JOURNAL OF MULTIVARIATE ANALYSIS 16, 393–411 (1985)

Prediction of Multivariate Time Series by Autoregressive Model Fitting

RICHARD LEWIS

Union Carbide Corporation

AND

GREGORY C. REINSEL

University of Wisconsin–Madison

Communicated by E. J. Hannan

Suppose the stationary r-dimensional multivariate time series $\{y_t\}$ is generated by an infinite autoregression. For lead times $h \geqslant 1$, the linear prediction of y_{t+h} based on $y_t, y_{t-1},...$ is considered using an autoregressive model of finite order k fitted to a realization of length T. Assuming that $k \to \infty$ (at some rate) as $T \to \infty$, the consistency and asymptotic normality of the estimated autoregressive coefficients are derived, and an asymptotic approximation to the mean square prediction error based on this autoregressive model fitting approach is obtained. The asymptotic effect of estimating autoregressive parameters is found to inflate the minimum mean square prediction error by a factor of $(1 + kr/T)$.

1. INTRODUCTION

A problem of considerable interest in multivariate time series analysis is the prediction of future values of a stationary multivariate time series $\{y_t, t=0, \pm 1, \pm 2,...\}$, based on a realization $y_1, y_2,..., y_T$ from the process. When the process $\{y_t\}$ is generated by a model with a known parametric structure, such as a finite parameter autoregressive moving average model, estimates of the unknown parameters in the model are used in the prediction of future values. The asymptotic properties of prediction errors using such parametric models with estimated parameters have recently been investigated by several authors, including Akaike [2],

Received October 11, 1982; revised August 16, 1983.

AMS 1970 subject classifications: primary 62M20; secondary 62M10.

Key words and phrases: prediction mean square error, autoregressive model fitting, multivariate time series.

0047-259X/85 $3.00

Bloomfield [8] and Yamamoto [22] in the univariate case, and Baillie [5], Reinsel [15] and Yamamoto [23] in the multivariate case. In the practical case, however, where the precise form of parametric model appropriate for the process $\{y_t\}$ is not known, several authors, such as Parzen [13, 14] and Bhansali [7], have considered the "nonparametric" approach of predicting future values by autoregressive models fitted to the series of T observations, based only on the very mild assumption of an infinite order autoregressive model for the process $\{y_t\}$. This "autoregressive model fitting" approach has also been applied by Akaike [1] and Parzen [14] to the problem of spectral density estimation, with considerable success. In the univariate case, Berk [6] derived the asymptotic distribution of spectral density estimators obtained from fitting autoregressive models of order k to a series of T observations, under the assumption that k increases (at some rate) simultaneously with the realization length T. Bhansali [7] has adopted Berk's [6] approach and applied results of Berk to the problem of prediction of future values in the univariate case.

In this paper we shall develop multivariate generalizations of some of the univariate results of Berk, and apply these to the problem of multivariate prediction. We first derive the asymptotic distribution of estimated autoregressive coefficients, obtained from fitting an autoregressive model of order k to a series of T observations from an infinite order autoregressive process, as k and $T \to \infty$. The asymptotic distribution of corresponding $h \geqslant 1$ step ahead prediction errors based on the fitted autoregressive model of order k is then determined, under the simplifying assumption that the series used for estimation of parameters and the series used for prediction are generated from two independent processes which have the same stochastic structure. Based on this result, an approximation to the h step ahead prediction mean square error matrix is proposed, and a sampling experiment is considered to investigate the accuracy of this approximation in finite samples.

2. The Model and Parameter Estimation

Let $\{y_t, t=0, \pm 1, \pm 2,...\}$ be a vector-valued linear process,

$$y_t = \varepsilon_t + \sum_{j=1}^{\infty} B_j \varepsilon_{t-j}, \tag{2.1}$$

where $y_t = (y_{1t}, y_{2t},..., y_{rt})'$ and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t},..., \varepsilon_{rt})'$ are $(r \times 1)$ random vectors, and $\{\varepsilon_t, t=0, \pm 1, \pm 2,...\}$ is a sequence of independent identically distributed random vectors with mean 0 and positive definite covariance matrix Σ. Defining $\|B_j\|^2 = \mathrm{tr}(B_j' B_j)$ and $B(z) = \sum_{j=0}^{\infty} B_j z^j$, where $B_0 = I_r$, the $(r \times r)$ identity matrix, we assume throughout this paper that $\sum_{j=0}^{\infty} \|B_j\| < \infty$ and $\det\{B(z)\} \neq 0$ for $|z| \leqslant 1$. Under these assumptions we can also express (2.1) as an infinite autoregression,

$$y_t - \sum_{j=1}^{\infty} A_j y_{t-j} = \varepsilon_t, \tag{2.2}$$

where $\sum_{j=1}^{\infty} \|A_j\| < \infty$ and $A(z) = I_r - \sum_{j=1}^{\infty} A_j z^j = B(z)^{-1}$ satisfies $\det\{A(z)\} \neq 0$ for $|z| \leqslant 1$. We note here that the class of stationary invertible ARMA (p, q) models, $y_t - \sum_{j=1}^{p} \Phi_j y_{t-j} = \varepsilon_t - \sum_{j=1}^{q} \Theta_j \varepsilon_{t-j}$, is included in the class of models described above.

Denote the autocovariances of the process $\{y_t\}$ by $\Gamma(j) = E(y_t y_{t+j}')$, $j = 0, \pm 1, \pm 2, \ldots$. Then $\Gamma(-j) = \Gamma(j)'$ and we can also express $\Gamma(j)$ as

$$\Gamma(j) = \int_{-\pi}^{\pi} e^{ij\lambda} f(\lambda)\, d\lambda,$$

where $f(\lambda) = (2\pi)^{-1} A^{-1}(e^{i\lambda}) \Sigma A'^{-1}(e^{-i\lambda}) = (2\pi)^{-1} B(e^{i\lambda}) \Sigma B'(e^{-i\lambda})$, $-\pi \leqslant \lambda \leqslant \pi$, is the spectral density matrix of the process $\{y_t\}$.

The minimum mean square error linear predictor of y_{t+1} based on $y_t, y_{t-1}, \ldots, y_{t-k+1}$ is given by

$$y_{t,k}^*(1) = A_{1k} y_t + A_{2k} y_{t-1} + \cdots + A_{kk} y_{t-k+1}, \tag{2.3}$$

where the A_{jk}, $j = 1, \ldots, k$, satisfy the "Yule–Walker" equations

$$(A_{1k}, A_{2k}, \ldots, A_{kk}) = \Gamma_{1,k}' \Gamma_k^{-1},$$

where $\Gamma_{1,k}' = (\Gamma(1)', \Gamma(2)', \ldots, \Gamma(k)')$, and Γ_k is a $kr \times kr$ matrix whose (m, n)th $(r \times r)$ block of elements is $\Gamma(m-n)$, $m, n = 1, \ldots, k$. We let $\Sigma_k = E[(y_{t+1} - y_{t,k}^*(1))(y_{t+1} - y_{t,k}^*(1))']$ denote the mean square error of the predictor $y_{t,k}^*(1)$.

