}}$. Let these subhypercubes be denoted by $B_i, i = 1, \dots, m^*$. Then, it can be seen that the number of $B_i$'s, $m^*$, does not exceed $(\lfloor u^{\{l+(1/2)\}q^{-1}} \rfloor + 1)^{k_n}$. Define $G_i = S_n \cap B_i$ and let $\{G_{\zeta_s} : s = 1, \dots, \bar{m}\}$ denote all nonempty $G_i$'s. Since $S_n \subseteq H^{k_n}$, $S_n = \cup_{i=1}^{\bar{m}} G_{\zeta_i}$. The arguments similar to those used in (2.11) and (2.12) of Ing and Wei (2003) yield

$$P\left(\inf_{\|\mathbf{h}\|=1} \sum_{i=0}^{\nu k_n-1} (\mathbf{h}'\phi_i)^2 < u^{-q^{-1}}, \sum_{i=0}^{\nu k_n-1} \|\phi_i\|^2 \le u^{2lq^{-1}} k_n^{-1}\right)$$

$$\le \sum_{j=1}^{\bar{m}} \mathrm{E}\left(\prod_{i=0}^{\nu k_n-1} I_{D_{j,i}}\right), \tag{A.8}$$

where $I_{D_{j,i}}$ is the indicator function for the event $D_{j,i}$ and $D_{j,i} = \{|\mathbf{l}_j'\phi_i| \le 3u^{-1/2q}\}$, with $\mathbf{l}_j$ being a vector arbitrarily chosen from $G_{\zeta_j}$. Obviously,

$$\mathrm{E}\left(\prod_{i=0}^{\nu k_n-1} I_{D_{j,i}}\right)$$

$$= \mathrm{E}\Bigg\{\prod_{i=0}^{\nu k_n-2} I_{D_{j,i}}$$

$$P(|\mathbf{l}_j' M_{\lfloor n\delta_2\rfloor+g_n(\nu k_n-1)} + \mathbf{l}_j' R_{\lfloor n\delta_2\rfloor+g_n(\nu k_n-1)}| \le 3u^{-1/2q} | \varepsilon_r, r \le \iota_n)\Bigg\}, \tag{A.9}$$

where $\iota_n = \lfloor n\delta_2 \rfloor + g_n(\nu k_n - 2)$. It is shown in Lemma A.2 that for all sufficiently large $n$ and $j_1 = 0, \ldots, \nu k_n - 1$,

$$\operatorname{var}(\mathbf{l}_j' M_{\lfloor n\delta_2 \rfloor + g_n j_1}) \geq \frac{C^*}{k_n^{2d-1}}. \tag{A.10}$$

By (A.9), (A.10), NS, and independence of $\varepsilon_j$'s, we have, for $u \geq [\bar{C} k_n^{(2d-1+\theta)}]^q$ and all sufficiently large $n$,

$$\begin{aligned} \mathrm{E}\left(\prod_{i=0}^{\nu k_n - 1} I_{D_{j,i}}\right) &\leq \mathrm{E}\left(\prod_{i=0}^{\nu k_n - 2} I_{D_{j,i}}\right) M \left(6 k_n^{d-(1/2)} \sigma C^{*^{-1/2}} u^{-1/2q}\right)^{\alpha} \\ &\leq M^{\nu k_n} \left(6 k_n^{d-(1/2)} \sigma C^{*^{-1/2}} u^{-1/2q}\right)^{\alpha \nu k_n}, \end{aligned} \tag{A.11}$$

where the second inequality is obtained from repeating the same argument $\nu k_n - 1$ times. By (A.8), (A.11), and taking

$$\nu \geq \left\lfloor \frac{2\left(d - \frac{1}{2} + \frac{\theta}{2}\right)(2l + 1 + 2q)}{\alpha\theta} \right\rfloor + 1,$$

one obtains, for sufficiently large $n$,

$$(I) \leq C M^{\nu k_n} \left(6 k_n^{d-(1/2)} \sigma C^{*^{-1/2}}\right)^{\alpha \nu k_n} \int_{[\bar{C} k_n^{(2d-1+\theta)}]^q}^{\infty} u^{-k_n(\alpha\nu - 2l - 1)/2q} du \leq C. \tag{A.12}$$

In view of (A.5), (A.7), and (A.12), the proof is complete. ■

To prove (A.10), we need an auxiliary lemma.

LEMMA A.1. *Assume that (1) with* $d \geq 1$ *and (2) hold. Define*

$$\mathbf{V}_n(d) = \left( \frac{1}{q_n^{d-(1/2)}} \sum_{j=0}^{q_n - 1} \kappa_j(d) \varepsilon_{n-j}, \ldots, \frac{1}{q_n^{1/2}} \sum_{j=0}^{q_n - 1} \kappa_j(1) \varepsilon_{n-j} \right)',$$

*where* $1 \leq q_n \leq n$ *and* $q_n \to \infty$ *as* $n \to \infty$. *Then,*

$$\lim_{n\to\infty} \mathrm{E}\left\{\mathbf{V}_n(d)\mathbf{V}_n'(d)\right\} = Z(d) = [z_{ij}]_{i,j=1,\ldots,d}, \tag{A.13}$$

*where* $z_{ij} = V_0^2(2d - i - j + 1)^{-1}\{(d-i)!(d-j)!\}^{-1}$, *with* $V_0^2 = \sigma^2(\sum_{j=0}^{\infty} b_j)^2$ *and the convention that* $0! = 1$. *In addition,* $Z(d)$ *is positively definite for all* $d = 1, 2, \ldots$.

**Proof.** (A.13) can be obtained by (3), (A.6), and a straightforward calculation. The details are omitted. To show that $Z(d)$ is positively definite, write

$$Z(d) = V_0^2 \operatorname{diag}(1/(d-1)!, \ldots, 1/0!)\Delta(d) H(d) \Delta'(d) \operatorname{diag}(1/(d-1)!, \ldots, 1/0!), \tag{A.14}$$

where the $(i, j)$-th element of $H(d)$ is given by $(i + j - 1)^{-1}$ and $\Delta(d) = [f_{i,j}]_{i,j=1,\ldots,d}$, with $f_{i,j} = 1$ if $i + j = d + 1$ and $f_{i,j} = 0$ if $i + j \neq d + 1$ . Since $H(d)$, known as Hilbert matrix, is positively definite (see Barria and Halmos, 1982, or Choi, 1983), in view of (A.14), the positive definiteness of $Z(d)$ follows. ■

LEMMA A.2. *Assume that (1) with $d \geq 1$, (2), and MO(i) hold. Then (A.10) follows.*