Based on a realization $y_1, y_2, \ldots, y_T$ of length T, the A_{jk}, $j = 1, \ldots, k$, are estimated by

$$\hat{A}(k) = (\hat{A}_{1k}, \hat{A}_{2k}, \ldots, \hat{A}_{kk}) = \hat{\Gamma}_{1,k}' \hat{\Gamma}_k^{-1}, \tag{2.4}$$

where $\hat{\Gamma}_{1,k} = (T-k)^{-1} \sum_{t=k}^{T-1} Y_{t,k} y_{t+1}'$, $\hat{\Gamma}_k = (T-k)^{-1} \sum_{t=k}^{T-1} Y_{t,k} Y_{t,k}'$, and $Y_{t,k} = (y_t', y_{t-1}', \ldots, y_{t-k+1}')'$. We also estimate Σ_k by

$$\hat{\Sigma}_k = (T-k)^{-1} \sum_{t=k}^{T-1} (y_{t+1} - \hat{y}_{t,k}(1))(y_{t+1} - \hat{y}_{t,k}(1))',$$

where

$$\hat{y}_{t,k}(1) = \sum_{j=1}^{k} \hat{A}_{jk} y_{t-j+1} \tag{2.5}$$

is an estimate of $y^*_{t,k}(1)$.

We are interested in the asymptotic behavior of $y_{t+1} - \hat{y}_{t,k}(1)$ (as well as the asymptotic behavior of general h step ahead prediction errors) as $T \to \infty$. However, since $\{y_t\}$ is expressible as the infinite autoregression (2.2), to obtain predictors which are asymptotically equivalent to the optimal predictor we shall let $k \to \infty$ (at some rate) as $T \to \infty$. In order to determine the asymptotic behavior of $y_{t+1} - \hat{y}_{t,k}(1)$ as k and $T \to \infty$ we will first establish the asymptotic properties of $\hat{A}(k) - A(k)$ as k and $T \to \infty$, where $A(k) = (A_1, A_2, ..., A_k)$.

For later convenience we now introduce the vec operator. That is, for any $(m \times n)$ matrix C define $\operatorname{vec}(C)$ to be the $(mn \times 1)$ column vector formed by stacking the columns of C below one another. A useful property of the vec operator is that $\operatorname{vec}(ABC) = (C' \otimes A) \operatorname{vec}(B)$, where A, B and C are conformable, and $\otimes$ denotes the Kronecker product. We also introduce the matrix norm $\|C\|_1^2 = \sup_{l \neq 0} l'C'Cl/l'l$, the largest eigenvalue of $C'C$. (If C is symmetric, then $\|C\|_1^2$ is the square of the largest, in absolute value, eigenvalue of C.) A useful inequality relating $\|\cdot\|^2$ and $\|\cdot\|_1^2$ is [21, p. 96]

$$\|AB\|^2 \leqslant \|A\|_1^2 \, \|B\|^2, \qquad \text{as well as} \quad \|AB\|^2 \leqslant \|A\|^2 \, \|B\|_1^2. \tag{2.6}$$

We are now ready to present our main results concerning the asymptotic behavior of $\hat{A}(k) - A(k)$. First note, however, that

$$\hat{A}(k) - A(k) = \hat{\Gamma}'_{1,k} \hat{\Gamma}_k^{-1} - A(k) \hat{\Gamma}_k \hat{\Gamma}_k^{-1} = \left\{ (T-k)^{-1} \sum_{t=k}^{T-1} \varepsilon_{t+1,k} Y'_{t,k} \right\} \hat{\Gamma}_k^{-1}, \tag{2.7}$$

where $\varepsilon_{t,k} = y_t - \sum_{j=1}^{k} A_j y_{t-j}$. In Theorem 1 we establish the "consistency" of $\hat{A}(k)$, while in Theorems 2, 3 and 4 we derive the asymptotic normality of $\hat{\alpha}(k) - \alpha(k) = \operatorname{vec}\{\hat{A}(k) - A(k)\}$.

THEOREM 1. *Let $\{y_t\}$ satisfy (2.1), and assume that*

(i) $E\,|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}| \leqslant \gamma_4 < \infty$, $1 \leqslant i, j, k, l \leqslant r$;

(ii) *k is chosen as a function of T such that $k^2/T \to 0$ as $k, T \to \infty$;*

(iii) *k is chosen as a function of T such that $k^{1/2} \sum_{j=k+1}^{\infty} \|A_j\| \to 0$ as $k, T \to \infty$.*

Then

$$\|\hat{A}(k) - A(k)\| \xrightarrow{p} 0 \qquad \text{as} \quad T \to \infty.$$

Proof. From (2.7) and using (2.6), we have

$$\|\hat{A}(k) - A(k)\| \leqslant \|\hat{\Gamma}_k^{-1}\|_1 \, \|U_{1T}\| + \|\hat{\Gamma}_k^{-1}\|_1 \, \|U_{2T}\|, \tag{2.8}$$

where

$$U_{1T} = (T-k)^{-1} \sum_{t=k}^{T-1} (\varepsilon_{t+1,k} - \varepsilon_{t+1}) \, Y'_{t,k}, \qquad U_{2T} = (T-k)^{-1} \sum_{t=k}^{T-1} \varepsilon_{t+1} Y'_{t,k}.$$

Also, $\|\hat{\Gamma}_k^{-1}\|_1 \leqslant \|\Gamma_k^{-1}\|_1 + \|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1$ where, as in the univariate case [6, p. 491], $\|\Gamma_k^{-1}\|_1$ is uniformly bounded above by a positive constant F for all k, and we now indicate that $\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1 \to^p 0$ as $T \to \infty$ under assumption (ii). First, from Hannan [11, Chap. 4] we can establish that $\|\hat{\Gamma}_k - \Gamma_k\|_1 \to^p 0$ since $E(\|\hat{\Gamma}_k - \Gamma_k\|_1^2) \leqslant E(\|\hat{\Gamma}_k - \Gamma_k\|^2) \leqslant \text{constant} \cdot (kr)^2/(T-k) \to 0$ as $T \to \infty$ by assumption (ii). Then from

$$\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1 = \|\hat{\Gamma}_k^{-1}(\hat{\Gamma}_k - \Gamma_k) \, \Gamma_k^{-1}\|_1 \leqslant F(\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1 + F) \, \|\hat{\Gamma}_k - \Gamma_k\|_1,$$

we have $0 \leqslant Z_{k,T} = \|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1 / F(\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1 + F) \leqslant \|\hat{\Gamma}_k - \Gamma_k\|_1$, so that $Z_{k,T} \to^p 0$ as $T \to \infty$. Hence it follows [12] that $\|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1 = F^2 Z_{k,T}/(1 - F Z_{k,T})$ also converges in probability to zero as $T \to \infty$. Now we also have