**Proof.** We only prove the case of $j_1 = 0$ since the other cases can be obtained similarly. By noticing that $\|\mathbf{l}_j\| = 1$, and the definition of $\lambda_{\min}(.)$, we have

$$\begin{aligned}
\operatorname{var}(\mathbf{l}_j' M_{\lfloor n\delta_2 \rfloor}) &\geq \lambda_{\min}\left\{\mathrm{E}(M_{\lfloor n\delta_2 \rfloor} M'_{\lfloor n\delta_2 \rfloor})\right\} \\
&\geq \lambda_{\min}\left\{\begin{pmatrix} \mathrm{E}(M^{(1)}_{\lfloor n\delta_2 \rfloor} M^{(1)'}_{\lfloor n\delta_2 \rfloor}) & \mathbf{0}_{(k_n-d)\times d} \\ \mathbf{0}'_{d\times(k_n-d)} & \mathrm{E}(M^{(2)}_{\lfloor n\delta_2 \rfloor} M^{(2)'}_{\lfloor n\delta_2 \rfloor}) \end{pmatrix}\right\} \\
&\quad - \left\| \mathrm{E}(M_{\lfloor n\delta_2 \rfloor} M'_{\lfloor n\delta_2 \rfloor}) - \begin{pmatrix} \mathrm{E}(M^{(1)}_{\lfloor n\delta_2 \rfloor} M^{(1)'}_{\lfloor n\delta_2 \rfloor}) & \mathbf{0}_{(k_n-d)\times d} \\ \mathbf{0}'_{d\times(k_n-d)} & \mathrm{E}(M^{(2)}_{\lfloor n\delta_2 \rfloor} M^{(2)'}_{\lfloor n\delta_2 \rfloor}) \end{pmatrix} \right\| \\
&\geq \min\left\{\lambda_{\min}\left\{\mathrm{E}(M^{(1)}_{\lfloor n\delta_2 \rfloor} M^{(1)'}_{\lfloor n\delta_2 \rfloor})\right\}, \lambda_{\min}\left\{\mathrm{E}(M^{(2)}_{\lfloor n\delta_2 \rfloor} M^{(2)'}_{\lfloor n\delta_2 \rfloor})\right\}\right\} \\
&\quad - \left\| \begin{pmatrix} \mathbf{0}_{(k_n-d)\times(k_n-d)} & \mathrm{E}(M^{(1)}_{\lfloor n\delta_2 \rfloor} M^{(2)'}_{\lfloor n\delta_2 \rfloor}) \\ \mathrm{E}(M^{(2)}_{\lfloor n\delta_2 \rfloor} M^{(1)'}_{\lfloor n\delta_2 \rfloor}) & \mathbf{0}_{d\times d} \end{pmatrix} \right\|, \qquad \textbf{(A.15)}
\end{aligned}$$

where the matrix norms are defined in Remark 3,

$$M^{(1)}_{\lfloor n\delta_2 \rfloor} = \left(\sum_{i=0}^{g_n-1} b_i \varepsilon_{\lfloor n\delta_2 \rfloor - i}, \ldots, \sum_{i=0}^{g_n-k_n+d} b_i \varepsilon_{\lfloor n\delta_2 \rfloor - k_n + d + 1 - i}\right)',$$

and

$$M^{(2)}_{\lfloor n\delta_2 \rfloor} = \left(\frac{1}{N^{d-(1/2)}} \sum_{s=0}^{g_n-1} \kappa_s(d) \varepsilon_{\lfloor n\delta_2 \rfloor - s}, \ldots, \frac{1}{N^{1/2}} \sum_{s=0}^{g_n-1} \kappa_s(1) \varepsilon_{\lfloor n\delta_2 \rfloor - s}\right)'.$$

By (2) and (3), it is not difficult to see that, for some $C_1^* > 0$ and all sufficiently large $n$,

$$\lambda_{\min}\left\{\mathrm{E}(M^{(1)}_{\lfloor n\delta_2 \rfloor} M^{(1)'}_{\lfloor n\delta_2 \rfloor})\right\} \geq C_1^*. \qquad \textbf{(A.16)}$$

Also observe that

$$\begin{aligned}
\mathrm{E}(M^{(2)}_{\lfloor n\delta_2 \rfloor} M^{(2)'}_{\lfloor n\delta_2 \rfloor}) &= \operatorname{diag}\left(\left(\frac{g_n}{N}\right)^{d-(1/2)}, \ldots, \left(\frac{g_n}{N}\right)^{1/2}\right) Z_n(d) \\
&\quad \times \operatorname{diag}\left(\left(\frac{g_n}{N}\right)^{d-(1/2)}, \ldots, \left(\frac{g_n}{N}\right)^{1/2}\right),
\end{aligned}$$

where, with

$$\bar{M}^{(2)}_{\lfloor n\delta_2 \rfloor} = \left(\frac{1}{g_n^{d-(1/2)}} \sum_{s=0}^{g_n-1} \kappa_s(d) \varepsilon_{\lfloor n\delta_2 \rfloor - s}, \ldots, \frac{1}{g_n^{1/2}} \sum_{s=0}^{g_n-1} \kappa_s(1) \varepsilon_{\lfloor n\delta_2 \rfloor - s}\right)',$$

$Z_n(d) = \mathrm{E}(\bar{M}^{(2)}_{\lfloor n\delta_2 \rfloor} \bar{M}^{(2)'}_{\lfloor n\delta_2 \rfloor})$. By Lemma A.1, $\lim_{n\to\infty} Z_n = Z(d)$, and hence, for some $C_2^* > 0$ and all sufficiently large $n$,

$$\lambda_{\min}\left\{\mathrm{E}(M^{(2)}_{\lfloor n\delta_2 \rfloor} M^{(2)'}_{\lfloor n\delta_2 \rfloor})\right\} \geq C_2^* k_n^{1-2d}. \qquad \textbf{(A.17)}$$

Moreover, armed with (A.6), it can be shown that, for some $C_3^* > 0$ and all sufficiently large $n$,

$$\left\| \begin{pmatrix} \mathbf{0}_{(k_n-d)\times(k_n-d)} & \mathrm{E}(M^{(1)}_{\lfloor n\delta_2\rfloor} M^{(2)\prime}_{\lfloor n\delta_2\rfloor}) \\ \mathrm{E}(M^{(2)}_{\lfloor n\delta_2\rfloor} M^{(1)\prime}_{\lfloor n\delta_2\rfloor}) & \mathbf{0}_{d\times d} \end{pmatrix} \right\| \le C_3^* \left(\frac{k_n}{N}\right)^{1/2}. \tag{A.18}$$

Consequently, the desired result follows from (A.15)–(A.18) and the assumption that $K_n = o(n^{1/(4d-1)})$. ■

The following lemma is required in the proof of Theorem 1.