$$\begin{aligned}
E(\|U_{1T}\|) &\leqslant (T-k)^{-1} \sum_{t=k}^{T-1} E(\|(\varepsilon_{t+1,k} - \varepsilon_{t+1}) \, Y'_{t,k}\|) \\
&\leqslant \{E(\|Y_{t,k}\|^2) \, E(\|(\varepsilon_{t+1,k} - \varepsilon_{t+1})\|^2)\}^{1/2} \\
&= \{k \operatorname{tr}(\Gamma(0))\}^{1/2} \left\{E\left(\left\| \sum_{j=k+1}^{\infty} A_j y_{t-j+1} \right\|^2\right)\right\}^{1/2} \\
&\leqslant \{k \operatorname{tr}(\Gamma(0))\}^{1/2} \left\{ \sum_{i=k+1}^{\infty} \sum_{j=k+1}^{\infty} \|\Gamma(i-j)\| \, \|A_i\| \, \|A_j\| \right\}^{1/2} \\
&\leqslant \text{constant} \cdot k^{1/2} \sum_{j=k+1}^{\infty} \|A_j\|,
\end{aligned} \tag{2.9}$$

noting that the $\|\Gamma(i-j)\|$ are uniformly bounded since $\sum_{j=-\infty}^{\infty} \|\Gamma(j)\| < \infty$, which follows from the condition $\sum_{j=0}^{\infty} \|B_j\| < \infty$. Thus by assumption (iii), $\|U_{1T}\| \to^p 0$, and it follows that the first term on the right side of (2.8) converges to zero in probability. Finally, since ε_{t+1} and $Y_{t,k}$ are independent,

$$\begin{aligned}
E(\|U_{2T}\|^2) &= (T-k)^{-2} \sum_{t=k}^{T-1} E(\varepsilon'_{t+1} \varepsilon_{t+1}) \, E(Y'_{t,k} Y_{t,k}) \\
&= k(T-k)^{-1} \operatorname{tr}(\Sigma) \operatorname{tr}\{\Gamma(0)\} \to 0 \qquad \text{as} \quad T \to \infty
\end{aligned} \tag{2.10}$$

by assumption (ii). Therefore $\|U_{2T}\| \to^p 0$ as $T \to \infty$ and it follows that the second term on the right side of (2.8) also converges to zero in probability.

In the next theorem we show that $\varepsilon_{t+1,k}$ and $\hat{\Gamma}_k^{-1}$ in expression (2.7) for $\hat{A}(k) - A(k)$ may be replaced by ε_{t+1} and Γ_k^{-1} when obtaining the asymptotic distribution of $\hat{\alpha}(k) - \alpha(k) = \operatorname{vec}\{\hat{A}(k) - A(k)\}$. In Theorem 3 the asymptotic distribution of the asymptotically equivalent expression for $\hat{\alpha}(k) - \alpha(k)$ is then derived.

THEOREM 2. *Let* $\{y_t\}$ *satisfy* (2.1), *and assume that*

(i) $E|\varepsilon_{it}\varepsilon_{jt}\varepsilon_{kt}\varepsilon_{lt}| \leqslant \gamma_4 < \infty$, $1 \leqslant i, j, k, l \leqslant r$;

(ii) *k is chosen as a function of T such that* $k^3/T \to 0$ *as* $k, T \to \infty$;

(iii) *k is chosen as a function of T such that*

$$T^{1/2} \sum_{j=k+1}^{\infty} \|A_j\| \to 0 \qquad \text{as} \quad k, T \to \infty;$$

(iv) $\{l(k)\}$ *is a sequence of* $(kr^2 \times 1)$ *vectors such that*

$$0 < M_1 \leqslant \|l(k)\|^2 = l(k)'\, l(k) \leqslant M_2 < \infty \qquad \text{for} \quad k = 1, 2, \ldots .$$

Then

$$(T-k)^{1/2}\, l(k)'(\hat{\alpha}(k) - \alpha(k)) - (T-k)^{1/2}\, l(k)' \operatorname{vec}\left[\left\{(T-k)^{-1} \sum_{t=k}^{T-1} \varepsilon_{t+1} Y'_{t,k}\right\} \Gamma_k^{-1}\right]$$

converges in probability to zero as $T \to \infty$.

Proof. From (2.7) we have that

$$\begin{aligned}
&(T-k)^{1/2}\, l(k)'(\hat{\alpha}(k) - \alpha(k)) \\
&\quad - (T-k)^{1/2}\, l(k)' \operatorname{vec}\left[\left\{(T-k)^{-1} \sum_{t=k}^{T-1} \varepsilon_{t+1} Y'_{t,k}\right\} \Gamma_k^{-1}\right] \\
&= l(k)'\{((\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \otimes I_r) \operatorname{vec}[(T-k)^{1/2}\, U_{1T}] \\
&\quad + ((\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}) \otimes I_r) \operatorname{vec}[(T-k)^{1/2}\, U_{2T}] \\
&\quad + (\Gamma_k^{-1} \otimes I_r) \operatorname{vec}[(T-k)^{1/2}\, U_{1T}]\} \\
&= w_{1T} + w_{2T} + w_{3T},
\end{aligned}$$

where U_{1T} and U_{2T} are as defined in the proof of Theorem 1, and w_{1T}, w_{2T} and w_{3T} are defined in an obvious manner. Using (2.6) we have

$$|w_{1T}| \leqslant \|l(k)\|\, k^{1/2}\, \|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1\, \|k^{-1/2}(T-k)^{1/2}\, U_{1T}\|,$$

where $\|l(k)\| \leqslant M_2^{1/2}$, and, similar to the argument following (2.8), it can be shown that $k^{1/2} \|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1 \to^p 0$ under assumption (ii). Also, by (2.9),

$$E(\|k^{-1/2}(T-k)^{1/2} U_{1T}\|) \leqslant \text{constant} \cdot (T-k)^{1/2} \sum_{j=k+1}^{\infty} \|A_j\| \to 0 \quad \text{as } T \to \infty$$

by assumption (iii), and thus $w_{1T} \to^p 0$ as $T \to \infty$. Similarly,

$$|w_{2T}| \leqslant \|l(k)\| \, k^{1/2} \, \|\hat{\Gamma}_k^{-1} - \Gamma_k^{-1}\|_1 \, \|k^{-1/2}(T-k)^{1/2} U_{2T}\|,$$

where $\|k^{-1/2}(T-k)^{1/2} U_{2T}\|$ is bounded in probability, by (2.10). Thus w_{2T} also converges to zero in probability as $T \to \infty$. Finally,

$$\begin{aligned} |w_{3T}| &= \left| (T-k)^{-1/2} \sum_{t=k}^{T-1} l(k)'(\Gamma_k^{-1} Y_{t,k} \otimes I_r)(\varepsilon_{t+1,k} - \varepsilon_{t+1}) \right| \\ &\leqslant (T-k)^{-1/2} \, r^{1/2} \left\| \sum_{t=k}^{T-1} (\varepsilon_{t+1,k} - \varepsilon_{t+1}) \, l(k)'(\Gamma_k^{-1} Y_{t,k} \otimes I_r) \right\|, \end{aligned}$$

so that similar to the result in (2.9), defining $v'_{t,k} = l(k)'(\Gamma_k^{-1} Y_{t,k} \otimes I_r)$, we have

$$E\,|w_{3T}| \leqslant \text{constant} \cdot \{E(\|v_{t,k}\|^2)\}^{1/2} (T-k)^{1/2} \sum_{j=k+1}^{\infty} \|A_j\| \to 0 \qquad \text{as} \quad T \to \infty$$

by assumption (iii), since $E(\|v_{t,k}\|^2) = l(k)'(\Gamma_k^{-1} \otimes I_r) \, l(k) \leqslant \|\Gamma_k^{-1}\|_1 M_2$ is uniformly bounded under assumption (iv). Therefore $w_{3T} \to^p 0$ as $T \to \infty$.

THEOREM 3. *Let* $\{y_t\}$ *satisfy* (2.1), *and assume that conditions* (i)–(ii) *and* (iv) *of Theorem* 2 *hold. Also let*

$$s_T = (T-k)^{1/2} \, l(k)' \operatorname{vec}\left[\left\{ (T-k)^{-1} \sum_{t=k}^{T-1} \varepsilon_{t+1} Y'_{t,k} \right\} \Gamma_k^{-1} \right]$$

with

$$v_T^2 = \operatorname{Var}(s_T) = l(k)'(\Gamma_k^{-1} \otimes \Sigma) \, l(k).$$

Then

$$s_T / v_T \xrightarrow{d} N(0, 1) \qquad \textit{as} \quad T \to \infty.$$

Proof. First let $y_t(m) = \sum_{j=0}^{m} B_j \varepsilon_{t-j}$, and

$$s_{Tm} = (T-k)^{1/2} \, l(k)' \operatorname{vec}\left[\left\{ (T-k)^{-1} \sum_{t=k}^{T-1} \varepsilon_{t+1} Y'_{t,k}(m) \right\} \Gamma_k^{-1} \right], \qquad m = 1, 2, \ldots,$$

where $Y_{t,k}(m) = (y_t'(m), y_{t-1}'(m), \ldots, y_{t-k+1}'(m))'$. Then set

$$s_T/v_T = s_{Tm}/v_{Tm} + (s_T/v_T - s_{Tm}/v_{Tm}) = s_{Tm}/v_{Tm} + z_{Tm},$$

where

$$v_{Tm}^2 = \operatorname{Var}(s_{Tm}) = l(k)'(\Gamma_k^{-1}\Gamma_k(m)\,\Gamma_k^{-1} \otimes \Sigma)\,l(k),$$

and

$$\Gamma_k(m) = E(Y_{t,k}(m)\,Y_{t,k}'(m)).$$

By Anderson [4, Corollary 7.7.1], s_T/v_T converges in distribution to $N(0, 1)$ as $T \to \infty$ since we can establish that

(a) $E(z_{Tm}^2) \leqslant M_m < \infty$ for all $T \geqslant 1$, where $\lim_{m \to \infty} M_m = 0$, and

(b) $s_{Tm}/v_{Tm} \to^d N(0, 1)$ as $T \to \infty$, for each $m \geqslant 1$.

To verify that condition (a) holds, we set $z_{Tm} = z_{1Tm} + z_{2Tm}$, where $z_{1Tm} = (s_{Tm}/v_{Tm})(v_{Tm}/v_T - 1)$ and $z_{2Tm} = (s_T - s_{Tm})/v_T$. Now we have $E(z_{1Tm}^2) = (v_{Tm}/v_T - 1)^2$, and $v_{Tm}^2/v_T^2 = 1 - [l'((\Gamma_k - \Gamma_k(m)) \otimes \Sigma)\,l/l'l]/[l'(\Gamma_k \otimes \Sigma)\,l/l'l]$, where $l = (\Gamma_k^{-1} \otimes I_r)\,l(k)$, and $l'(\Gamma_k \otimes \Sigma)\,l/l'l$ is uniformly bounded below by a positive constant [6, p. 491]. That condition (a) is satisfied for z_{1Tm} then follows from the fact that $\|f(\lambda) - f_m(\lambda)\|^2$ converges to zero uniformly in λ as $m \to \infty$, where $f_m(\lambda) = (2\pi)^{-1}(\sum_{j=0}^m B_j e^{i\lambda j})\,\Sigma(\sum_{j=0}^m B_j' e^{-i\lambda j})$ is the spectral density matrix of $\{y_t(m)\}$. Thus, defining $h = (e^{i\lambda}, e^{i2\lambda}, \ldots, e^{ik\lambda})'$, it follows that

$$|l'((\Gamma_k - \Gamma_k(m)) \otimes \Sigma)\,l/l'l| \leqslant \int_{-\pi}^{\pi} |l'(h\bar{h}' \otimes (f(\lambda) - f_m(\lambda)) \otimes \Sigma)\,l|\,d\lambda/l'l \to 0 \tag{2.11}$$

uniformly in k as $m \to \infty$, and hence $\lim_{m \to \infty} v_{Tm}^2/v_T^2 = 1$ uniformly in T and condition (a) holds for z_{1Tm}. In a similar way condition (a) can be established for $z_{2Tm} = (s_T - s_{Tm})/v_T$ based on the fact that the spectral density matrix of the process $\{y_t - y_t(m)\}$ also converges to zero uniformly in λ as $m \to \infty$.

To verify condition (b), for each $m = 1, 2, \ldots$, we can write $s_{Tm}/v_{Tm} = \sum_{t=1}^T X_t(T)$, where

$$X_t(T) = (T-k)^{-1/2}\,l(k)'(\Gamma_k^{-1} Y_{t-1,k}(m) \otimes I_r)\,\varepsilon_t/v_{Tm}, \qquad \text{for} \quad k+1 \leqslant t \leqslant T,$$

and $X_t(T) = 0$ for $1 \leqslant t \leqslant k$. We note that $X_s(T)$ and $X_t(T)$ are uncorrelated for $s \neq t$, with $E(X_t(T)) = 0$ and $\operatorname{Var}(X_t(T)) = (T-k)^{-1}$ for $k+1 \leqslant t \leqslant T$. By Theorem 9.1.5 of Chung [9], $\{S_n(T) = \sum_{t=1}^n X_t(T), 0 \leqslant n \leqslant T\}$, where

$S_0(T)=0$ a.e. is a martingale sequence for each $T \geqslant 1$. Then since we can establish the following conditions,