LEMMA A.3. *Assume (1), (2), and for some* $q \ge 2$, $\sup_{0<t<\infty} \mathrm{E}(|\varepsilon_t|^{2q}) < \infty$. *Then, for* $1 \le K_n \le n-1$ *and* $1 \le l \le d$,

$$\max_{1\le k\le K_n} \frac{\mathrm{E}\left\| N^{-l-(1/2)} \sum_{j=K_n}^{n-1} \mathbf{z}_j(k) y_j(l) \right\|^q}{\left(\frac{k}{N}\right)^{q/2}} \le C. \tag{A.19}$$

**Proof.** By the convexity of $x^q$, $x > 0$,

$$\mathrm{E}\left\| N^{-(1/2)-l} \sum_{j=K_n}^{n-1} \mathbf{z}_j(k) y_j(l) \right\|^q$$

$$\le \left(\frac{k}{N}\right)^{q/2} (k)^{-1} \sum_{s=0}^{k-1} \mathrm{E}\left\{ N^{-lq} \left| \sum_{j=K_n}^{n-1} z_{j-s} y_j(l) \right|^q \right\}. \tag{A.20}$$

In view of (A.20), it remains to be shown that, for $s = 0, 1, \ldots, K_n - 1$,

$$\mathrm{E}\left\{ N^{-lq} \left| \sum_{j=K_n}^{n-1} z_{j-s} y_j(l) \right|^q \right\} \le C. \tag{A.21}$$

We only show (A.21) for the case $s = 0$, since the other cases can be obtained similarly. Changing the order of summation, we obtain

$$\begin{aligned} \sum_{j=K_n}^{n-1} z_j y_j(l) = {} & \sum_{t=1}^{n-1} \left( \sum_{j=K_n\vee t}^{n-1} \kappa_{j-t}(l) b_{j-t} \right) \varepsilon_t^2 + \sum_{s=2}^{n-1} \left\{ \sum_{t=1}^{s-1} \left( \sum_{j=K_n\vee s}^{n-1} \kappa_{j-t}(l) b_{j-s} \right) \varepsilon_t \right\} \varepsilon_s \\ & + \sum_{t=2}^{n-1} \left\{ \sum_{s=1}^{t-1} \left( \sum_{j=K_n\vee t}^{n-1} \kappa_{j-t}(l) b_{j-s} \right) \varepsilon_s \right\} \varepsilon_t \\ \equiv {} & (I) + (II) + (III). \end{aligned} \tag{A.22}$$

By (A.6) and (3), one has, for all $1 \le t < n-1$,

$$\left| \sum_{j=K_n\vee t}^{n-1} \kappa_{j-t}(l) b_{j-t} \right| \le \begin{cases} C & l = 1, 2, \\ C n^{l-2} & l \ge 3. \end{cases} \tag{A.23}$$

Equation (A.23), the assumption that $\sup_{0<t<\infty} \mathrm{E}(|\varepsilon_t|^{2q}) < \infty$, and Minkowski's inequality together imply

$$N^{-lq} \mathrm{E}|(I)|^q \le C. \tag{A.24}$$

Since both (II) and (III) are martingale transformations, by (A.6), (3), Minkowski's inequality, and applying Wei (1987, Lem. 2) repeatedly, we obtain

$$\mathrm{E}|(II)|^q \leq Cn^{lq} \quad \text{and} \quad \mathrm{E}|(III)|^q \leq Cn^{lq}. \tag{A.25}$$

Consequently, (A.21) follows from (A.22), (A.24), and (A.25). ■

**Proof of Theorem 1.** We only prove the case of $d \geq 1$, since the case of $d = 0$ can be shown by an argument similar to that used in the proof of Theorem 2 in Ing and Wei (2003). We start by verifying (16). In view of the definition of $\hat{S}_{n,d}(k)$, it suffices to consider the case of $k \geq d+1$. First note that for all $d+1 \leq k \leq K_n$,

$$\begin{aligned}
&\mathrm{E}\|\hat{S}_n(k) - \hat{S}_{d,n}(k)\|^{q_1} \\
&= \mathrm{E}\left\| \begin{pmatrix} \frac{1}{N}\sum_{j=K_n}^{n-1} \mathbf{z}_j(k-d)\mathbf{z}_j'(k-d) - \Gamma(k-d) & \frac{1}{N}\sum_{j=K_n}^{n-1} \mathbf{z}_j(k-d)U_{j,n}'(d) \\ \frac{1}{N}\sum_{j=K_n}^{n-1} U_{j,n}(d)\mathbf{z}_j'(k-d) & \mathbf{0}_{d\times d} \end{pmatrix} \right\|^{q_1} \\
&\leq C\left(\frac{k^{q_1}}{N^{q_1/2}}\right),
\end{aligned} \tag{A.26}$$

where the inequality follows from Lemma A.3 and an analogy with Lemma 2 of Ing and Wei (2003). In addition, Lemma 1 yields, for any $r > 0$ and $1 > \theta > 0$,

$$\mathrm{E}\|\hat{S}_n^{-1}(K_n)\|^r = O(K_n^{(2d+\theta)r}). \tag{A.27}$$

By Lemma 1 (taking $k_n = d$) and (2), one has, for any $r > 0$,

$$\mathrm{E}\|\hat{S}_{d,n}^{-1}(K_n)\|^r = O(1). \tag{A.28}$$

Since

$$\|\hat{S}_n^{-1}(k) - \hat{S}_{d,n}^{-1}(k)\|^q \leq \|\hat{S}_n^{-1}(K_n)\|^q \|\hat{S}_n(k) - \hat{S}_{d,n}(k)\|^q \|\hat{S}_{d,n}^{-1}(K_n)\|^q, \tag{A.29}$$

(A.26)–(A.28) and Hölder's inequality imply, for all $d+1 \leq k \leq K_n$ and all sufficiently large $n$,

$$\begin{aligned}
\mathrm{E}\|\hat{S}_n^{-1}(k) - \hat{S}_{d,n}^{-1}(k)\|^q &\leq C\left(K_n^{(2d+\theta)q}\right)\left(\mathrm{E}\|\hat{S}_n(k) - \hat{S}_{d,n}(k)\|^{q_1}\right)^{q/q_1} \\
&\leq C\left(\frac{K_n^{4d+2+2\theta}}{N}\right)^{q/2}.
\end{aligned} \tag{A.30}$$

Set $2\theta \leq \delta_1$. Then (16) follows from (A.30) and MO(iii). Moreover, (17) is an immediate consequence of (16), (A.28), and the fact that

$$\|\hat{S}_n^{-1}(k)\|^q \leq C(\|\hat{S}_{d,n}^{-1}(k)\|^q + \|\hat{S}_n^{-1}(k) - \hat{S}_{d,n}^{-1}(k)\|^q).$$

Finally, (18) is guaranteed by (17), (A.26), (A.28), (A.29), and Hölder's inequality.