(c) $\sup_{t \leqslant T} X_t^2(T) \to^p 0$ as $T \to \infty$, and

(d) $\sum_{t=1}^{n_T(\tau)} X_t^2(T) \to^p \tau$, $0<\tau \leqslant 1$, as $T \to \infty$,

where $n_T(\tau)=\max_{n \leqslant T}\{n: E[(\sum_{t=1}^n X_t(T))^2] \leqslant \tau\}=[\tau(T-k)+k]$, and $[x]$ denotes the greatest integer $\leqslant x$, it follows that $s_{Tm}/v_{Tm} \to^d N(0,1)$ as $T \to \infty$ by Theorem 2 of Scott [17]. Condition (c) holds since for any $\delta>0$, using (2.6) and the independence of ε_t and $Y_{t-1,k}(m)$, we have

$$\begin{aligned} P(\sup_{t \leqslant T} X_t^2(T) \geqslant \delta) &\leqslant \sum_{t=k+1}^{T} P(X_t^2(T) \geqslant \delta) \leqslant \delta^{-2}(T-k)\, E(X_t^4(T)) \\ &\leqslant \delta^{-2}(T-k)^{-1}\, v_{Tm}^{-4}\, \|l(k)\|^4\, \|\Gamma_k^{-1}\|_1^4\, E(\|\varepsilon_t\|^4)\, E(\|Y_{t-1,k}(m)\|^4) \\ &\leqslant \text{constant} \cdot (T-k)^{-1}\, k^2 E[(y_t'(m)\, y_t(m))^2] \\ &\leqslant \text{constant} \cdot k^2/(T-k) \to 0 \end{aligned} \tag{2.12}$$

as $T \to \infty$ by assumption (ii). For condition (d), we have

$$\begin{aligned} E\left(\sum_{t=1}^{n_T(\tau)} X_t^2/(T)\right) &= (n_T(\tau)-k)\, E(X_t^2(T)) \\ &= ([\tau(T-k)+k]-k)/(T-k) \to \tau \qquad \text{as} \quad T \to \infty, \end{aligned}$$

and

$$\operatorname{Var}\left(\sum_{t=1}^{n_T(\tau)} X_t^2(T)\right) \leqslant \text{constant} \cdot k^3/(T-k) \to 0 \qquad \text{as} \quad T \to \infty$$

by assumption (ii), since $X_s^2(T)$ and $X_t^2(T)$ are independent for $|s-t|>k+m$, and from (2.12), $\operatorname{Var}(X_t^2(T)) \leqslant E(X_t^4(T)) \leqslant \text{constant} \cdot k^2/(T-k)^2$. We thus have that condition (d) holds, and the theorem is established.

The following is an immediate consequence of Theorems 2 and 3.

THEOREM 4. *Let $\{y_t\}$ satisfy (2.1), and assume that conditions (i)–(iv) of Theorem 2 hold. Then*

$$(T-k)^{1/2}\, l(k)'(\hat{\alpha}(k)-\alpha(k))/v_T \xrightarrow{d} N(0,1) \qquad \textit{as} \quad T \to \infty,$$

where $v_T^2 = l(k)'(\Gamma_k^{-1} \otimes \Sigma)\, l(k)$.

In particular, let us define the infinite-dimensional matrix $\Gamma_\infty = \{\Gamma(m-n)\}$, $m, n=1,2,...,$ and let Γ_∞^{mn} denote the (m,n)th $(r \times r)$ block element of Γ_∞^{-1} which can be shown to be given by

$$\Gamma_\infty^{mn} = \sum_{j=0}^{m-1} A_j' \Sigma^{-1} A_{j+n-m} = \sum_{j=0}^{n-1} A_{j+m-n}' \Sigma^{-1} A_j, \qquad m, n = 1, 2, ...,$$

where we define $A_j = 0$ for $j < 0$, and $A_0 = -I_r$. Then for any fixed integer k_0, we have that the asymptotic distribution of $(T-k)^{1/2}\,\mathrm{vec}[(\hat{A}_{1k}, \hat{A}_{2k},..., \hat{A}_{k_0k}) - (A_1, A_2,..., A_{k_0})]$ is multivariate normal with mean vector 0 and covariance matrix $V \otimes \Sigma$, where $V = \{V_{mn}\} = \{\Gamma_\infty^{mn}\}$, $m, n = 1,..., k_0$, is the upper left $(k_0 r^2 \times k_0 r^2)$ corner of Γ_∞^{-1}. This result follows directly from Theorem 4 based on the fact that the (m, n)th $(r \times r)$ block element of Γ_k^{-1} is given by

$$\Gamma_k^{mn} = \sum_{j=0}^{m-1} A'_{j,j+k-m} \Sigma_{j+k-m}^{-1} A_{j+n-m,j+k-m}, \qquad m, n = 1,..., k,$$

where, under the conditions of the theorem, the A_{jk} converge to A_j and Σ_k^{-1} converges to Σ^{-1}, pointwise, as $k \to \infty$, and hence Γ_k^{mn} converges to Γ_∞^{mn} as $T \to \infty$, for $m, n = 1,..., k_0$.

3. Asymptotic Properties of Predictors Based on Autoregressive Model Fitting

Let $\{y_t\}$ be a stochastic process generated by (2.1), and let $\{x_t\}$ be a process which is independent of $\{y_t\}$, but has the same stochastic structure. We want to study the asymptotic properties of one-step ahead predictors of the form (2.5), as well as related multistep predictors. We assume the estimates $\hat{A}(k) = (\hat{A}_{1k},..., \hat{A}_{kk})$ have been obtained as in (2.4), but, as is commonly assumed in studies of this type (see, for example, [22, 7]), using a realization $x_1, x_2,..., x_T$ from the independent process $\{x_t\}$. This problem has previously been considered by Bhansali [7] for the univariate case.

We define the $(kr \times kr)$ matrix $A_{(k)}$ and the infinite-dimensional matrix $A_{(\infty)}$ by

$$A_{(k)} = \begin{bmatrix} A_1 & A_2 & \cdots & & A_k \\ I_r & 0 & \cdots & & 0 \\ 0 & I_r & \cdots & & 0 \\ \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & I_r & 0 \end{bmatrix}, \quad \text{and} \quad A_{(\infty)} = \begin{bmatrix} A_1 & A_2 & \cdots & A_k & \cdots \\ I_r & 0 & \cdots & 0 & \cdots \\ 0 & I_r & \cdots & 0 & \cdots \\ \vdots & \vdots & & \vdots & \end{bmatrix}.$$

We also define the $(kr \times r)$ matrix $E_k = (I_r, 0,..., 0)'$, the matrix $E_\infty = (I_r, 0, 0,...)'$ and the infinite-dimensional vector $Y_{t,\infty} = (y'_t, y'_{t-1},...)'$. Then similar to the development in the finite order multivariate autoregressive case (see [15, p. 329]), it follows that an explicit expression for the minimum mean square error predictor of the future value y_{t+h} based on y_t, y_{t-1},... is given by

$$y_t^*(h) = E'_\infty A_{(\infty)}^h Y_{t,\infty}, \tag{3.1}$$

with prediction error equal to $y_{t+h}-y_t^*(h)=\sum_{j=0}^{h-1} B_j\varepsilon_{t+h-j}$ and associated mean square prediction error matrix

$$\Sigma(h)=E[(y_{t+h}-y_t^*(h))(y_{t+h}-y_t^*(h))']=\sum_{j=0}^{h-1} B_j\Sigma B_j'.$$

As mentioned in Section 2, an estimate of the minimum mean square error linear predictor of y_{t+1} given $y_t, y_{t-1},..., y_{t-k+1}$ is $\hat{y}_{t,k}(1)=\sum_{j=1}^{k}\hat{A}_{jk}\,y_{t-j+1}=E_k'\hat{A}_{(k)}Y_{t,k}$, where $\hat{A}_{(k)}$ is a matrix of the same form as $A_{(k)}$, but with the estimates $\hat{A}_{jk}$ in place of the A_j, $j=1,..., k$. Generalizing this to h step ahead prediction yields the predictor

$$\hat{y}_{t,k}(h)=E_k'\hat{A}_{(k)}^h Y_{t,k}, \qquad h=1, 2,... .$$

We note here that if $\{y_t\}$ were a kth order autoregressive process, then $y_{t,k}(h)=E_k'A_{(k)}^h Y_{t,k}$ would be the minimum mean square error predictor of y_{t+h} given y_t, y_{t-1},... . Thus $\hat{y}_{t,k}(h)$ is the "natural" predictor of y_{t+h} based on fitting an autoregressive model of order k. We are interested in determining an asymptotic approximation, as k and $T\to\infty$, to the mean square error matrix of the predictor $\hat{y}_{t,k}(h)$,

$$\Sigma_k(h)=E[(y_{t+h}-\hat{y}_{t,k}(h))(y_{t+h}-\hat{y}_{t,k}(h))'].$$

To obtain an approximation to $\Sigma_k(h)$, we first note that

$$y_{t+h}-\hat{y}_{t,k}(h)=\sum_{j=0}^{h-1} B_j\varepsilon_{t+h-j}-(\hat{y}_{t,k}(h)-y_t^*(h)),$$

where the two terms on the right side are independent since the ε_{t+h-j}, $0\leqslant j\leqslant h-1$, are independent of $Y_{t,\infty}$ and $\hat{A}(k)$. Thus

$$\Sigma_k(h)=\Sigma(h)+E[(\hat{y}_{t,k}(h)-y_t^*(h))(\hat{y}_{t,k}(h)-y_t^*(h))']. \tag{3.2}$$

For the second term on the right side of (3.2) we have

$$\begin{aligned}\hat{y}_{t,k}(h)-y_t^*(h)&=(\hat{y}_{t,k}(h)-y_{t,k}(h))-(y_t^*(h)-y_{t,k}(h))\\&=E_k'(\hat{A}_{(k)}^h-A_{(k)}^h)\,Y_{t,k}-(E_\infty' A_{(\infty)}^h Y_{t,\infty}-E_k'A_{(k)}^h Y_{t,k}).\end{aligned} \tag{3.3}$$

Now, assuming that k is chosen as a function of T such that $T^{1/2}\sum_{j=k+1}^{\infty}\|A_j\|\to 0$ as k, $T\to\infty$, it follows that

$$(T/k)^{1/2}(E_\infty' A_{(\infty)}^h Y_{t,\infty}-E_k'A_{(k)}^h Y_{t,k})\to 0 \qquad \text{in mean square as } k,\ T\to\infty. \tag{3.4}$$

For example, considering the case $h=1$, from the derivation in (2.9) we have

$$E(\|(T/k)^{1/2}(E'_{\infty}A_{(\infty)}Y_{t,\infty}-E'_kA_{(k)}Y_{t,\infty})\|^2)$$
$$=E\left(\left\|(T/k)^{1/2}\sum_{j=k+1}^{\infty}A_jy_{t-j+1}\right\|^2\right)$$
$$\leqslant \text{constant}\cdot\left(T^{1/2}\sum_{j=k+1}^{\infty}\|A_j\|\right)^2\to 0 \qquad \text{as} \quad k, T\to\infty.$$

Also, for the first term on the right side of (3.3) we have the following asymptotic result.

THEOREM 5. *Let $\{y_t\}$ satisfy* (2.1), *and assume that conditions* (i)–(iii) *of Theorem* 2 *are satisfied. Also let $\{x_t\}$ be a stochastic process which is independent of $\{y_t\}$, but has the same stochastic structure. Then*

$$(T/k)^{1/2}E'_k(\hat{A}^h_{(k)}-A^h_{(k)})Y_{t,k}\xrightarrow{d}N(0,r\Sigma(h)) \qquad as \quad T\to\infty,$$

where the estimates $\hat{A}(k)=(\hat{A}_{1k},..., \hat{A}_{kk})$ have been obtained as in (2.4), *but using a realization $x_1,..., x_T$ from the process $\{x_t\}$.*

Proof. Noting that $\hat{A}^h_{(k)}-A^h_{(k)}=\sum_{j=0}^{h-1}\hat{A}^j_{(k)}(\hat{A}_{(k)}-A_{(k)})A^{h-j-1}_{(k)}$ [16] and that $\hat{A}_{(k)}-A_{(k)}=E_k(\hat{A}(k)-A(k))$, we have

$$(T/k)^{1/2}E'_k(\hat{A}^h_{(k)}-A^h_{(k)})Y_{t,k}$$
$$=(T/k)^{1/2}\sum_{j=0}^{h-1}(Y'_{t,k}A'^{h-j-1}_{(k)}\otimes B_j)(\hat{\alpha}(k)-\alpha(k))$$
$$+(T/k)^{1/2}\sum_{j=0}^{h-1}(Y'_{t,k}A'^{h-j-1}_{(k)}\otimes(\hat{B}_{jk}-B_j))(\hat{\alpha}(k)-\alpha(k)), \quad (3.5)$$

where $\hat{B}_{jk}=E'_k\hat{A}^j_{(k)}E_k$.

Now let $l=(l_1,..., l_r)'$ be an arbitrary $(r\times 1)$ vector and define

$$l(k)'=l'k^{-1/2}\sum_{j=0}^{h-1}(Y'_{t,k}A'^{h-j-1}_{(k)}\otimes B_j)=\sum_{j=0}^{h-1}(k^{-1/2}Y'_{t,k}A'^{h-j-1}_{(k)}\otimes l'B_j).$$

Then an arbitrary linear combination of the first term on the right side of (3.5) can be expressed as $T^{1/2}l(k)'(\hat{\alpha}(k)-\alpha(k))$, which is in a form to which Theorem 4 may be applied, conditional on the values $Y_{t,\infty}=(y'_t, y'_{t-1},...)'$. To verify that condition (iv) of Theorem 4 is satisfied for the above random sequence $l(k)$, we note that

$$\|l(k)\|^2 = \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} (k^{-1} Y'_{t,k} A'^{h-i-1}_{(k)} A^{h-j-1}_{(k)} Y_{t,k})(l' B_i B'_j l)$$

$$\xrightarrow{\text{a.s.}} \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} \operatorname{tr}\{\Gamma(i-j)\}\, l' B_i B'_j l$$

$$= E\left(\left\| \sum_{j=0}^{h-1} (y_{t-j} \otimes l' B_j) \right\|^2\right) > 0 \qquad \text{as} \quad k \to \infty,$$

since $k^{-1} Y'_{t+h-i-1,k} Y_{t+h-j-1,k} \to^{\text{a.s.}} \operatorname{tr}\{\Gamma(i-j)\}$ as $k \to \infty$ [11, p. 210], and