The second part of this theorem follows from (A.30), MO(ii), and an argument similar to that used to verify Theorem 2(ii) in Ing and Wei (2003). The details are skipped. ■

# APPENDIX B

LEMMA B.1. *Assume (1) with* $d \geq 1$, *(2), and* $\sup_{0<t<\infty} \mathrm{E}|\varepsilon_t|^q < \infty, q \geq 2$. *Then, for* $K_n \geq d$ *and* $K_n = o(n)$,

$$\max_{K_n \leq j \leq n-1} \mathrm{E}\|U_{j,n}(d)\|^q = O(1), \tag{B.1}$$

*and*

$$\max_{K_n \leq l_1 \leq l_2 \leq n-1} \mathrm{E}\left\| \frac{1}{\sqrt{l_2 - l_1 + 1}} \sum_{j=l_1}^{l_2} U_{j,n}(d)\varepsilon_{j+1} \right\|^q = O(1), \tag{B.2}$$

*where* $U_{j,n}(d)$ *is defined after (23).*

**Proof.** Equalities (B.1) and (B.2) can be shown by (A.6) and an argument similar to that used in the proof of Lemma 4 of Ing and Wei (2003). The details are omitted. ■

LEMMA B.2.

$$\mathrm{E}\{\sigma^2(\iota' F^{-1}\zeta)^2\} = \sigma^2 d(d+1), \tag{B.3}$$

*where* $\iota$, $F$, *and* $\zeta$ *are defined in (26).*

**Proof.** Since it is difficult to obtain (B.3) through direct calculations, we adopt an indirect approach. Consider an AR($d$) model,

$$x_t = \beta_1 x_{t-1} + \cdots + \beta_d x_{t-d} + \epsilon_t, \tag{B.4}$$

where, for $i = 1, \ldots, d$, $\beta_i = (-1)^{i+1} C_i^d$ with $C_{m_2}^{m_1} = m_1!/[m_2!(m_1 - m_2)!]$, $\epsilon_t$'s are i.i.d. normal random variables with zero means and variances $\sigma^2$, and $x_t = 0$ for $t \leq 0$. Having observed $x_1, \ldots, x_i$, the least squares estimator of $\beta(d) = (\beta_1, \ldots, \beta_d)'$, $\hat{\beta}_i(d)$, is given by $(\sum_{j=d}^{i-1} \mathbf{x}_j(d)\mathbf{x}_j'(d))\hat{\beta}_i(d) = \sum_{j=d}^{i-1} \mathbf{x}_j(d) x_{j+1}$, where $\mathbf{x}_j(d) = (x_j, \ldots, x_{j-d+1})'$. By Chan and Wei (1988), it can be shown that

$$n\{\mathbf{x}_n'(d)(\hat{\beta}_n(d) - \beta(d))\}^2 \Rightarrow \sigma^2(\iota' F^{-1}\zeta)^2, \quad \text{as } n \to \infty. \tag{B.5}$$

Since model (B.4) is a special case of model (1), by Lemma 1 (taking $k_n = d$), (B.1), (B.2), and Hölder's inequality, there is a positive integer $h^*$ such that, for any $\delta_1 > 0$,

$$\sup_{n \geq h^*} \mathrm{E}|n\{\mathbf{x}_n'(d)(\hat{\beta}_n(d) - \beta(d))\}^2|^{1+\delta_1} \leq C, \tag{B.6}$$

which implies that the sequence $\{n\{\mathbf{x}_n'(d)(\hat{\beta}_n(d) - \beta(d))\}^2\}_{n \geq h^*}$ is uniformly integrable (see Chow and Teicher, 1997, Ex. 4.2.6). This fact, (B.5), and Billingsley (1968, Thm. 5.4) give

$$\lim_{n\to\infty} \mathrm{E}[n\{\mathbf{x}_n'(d)(\hat{\beta}_n(d) - \beta(d))\}^2] = \sigma^2 \mathrm{E}\{(\iota' F^{-1}\zeta)^2\}. \tag{B.7}$$

In addition, by Theorem 5 of Wei (1987),

$$\left[\frac{1}{\log n} \sum_{i=h^*}^{n} \{\mathbf{x}_i'(d)(\hat{\beta}_i(d) - \beta(d))\}^2\right] - \sigma^2 d(d+1) = o_p(1). \tag{B.8}$$

According to (B.6) and Minkowski's inequality, for any $\delta_1 > 0$,

$$\sup_{n \geq h^*} \mathrm{E}\left| \frac{1}{\log n} \sum_{i=h^*}^{n} \{\mathbf{x}_i'(d)(\hat{\beta}_i(d) - \beta(d))\}^2 \right|^{1+\delta_1} \leq C, \tag{B.9}$$

which, together with (B.8), yields

$$\lim_{n \to \infty} \frac{1}{\log n} \sum_{i=h^*}^{n} \mathrm{E}\{\mathbf{x}_i'(d)(\hat{\beta}_i(d) - \beta(d))\}^2 = \sigma^2 d(d+1). \tag{B.10}$$

Consequently, (B.3) follows from (B.7) and (B.10). ■

**Proof of Lemma 2.** By the arguments used in Phillips (1987) and Chan and Wei (1988),

$$\left\| \left( \frac{1}{n-d} \sum_{j=d}^{n-1} \bar{U}_{j,n}(d) \bar{U}_{j,n}'(d) \right)^{-1} \right\| = O_p(1).$$

This and Lemma B.1 imply

$$N^{1/2} \mathbf{f}_{1,n}(d) - (n-d)^{1/2} \bar{\mathbf{f}}_{1,n}(d) = o_p(1),$$

which, together with (26), Slutsky's lemma, and the continuous mapping theorem, yields

$$N \mathbf{f}_{1,n}^2(d) \Rightarrow \sigma^2 (\iota' F^{-1} \zeta)^2. \tag{B.11}$$

Moreover, by Lemmas 1 and B.1, Hölder's inequality, and the moment condition imposed on $\{\varepsilon_t\}$, one has

$$\sup_{n \geq h^*} \mathrm{E}|N \mathbf{f}_{1,n}^2(d)|^{1+\delta_1} < \infty,$$

where $h^*$ is some positive integer and $\delta_1$ is some positive number. Consequently, the desired result follows from the above inequality, (B.3), and (B.11). ■

To verify Theorem 2, we need several auxiliary lemmas.