$$k^{-1} Y'_{t,k} A'^{h-i-1}_{(k)} A^{h-j-1}_{(k)} Y_{t,k} - k^{-1} Y'_{t+h-i-1,k} Y_{t+h-j-1,k} \xrightarrow{\text{a.s.}} 0 \qquad \text{as} \quad k \to \infty,$$

where we note that, with $m = h - j - 1$, $A^m_{(k)} Y_{t,k} = (y'_{t,k}(m),..., y'_{t,k}(1), y'_t,..., y'_{t-k+m+1})'$, so that $A^m_{(k)} Y_{t,k} - Y_{t+m,k} = (y'_{t,k}(m) - y'_{t+m},..., y'_{t,k}(1) - y'_{t+1}, 0',..., 0')'$ consists of only $m \leqslant h-1$ nonzero terms. Thus, since $\|l(k)\|^2$ converges almost surely to a positive constant as $k \to \infty$, conditional on $Y_{t,\infty} = (y'_t, y'_{t-1},...)'$, we have by Theorem 4 that

$$\frac{T^{1/2} l(k)'(\hat{\alpha}(k) - \alpha(k))}{l(k)'(\Gamma_k^{-1} \otimes \Sigma)\, l(k)} \xrightarrow{d} N(0, 1) \qquad \text{as} \quad T \to \infty \tag{3.6}$$

for almost all realizations of the process $\{y_t\}$. Since the above limiting distribution does not depend on $Y_{t,\infty}$ it follows that (3.6) also holds unconditionally, because, letting Z_T denote the quantity on the left side of (3.6), by the Dominated Convergence Theorem and the almost sure convergence of the conditional distributions, we have

$$\lim_{T \to \infty} P\{Z_T \leqslant z\} = \lim_{T \to \infty} E[P\{Z_T \leqslant z \mid Y_{t,\infty}\}] = E[\lim_{T \to \infty} P\{Z_T \leqslant z \mid Y_{t,\infty}\}],$$

where the last term above equals the standard normal distribution function by (3.6).

We now consider the probability limit of the term $l(k)'(\Gamma_k^{-1} \otimes \Sigma)\, l(k)$. We have

$$\begin{aligned} l(k)'(\Gamma_k^{-1} \otimes \Sigma)\, l(k) = {} & \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} (k^{-1} Y'_{t+h-i-1,k} \Gamma_k^{-1} Y_{t+h-j-1,k})(l' B_i \Sigma B'_j l) \\ & + \sum_{i=0}^{h-1} \sum_{j=0}^{h-1} (k^{-1} Y'_{t,k} A'^{h-i-1}_{(k)} \Gamma_k^{-1} A^{h-j-1}_{(k)} Y_{t,k} \\ & - k^{-1} Y'_{t+h-i-1,k} \Gamma_k^{-1} Y_{t+h-j-1,k})(l' B_i \Sigma B'_j l). \end{aligned} \tag{3.7}$$

For the first term on the right side of (3.7) we have

$$\begin{aligned} k^{-1} Y'_{t+h-i-1,k} \Gamma_k^{-1} Y_{t+h-j-1,k} &\xrightarrow{p} r, \qquad i=j \\ &\xrightarrow{p} 0, \qquad i \neq j \end{aligned} \qquad \text{as } k \to \infty,$$

since, for $0 \leqslant i \leqslant j \leqslant h-1$, letting $\Gamma_Y(n) = E(Y_{t,k} Y'_{t+n,k})$,

$$\begin{aligned} E(k^{-1} Y'_{t+h-i-1,k} \Gamma_k^{-1} Y_{t+h-j-1,k}) &= k^{-1} \operatorname{tr}\{E(\Gamma_k^{-1} Y_{t,k} Y'_{t+j-i,k})\} \\ &= k^{-1} \operatorname{tr}\{\Gamma_k^{-1} \Gamma_Y(j-i)\}, \end{aligned}$$

which equals $k^{-1} \operatorname{tr}(\Gamma_k^{-1} \Gamma_k) = r$ when $i = j$, and which converges to zero as $k \to \infty$ when $i < j$, noting that the last $r(k-n)$ columns of $\Gamma_k^{-1} \Gamma_Y(n)$ equal the first $r(k-n)$ columns of I_{kr}. Furthermore, for $0 \leqslant i \leqslant j \leqslant h-1$,

$$\begin{aligned} &\operatorname{Var}(k^{-1} Y'_{t+h-i-1,k} \Gamma_k^{-1} Y_{t+h-j-1,k}) \\ &\quad = k^{-2}[\operatorname{tr}(\Gamma_k^{-1} \Gamma_k \Gamma_k^{-1} \Gamma_k) + \operatorname{tr}(\Gamma_k^{-1} \Gamma_Y(j-i)\, \Gamma_k^{-1} \Gamma_Y(j-i)) + C_k(j-i)] \\ &\quad = k^{-2}[kr + \operatorname{tr}((\Gamma_k^{-1} \Gamma_Y(j-i))^2) + C_k(j-i)] \to 0 \qquad \text{as} \quad k \to \infty, \end{aligned}$$

where $C_k(j-i)$, which is of order k [11, p. 211], is a term containing fourth cumulants of the process $\{y_t\}$. Hence the first term on the right side of (3.7) converges in probability to

$$\sum_{j=0}^{h-1} r(l' B_j \Sigma B_j' l) = l'(r\Sigma(h))\, l \qquad \text{as} \quad k \to \infty.$$

Furthermore, it is not difficult to show that the second term on the right side of (3.7) converges to zero in absolute mean (and hence in probability) as $k \to \infty$, noting the comment which precedes (3.6). Thus the denominator in (3.6), $l(k)'(\Gamma_k^{-1} \otimes \Sigma)\, l(k)$, converges in probability to $l'(r\Sigma(h))\, l$ as $k \to \infty$, and it follows that the first term on the right side of (3.5) converges in distribution to $N(0, r\Sigma(h))$ as $T \to \infty$.

Finally, we show that the second term on the right side of (3.5) converges to zero in probability. We have

$$\begin{aligned} &(T/k)^{1/2} \sum_{j=0}^{h-1} (Y'_{t,k} A'^{h-j-1}_{(k)} \otimes (\hat{B}_{jk} - B_j))(\hat{\alpha}(k) - \alpha(k)) \\ &\quad = \sum_{j=0}^{h-1} (\hat{B}_{jk} - B_j)(T/k)^{1/2} (Y'_{t,k} A'^{h-j-1}_{(k)} \otimes I_r)(\hat{\alpha}(k) - \alpha(k)), \end{aligned}$$

where $B_j = E_k' A^j_{(k)} E_k$ for $j \leqslant h-1 \leqslant k$. Now, for $0 \leqslant j \leqslant h-1$, $\hat{B}_{jk} - B_j \to^p 0$ as $T \to \infty$ by Theorem 4 and Serfling's [18, p. 122] Theorem A concerning the asymptotic distribution of a function of an asymptotically normal random vector. Also, using an argument similar to that which established the asymptotic normality of the first term on the right side of (3.5), we can show that, for $0 \leqslant j \leqslant h-1$,

$$(T/k)^{1/2}(Y'_{t,k} A'^{h-j-1}_{(k)} \otimes I_r)(\hat{\alpha}(k) - \alpha(k)) \xrightarrow{d} N(0, r\Sigma) \qquad \text{as} \quad T \to \infty.$$

It follows that

$$(T/k)^{1/2} \sum_{j=0}^{h-1} (Y'_{t,k} A'^{h-j-1}_{(k)} \otimes (\hat{B}_{jk} - B_j))(\hat{\alpha}(k) - \alpha(k)) \xrightarrow{p} 0 \qquad \text{as} \quad T \to \infty,$$

and the proof of the theorem is complete.