LEMMA B.3. *Assume that (1) holds and* $\sup_{0<t<\infty} \mathrm{E}|\varepsilon_t|^q < \infty, q \geq 2$. *Then,*

$$\sup_{j \geq 1} \max_{1 \leq k \leq j} k^{-q/2} \mathrm{E}\|\mathbf{z}_j(k)\|^q \leq C, \tag{B.12}$$

*and*

$$\sup_{l_2 \geq l_1 \geq 1} \max_{1 \leq k \leq l_1} k^{-q/2} \mathrm{E}\left\| \frac{1}{\sqrt{l_2 - l_1 + 1}} \sum_{j=l_1}^{l_2} \mathbf{z}_j(k) \varepsilon_{j+1} \right\|^q \leq C. \tag{B.13}$$

**Proof.** Inequalities (B.12) and (B.13) can be shown by an argument used in the proof of Lemma 4 in Ing and Wei (2003). We skip the details. ■

LEMMA B.4. *Under the same assumptions as in Lemma B.1, but with the moment condition replaced with* $\sup_{0<t<\infty} \mathrm{E}|\varepsilon_t|^{2q} < \infty, q \geq 2$, *one has, for all* $0 \leq k \leq K_n$,

$$\mathrm{E}\left\| \frac{1}{\sqrt{N}} \sum_{j=K_n}^{n-1} U_{j,n}(d)(\varepsilon_{j+1,k} - \varepsilon_{j+1}) \right\|^q \leq C \left( \sum_{i=1}^{\infty} |a_i - a_i(k)| \right)^q \leq C \left( \sum_{i \geq k+1}^{\infty} |a_i| \right)^q. \tag{B.14}$$

**Proof.** First note that

$$\mathrm{E}\left\|\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1} U_{j,n}(d)(\varepsilon_{j+1,k}-\varepsilon_{j+1})\right\|^q$$

$$=\mathrm{E}\left\{\sum_{l=1}^{d}\left(\sum_{j=K_n}^{n-1}\frac{y_j(l)}{N^l}(\varepsilon_{j+1,k}-\varepsilon_{j+1})\right)^2\right\}^{q/2}$$

$$\leq d^{-1+(q/2)}\sum_{l=1}^{d}\mathrm{E}\left|\sum_{j=K_n}^{n-1}\frac{y_j(l)}{N^l}(\varepsilon_{j+1,k}-\varepsilon_{j+1})\right|^q$$

$$\leq d^{-1+(q/2)}\sum_{l=1}^{d}\mathrm{E}\left|\sum_{m=1}^{n-1}(a_m-a_m(k))\left(\sum_{j=K_n\vee m}^{n-1}\frac{y_j(l)}{N^l}z_{j+1-m}\right)\right|^q$$

$$\leq d^{-1+(q/2)}\sum_{l=1}^{d}\left[\sum_{m=1}^{n-1}|a_m-a_m(k)|\left(\mathrm{E}\left|\sum_{j=K_n\vee m}^{n-1}\frac{y_j(l)}{N^l}z_{j+1-m}\right|^q\right)^{1/q}\right]^q,$$

where the first inequality is due to the convexity of $x^{q/2}, x\geq 0$, the second inequality follows from (22) and changing the order of summation, and the last one is due to Minkowski's inequality. Moreover, by an argument similar to that used to obtain (A.21), one has, for all $1\leq l\leq d$ and $1\leq m\leq n-1$,

$$\mathrm{E}\left|\sum_{j=K_n\vee m}^{n-1}\frac{y_j(l)}{N^l}z_{j+1-m}\right|^q\leq C.$$

As a result, the first inequality of (B.14) follows. The second inequality of (B.14) is an immediate consequence of Lemma 4 of Berk (1974). ■

LEMMA B.5. *Assume (1) and (2). Then, for $K_n=o(n)$, there are sequences of positive numbers $\{\omega_n\}$ and $\{\chi_n\}$, with $\omega_n=o(n^{-1})$ and $\chi_n=o(n^{-2})$, such that, for all $0\leq k\leq K_n$,*

$$\left|\mathrm{E}\left(\varepsilon_{n+1,k}-\varepsilon_{n+1}\right)^2-\|\mathbf{a}-\mathbf{a}(k)\|_z^2\right|\leq\omega_n\sum_{i=1}^{\infty}(a_i-a_i(k))^2+\chi_n. \tag{B.15}$$

**Proof.** Denote $a_i-a_i(k)$ by $r_i(k)$. Then, algebraic manipulations yield

$$\mathrm{E}(\varepsilon_{n+1,k}-\varepsilon_{n+1})^2=\mathrm{E}\left(\sum_{i=1}^{n}r_i(k)z_{n+1-i,\infty}\right)^2-\mathrm{E}\left\{\sum_{i=1}^{n}r_i(k)(z_{n+1-i,\infty}-z_{n+1-i})\right\}^2$$
$$=(I)-(II). \tag{B.16}$$

Since $K_n=o(n)$, there are $0<\varrho<1$ and $M^*\geq 1$ such that, for all $n\geq M^*$, $K_n\leq\varrho n$. Now, for all $n\geq M^*$ and $0\leq k\leq K_n$,

$$\sigma^{-2}(II)=\sum_{l=0}^{\infty}\left(\sum_{i=1}^{n}r_i(k)b_{n+1+l-i}\right)^2$$

$$\le 2\left\{\sum_{l=0}^{\infty}\left(\sum_{i=1}^{\varrho n} r_i(k) b_{n+1+l-i}\right)^2 + \sum_{l=0}^{\infty}\left(\sum_{i=\varrho n+1}^{n} r_i(k) b_{n+1+l-i}\right)^2\right\}$$

$$\le 2\left[\sum_{i_1=1}^{\varrho n}\sum_{i_2=1}^{\varrho n} |r_{i_1}(k) r_{i_2}(k)| \left(\sum_{l=0}^{\infty} |b_{n+1+l-i_1} b_{n+1+l-i_2}|\right)\right.$$

$$\left. + \sum_{l=0}^{\infty}\left\{\sum_{i=\varrho n+1}^{n} r_i^2(k) \sum_{i=\varrho n+1}^{n} b_{n+1+l-i}^2\right\}\right]$$

$$\le 2\left(\sum_{i=1}^{\varrho n} r_i^2(k)\right) \varrho n \left\{\sum_{j=(1-\varrho)n+1}^{\infty} b_j^2\right\} + 2\left(\sum_{i=\varrho n+1}^{n} r_i^2(k)\right)$$

$$\times \left\{\sum_{j=1}^{(1-\varrho)n} j b_j^2 + (1-\varrho) n \sum_{j=(1-\varrho)n+1}^{\infty} b_j^2\right\}$$

$$\le \omega_{1,n} \sum_{i=1}^{\infty} (a_i - a_i(k))^2 + C_1 \sum_{i=\varrho n+1}^{\infty} a_i^2, \tag{B.17}$$

where $\omega_{1,n} = 2n\sum_{j=(1-\varrho)n+1}^{\infty} b_j^2$ and $C_1 = 2\sum_{j=1}^{\infty} j b_j^2$. Note that according to (3), $\omega_{1,n} = o(n^{-1})$ and $C_1 < \infty$. Similarly, for all $n \ge M^*$ and $0 \le k \le K_n$, there are $C_2, C_3 > 0$ such that

$$\sigma^{-2}|(I) - \|\mathbf{a} - \mathbf{a}(k)\|_z^2| \le C_2 \sum_{i=\varrho n+1}^{\infty} a_i^2 + C_3 \left\{n^{-1} \sum_{l=(1-\varrho)n+1}^{\infty} |\gamma_l|\right\} \sum_{i=k+1}^{\infty} |a_i|, \tag{B.18}$$

where $\gamma_l$ is defined after (4). Since (2) and (3) imply

$$\sum_{i=\varrho n+1}^{\infty} a_i^2 = o(n^{-2}) \quad \text{and} \quad \sum_{l=(1-\varrho)n+1}^{\infty} |\gamma_l| = o(n^{-1}),$$

(B.15) follows from (B.16)–(B.18). ■

LEMMA B.6. *Assume (1), (2), and, for some* $q \ge 2$, $\sup_{0<t<\infty} \mathrm{E}|\varepsilon_t|^{2q} < \infty$. *Then, for* $1 \le K_n \le n-1$ *and all* $1 \le k \le K_n$,

$$\mathrm{E}\left\|\frac{1}{\sqrt{N}} \sum_{j=K_n}^{n-1} \mathbf{z}_j(k)(\varepsilon_{j+1,k} - \varepsilon_{j+1})\right\|^q \le C k^{q/2} \left(\|\mathbf{a} - \mathbf{a}(k)\|_z^q + N^{-q/2}\right). \tag{B.19}$$

**Proof.** First note that

$$\mathrm{E}\left\|\frac{1}{\sqrt{N}} \sum_{j=K_n}^{n-1} \mathbf{z}_j(k)(\varepsilon_{j+1,k} - \varepsilon_{j+1})\right\|^q$$

$$\le C\left\{\mathrm{E}\left\|\frac{1}{\sqrt{N}} \sum_{j=K_n}^{n-1} (\mathbf{z}_j(k) - \mathbf{z}_{j,\infty}(k))(\varepsilon_{j+1,k} - \varepsilon_{j+1})\right\|^q\right.$$

$$+ \mathrm{E}\left\| \frac{1}{\sqrt{N}} \sum_{j=K_n}^{n-1} \mathbf{z}_{j,\infty}(k)(\varepsilon_{j+1,k} - \varepsilon^*_{j+1,k}) \right\|^q$$

$$+ \mathrm{E}\left\| \frac{1}{\sqrt{N}} \sum_{j=K_n}^{n-1} \mathbf{z}_{j,\infty}(k)(\varepsilon^*_{j+1,k} - \varepsilon_{j+1}) \right\|^q \Bigg\}$$

$$\equiv C\{(I) + (II) + (III)\}, \tag{B.20}$$

where $\varepsilon^*_{j+1,k} = \sum_{i=0}^{k} a_i(k) z_{j+1-i,\infty}$ with $a_0(k) = 1$. Without loss of generality, we can assume $\sup_{-\infty < t \le 0} \mathrm{E}|\varepsilon_t|^{2q} < \infty$ in the rest of the proof. By the convexity of $x^q, x \ge 0$, Minkowski's inequality, and the Cauchy-Schwarz inequality,

$$(I) \le k^{(q/2)-1} \sum_{l=1}^{k} N^{-q/2} \left\{ \sum_{j=K_n}^{n-1} (\mathrm{E}|z_{j+1-l} - z_{j+1-l,\infty}|^{2q})^{1/2q} \right.$$

$$\left. \times (\mathrm{E}|\varepsilon_{j+1,k} - \varepsilon_{j+1}|^{2q})^{1/2q} \right\}^q. \tag{B.21}$$

In view of (B.16), for all $K_n \le j \le n-1$ and $0 \le k \le K_n$,

$$\mathrm{E}(\varepsilon_{j+1,k} - \varepsilon_{j+1})^2 \le \mathrm{E}\left( \sum_{i=1}^{j} (a_i - a_i(k)) z_{j+1-i,\infty} \right)^2 \le C \sum_{i=1}^{\infty} (a_i - a_i(k))^2$$

$$\le C\|\mathbf{a} - \mathbf{a}(k)\|_z^2,$$

where the last two inequalities are ensured by (2). This fact and Wei (1987, Lem. 2) yield that, for all $K_n \le j \le n-1$ and $0 \le k \le K_n$,

$$\mathrm{E}|\varepsilon_{j+1,k} - \varepsilon_{j+1}|^{2q} \le C\{\mathrm{E}(\varepsilon_{j+1,k} - \varepsilon_{j+1})^2\}^q \le C\|\mathbf{a} - \mathbf{a}(k)\|_z^{2q} \le C. \tag{B.22}$$

Moreover, it follows from Wei (1987, Lem. 2) and (3) that, for all $1 \le l \le K_n$,

$$\sum_{j=K_n}^{n-1} (\mathrm{E}|z_{j+1-l} - z_{j+1-l,\infty}|^{2q})^{1/2q} \le C \sum_{j=K_n}^{n-1} \sum_{s=j+1-l}^{\infty} |b_s| \le C \sum_{j=1}^{\infty} j|b_j| < \infty, \tag{B.23}$$

which, together with (B.21) and (B.22), implies

$$(I) \le Ck^{q/2} N^{-q/2}. \tag{B.24}$$

By (B.12) and arguments similar to those used in (B.21)–(B.23),

$$(II) \le k^{(q/2)-1} \sum_{l=1}^{k} N^{-q/2} \left\{ \sum_{j=K_n}^{n-1} (\mathrm{E}|z_{j+1-l,\infty}|^{2q})^{1/2q} \right.$$

$$\left. \times \sum_{i=0}^{k} |a_i(k)| (\mathrm{E}|z_{j+1-i,\infty} - z_{j+1-i}|^{2q})^{1/2q} \right\}^q$$

$$\le Ck^{(q/2)-1} \sum_{l=1}^{k} N^{-q/2} \left( \sum_{i=0}^{k} |a_i(k)| \right)^q \le Ck^{q/2} N^{-q/2}, \tag{B.25}$$

where the last inequality follows from Lemma 4 of Berk (1974), which yields $\sup_{k\geq 1}\sum_{i=0}^{k}|a_i(k)| < \infty$. Moreover, by Lemma 3 of Ing and Wei (2003),

$$(III) \leq Ck^{q/2}\|\mathbf{a}-\mathbf{a}(k)\|_z^q. \tag{B.26}$$

Consequently, (B.19) is ensured by (B.20) and (B.24)–(B.26). ■

LEMMA B.7. *Assume (1), (2), NS, and* $\sup_{0<t<\infty} \mathrm{E}|\varepsilon_t|^{6+\theta_1} < \infty$ *for some* $\theta_1 > 0$. *Then, for* $K_n = o(n^{1/2})$,

$$\lim_{n\to\infty} \max_{\max\{1,d\}\leq k\leq K_n}\left|\frac{\mathrm{E}\{\mathbf{f}_{1,n}(d)+\mathbf{f}_{2,n}(k-d)+\mathcal{S}_n(k-d)\}^2}{L_n^d(k)}-1\right| = 0,$$

*where* $L_n^d(k)$ *is defined in Theorem 2.*

**Proof.** We only prove the case of $d \geq 1$, since the case of $d = 0$ can be shown by an argument similar to that used in the proof of Theorem 3 in Ing and Wei (2003). Define

$$(I) = \max_{d\leq k\leq K_n}\left|\frac{\mathrm{E}(\mathbf{f}_{1,n}(d)+\mathbf{f}_{2,n}(k-d)+\mathcal{S}_n(k-d))^2}{L_n^d(k)}-1\right|. \tag{B.27}$$

Then,

$$(I) \leq (II)+(III)+(IV)+(V)+(VI)+(VII), \tag{B.28}$$

where

$$(II) = \max_{d\leq k\leq K_n}\left|\frac{\mathrm{E}(\mathbf{f}_{1,n}^2(d)) - \frac{d(d+1)\sigma^2}{N}}{L_n^d(k)}\right|,$$

$$(III) = \max_{d\leq k\leq K_n}\left|\frac{\mathrm{E}(\mathbf{f}_{2,n}^2(k-d)) - \frac{(k-d)\sigma^2}{N}}{L_n^d(k)}\right|,$$

$$(IV) = \max_{d\leq k\leq K_n}\left|\frac{\mathrm{E}(\mathcal{S}_n^2(k-d)) - \|\mathbf{a}-\mathbf{a}(k-d)\|_z^2}{L_n^d(k)}\right|,$$

$$(V) = \max_{d\leq k\leq K_n}\left|\frac{2\mathrm{E}(\mathbf{f}_{1,n}(d)\mathbf{f}_{2,n}(k-d))}{L_n^d(k)}\right|,$$

$$(VI) = \max_{d\leq k\leq K_n}\left|\frac{2\mathrm{E}(\mathbf{f}_{1,n}(d)\mathcal{S}_n(k-d))}{L_n^d(k)}\right|,$$

$$(VII) = \max_{d\leq k\leq K_n}\left|\frac{2\mathrm{E}(\mathbf{f}_{2,n}(k-d)\mathcal{S}_n(k-d))}{L_n^d(k)}\right|.$$

By Lemma 2, it is easy to see that

$$\lim_{n\to\infty}(II)=0. \tag{B.29}$$

According to Lemma B.3 and an argument used in the proof of Theorem 3 in Ing and Wei (2003),

$$\lim_{n\to\infty}(III)=0, \tag{B.30}$$

and

$$\lim_{n\to\infty}(VII)=0. \tag{B.31}$$

Lemma B.5 and (22) imply

$$(IV)\le \max_{d\le k\le K_n}\left|\frac{w_n\sum_{i=1}^{\infty}(a_i-a_i(k-d))^2+\chi_n}{L_n^d(k)}\right|, \tag{B.32}$$

where $w_n=o(n^{-1})$ and $\chi_n=o(n^{-2})$. Moreover, since (2) yields, for some $C_1,C_2>0$,

$$C_1\|\mathbf{a}-\mathbf{a}(k-d)\|_z^2\le\sum_{i=1}^{\infty}(a_i-a_i(k-d))^2\le C_2\|\mathbf{a}-\mathbf{a}(k-d)\|_z^2, \tag{B.33}$$

one obtains from (B.32) and (B.33) that

$$\lim_{n\to\infty}(IV)=0. \tag{B.34}$$

To show

$$\lim_{n\to\infty}(V)=0 \tag{B.35}$$

and

$$\lim_{n\to\infty}(VI)=0, \tag{B.36}$$

consider

$$\mathbf{f}_{1,n}^*(d)=\left\{\frac{U_{n,n}^{*'}(d)}{\sqrt{N}}\left\{N^{-1}\sum_{j=K_n}^{n-\sqrt{n}-1}U_{j,n}(d)U_{j,n}'(d)\right\}^{-1}\right.$$

$$\left.\times N^{-1/2}\sum_{j=K_n}^{n-\sqrt{n}-1}U_{j,n}(d)\varepsilon_{j+1}\right\}I_{\{d\ge1\}},$$

$$\mathbf{f}_{2,n}^*(k-d)=\left\{\frac{\mathbf{z}_n^{*'}(k-d)}{\sqrt{N}}\Gamma^{-1}(k-d)N^{-1/2}\sum_{j=K_n}^{n-\sqrt{n}-1}\mathbf{z}_j(k-d)\varepsilon_{j+1}\right\}I_{\{k>d\}},$$

and

$$\mathcal{S}_n^*(k-d) = \sum_{i=1}^{\sqrt{n}/2} (a_i - a_i(k-d)) z_{n+1-i}^{**},$$

where

$$U_{n,n}^*(d) = \left( \frac{1}{N^{d-1/2}} \sum_{j=\sqrt{n}}^{n-1} \kappa_j(d)\varepsilon_{n-j}, \ldots, \frac{1}{N^{1/2}} \sum_{j=\sqrt{n}}^{n-1} \kappa_j(1)\varepsilon_{n-j} \right)',$$

$$\mathbf{z}_n^*(l) = \left( \sum_{j=0}^{\sqrt{n}-K_n} b_j \varepsilon_{n-j}, \ldots, \sum_{j=0}^{\sqrt{n}-K_n} b_j \varepsilon_{n-l+1-j} \right)', \; l \geq 1,$$

and $z_{n+1-i}^{**} = \sum_{j=0}^{\sqrt{n}/2} b_j \varepsilon_{n+1-i-j}$. By (2), (3), Minkowski's inequality, Lemma 4 of Berk (1974), the moment restriction imposed on $\{\varepsilon_t\}$, and analogies with (B.1) and (B.12), for any $0 < q \leq 6+\theta_1$,

$$\max_{1 \leq l \leq K_n} l^{-q/2} \mathrm{E} \left\| \mathbf{z}_n(l) - \mathbf{z}_n^*(l) \right\|^q = o(n^{-q/2}), \tag{B.37}$$

$$\mathrm{E} \left\| U_{n,n}(d) - U_{n,n}^*(d) \right\|^q = o(n^{-q/4}), \tag{B.38}$$

and

$$\max_{0 \leq l \leq K_n} \mathrm{E} |\mathcal{S}_n(l) - \mathcal{S}_n^*(l)|^q = o(n^{-q/2}). \tag{B.39}$$

Armed with (B.37)–(B.39), Lemmas 1, B.1, and B.3, Hölder's inequality, and some algebraic manipulations, we obtain

$$\lim_{n\to\infty} \max_{d \leq k \leq K_n} \mathrm{E} \left| \frac{\mathbf{f}_{1,n}(d)\mathbf{f}_{2,n}(k-d) - \mathbf{f}_{1,n}^*(d)\mathbf{f}_{2,n}^*(k-d)}{L_n^d(k)} \right| = 0 \tag{B.40}$$

and

$$\lim_{n\to\infty} \max_{d \leq k \leq K_n} \mathrm{E} \left| \frac{\mathbf{f}_{1,n}(d)\mathcal{S}_n(k-d) - \mathbf{f}_{1,n}^*(d)\mathcal{S}_n^*(k-d)}{L_n^d(k)} \right| = 0. \tag{B.41}$$

As a result, (B.35) and (B.36) follow from (B.40), (B.41), and the facts that for all $d \leq k \leq K_n$, $\mathrm{E}\{\mathbf{f}_{1,n}^*(d)\mathbf{f}_{2,n}^*(k-d)\} = \mathrm{E}\{\mathbf{f}_{1,n}^*(d)\mathcal{S}_n^*(k-d)\} = 0$. Finally, (B.27) is ensured by (B.28)–(B.31) and (B.34)–(B.36). ■

We are now ready to prove Theorem 2.

**Proof of Theorem 2.** We only prove the case of $d \geq 1$, since the proof of the case of $d = 0$ is similar and much simpler. By (A.26)–(A.28), the moment restriction imposed on $\{\varepsilon_t\}$, Lemmas B.1, B.3, B.4, and B.6, and Hölder's inequality, one has for all $d \leq k \leq K_n$ and any $0 < \theta < 1$,

$$
\begin{aligned}
&\mathrm{E}\{\mathbf{f}_n(k) - B_{1n}(k,d) - B_{2n}(k-d)\}^2 \\
&\quad \le \frac{C}{N}\left(\mathrm{E}\|\mathbf{s}_{n,n}(k)\|^{10+\delta_1}\right)^{2/(10+\delta_1)} \Big(\mathrm{E}\|\hat{S}_n^{-1}(K_n)\|^{(40+4\delta_1)/\delta_1} \\
&\qquad\qquad \times \mathrm{E}\|\hat{S}_{d,n}^{-1}(K_n)\|^{(40+4\delta_1)/\delta_1}\Big)^{\delta_1/(20+2\delta_1)} \\
&\qquad \times \Bigg(\mathrm{E}\|\hat{S}_n(k) - \hat{S}_{d,n}(k)\|^{(10+\delta_1)/2} \\
&\qquad\qquad \times \mathrm{E}\left\|\frac{1}{\sqrt{N}}\sum_{j=K_n}^{n-1} \mathbf{s}_{j,n}(k)\varepsilon_{j+1,k-d}\right\|^{(10+\delta_1)/2}\Bigg)^{4/(10+\delta_1)} \\
&\quad \le C\frac{k}{N}\frac{K_n^{4d+3+2\theta}}{N}.
\end{aligned}
$$

Taking $2\theta \le \delta_1$, it follows that

$$
\lim_{n\to\infty}\max_{d\le k\le K_n}\frac{\mathrm{E}\{\mathbf{f}_n(k) - B_{1n}(k,d) - B_{2n}(k-d)\}^2}{L_n^d(k)} = 0, \tag{B.42}
$$

Lemmas 1, B.1, and B.4 ensure that, for all $d \le k \le K_n$,

$$
\mathrm{E}\left\{B_{1n}(k,d) - \mathbf{f}_{1,n}(d)\right\}^2 \le CN^{-1}\left(\sum_{j=k-d+1}^{\infty}|a_i|\right)^2. \tag{B.43}
$$

If $a_i \ne 0$ for infinitely many $i$, then $k_n^*(d) \to \infty$ as $n \to \infty$, where $k_n^*(d)$ is defined after (32). This fact, (2), and (B.43) yield that

$$
\max_{d\le k\le K_n}\frac{\mathrm{E}\left\{B_{1n}(k,d) - \mathbf{f}_{1,n}(d)\right\}^2}{L_n^d(k)} \le C\frac{(\sum_{j=1}^{\infty}|a_i|)^2}{k_n^*(d)} \to 0
$$

as $n \to \infty$. On the other hand, if for some $0 \le k_0 < \infty$, $a_{k_0} \ne 0$ and $a_i = 0$ for all $i > k_0$ (note that $a_0 = 1$), then

$$
\max_{d\le k\le K_n}\frac{\mathrm{E}\left\{B_{1n}(k,d) - \mathbf{f}_{1,n}(d)\right\}^2}{L_n^d(k)} \le \begin{cases} \frac{C}{N\|\mathbf{a}-\mathbf{a}(k_0-1)\|_z^2} & k_0 \ge 1, \\ 0 & k_0 = 0. \end{cases}
$$

As a result,

$$
\lim_{n\to\infty}\max_{d\le k\le K_n}\frac{\mathrm{E}\left\{B_{1n}(k,d) - \mathbf{f}_{1,n}(d)\right\}^2}{L_n^d(k)} = 0. \tag{B.44}
$$

In addition, by Lemmas B.3 and B.6 and (2),

$$
\begin{aligned}
\mathrm{E}\left\{B_{2n}(k-d) - \mathbf{f}_{2,n}(k-d)\right\}^2 &\le \frac{(k-d)^2}{N}(\|\mathbf{a} - \mathbf{a}(k-d)\|_z^2 + N^{-1}) \\
&\le \frac{(k-d)^2}{N}\left(\sum_{j=k-d+1}^{\infty} a_j^2 + N^{-1}\right) \\
&\le \frac{C}{N}.
\end{aligned} \tag{B.45}
$$

Then (B.45) and an argument similar to that used to obtain (B.44) yield

$$\lim_{n\to\infty}\max_{d\le k\le K_n}\frac{\mathrm{E}\left\{B_{2n}(k-d)-\mathbf{f}_{2,n}(k-d)\right\}^2}{L_n^d(k)}=0. \qquad \textbf{(B.46)}$$

Consequently, (29) follows from (B.42), (B.44), (B.46), Lemma B.7, and the Cauchy-Schwarz inequality. ■

**Proof of Theorem 3.** Theorem 3 can be shown by the same argument as in the proof of Theorem 2, except that Theorem 1(ii) is used instead of (A.27) in verifying (B.42). The details are skipped. ■