We now state the main result of this section, concerning the asymptotic distribution of the predictor $\hat{y}_{t,k}(h)$.

THEOREM 6. *Let* $\{y_t\}$ *satisfy* (2.1), *and assume that all the conditions of Theorem* 5 *are satisfied. Then*

$$(T/k)^{1/2}(\hat{y}_{t,k}(h) - y_t^*(h)) \xrightarrow{d} N(0, r\Sigma(h)) \qquad \text{as} \quad k, T \to \infty \text{ for all } h \geqslant 1.$$

Proof. Writing $\hat{y}_{t,k}(h) - y_t^*(h)$ in the form (3.3), the proof follows immediately from result (3.4) and Theorem 5.

4. DISCUSSION OF RESULTS

Using the asymptotic distributional result of Theorem 6, we obtain an asymptotic approximation, as k, $T \to \infty$, to the mean square error matrix of $\hat{y}_{t,k}(h) - y_t^*(h)$ as

$$E[(\hat{y}_{t,k}(h) - y_t^*(h))(\hat{y}_{t,k}(h) - y_t^*(h))'] \approx (kr/T)\, \Sigma(h),$$

where $\Sigma(h)$ is the mean square error of the "optimal" predictor $y_t^*(h)$. Using the above asymptotic approximation and (3.2), we can approximate the mean square error of the predictor $\hat{y}_{t,k}(h)$ by

$$\Sigma_k(h) \approx \Sigma(h) + (kr/T)\, \Sigma(h) = (1 + kr/T)\, \Sigma(h). \tag{4.1}$$

Of course, this approximation is based on the covariance matrix of the limiting distribution of $(T/k)^{1/2}(\hat{y}_{t,k}(h) - y_t^*(h))$, which may not in general be guaranteed to equal the limit of the actual covariance matrices of this sequence of random vectors without additional considerations. Fuller and Hasza [10] have shown that these two asymptotic approaches do in fact yield equivalent prediction mean square error results in the case of prediction from a finite order univariate Gaussian autoregressive model. Also, in the univariate case for one-step ahead prediction, Shibata [19] has considered the asymptotic behavior of the quantity $(T/k)(\hat{A}(k)-A(k))' \Gamma_k(\hat{A}(k)-A(k))$, which is the mean square error of $(T/k)^{1/2}$ times the first term on the right side of (3.3) with $h=1$ and $r=1$, conditional on the values of $\hat{A}(k)$. He has shown that this quantity converges in probability to σ^2 as $T\to\infty$ under assumptions (ii) and (iii) of Theorem 2, and hence the mean square error of $(T/k)^{1/2}(\hat{y}_{t,k}(1)-y_t^*(1))$, conditional on the $\hat{A}(k)$, also converges in probability to σ^2 as $T\to\infty$. It seems likely that this argument can be extended to the present multivariate setting as well as to more than one-step ahead prediction errors. We note also that, in the case of one-step ahead prediction, $\Sigma_k(1)\approx(1+kr/T)\,\Sigma$ coincides with the asymptotic approximation obtained by Reinsel [15] under the assumptions that $\{y_t\}$ is a finite autoregression of known order k, and that k is fixed as $T\to\infty$.

A useful feature of the approximation (4.1) is its simplicity, which allows (4.1) to be both easily interpreted and computed. Note that (4.1) implies that the approximate asymptotic effect of parameter estimation in the autoregressive model fitting approach is to inflate the mean square prediction error $\Sigma(h)$ by a factor of $(1+kr/T)$. We also note that our results agree with those of Bhansali [7] in the univariate case, although for more than one-step ahead prediction we have extended these results by providing a simple, explicit expression in Theorem 6 for the covariance matrix of the asymptotic distribution of $(T/k)^{1/2}(\hat{y}_{t,k}(h)-y_t^*(h))$ which has not been previously obtained explicitly in the univariate case. As noted by Bhansali [7], in practice one may consider using the "finite sample" approximation to (4.1),

$$\Sigma(h)+T^{-1}\sum_{j,i=0}^{h-1}\operatorname{tr}\{A_k'^{\,h-1-j}\Gamma_k^{-1}A_k^{h-1-i}\Gamma_k\}(B_j\Sigma B_i'), \tag{4.2}$$

as derived by Reinsel [15] for the finite order multivariate autoregressive case. Although (4.1) is preferable on the basis of its simplicity, further investigation would be needed to compare the accuracy of the approximations (4.1) and (4.2). Also, in practical use estimates $\hat{B}_{jk}$ and $\hat{\Sigma}_k$ would need to be substituted in place of the B_j and Σ in (4.1) and (4.2).

Finally, we note that in practice one must choose the value of k to use for any given series length T. While we can provide no specific guidelines in this matter, the asymptotic approximation that has been obtained suggests that it may be reasonable to use Akaike's [3] FPE criterion, which was originally suggested for selecting the order of a finite autoregressive process by choosing the value of k which minimizes the determinant of the estimated one-step ahead mean square prediction error matrix, to determine a finite order approximation to a true infinite order autoregressive process. This has previously been noted by Bhansali [7] for the univariate case. One might also use the alternative CAT criterion suggested by Parzen [14]. The result (4.1) may also be convenient and useful when one is interested in choosing the "approximating" autoregressive order k which minimizes the mean square error of general $h \geqslant 1$ step ahead predictors.

5. Numerical Example

In this section we present the results of a sampling experiment conducted to investigate the finite sample properties of prediction errors based on the autoregressive model fitting procedure and compare their behavior with the theoretical asymptotic results obtained in the previous sections. We consider the bivariate ARMA (1, 1) model $y_t - \Phi y_{t-1} = \varepsilon_t - \theta \varepsilon_{t-1}$, with

$$\Phi = \begin{bmatrix} 1.2 & -0.5 \\ 0.6 & 0.3 \end{bmatrix}, \qquad \theta = \begin{bmatrix} -0.6 & 0.3 \\ 0.3 & 0.6 \end{bmatrix}, \qquad \Sigma = \begin{bmatrix} 1.00 & 0.50 \\ 0.50 & 1.25 \end{bmatrix}.$$

While this model represents an infinite order autoregressive model, we will consider properties of predictors obtained by fitting finite order autoregressive models. Note that if the order k of an "approximating" autoregressive model for this ARMA (1, 1) model is chosen by minimizing $\det\{(1 + kr/T)\,\Sigma_k\}$, a version of Akaike's [3] FPE criterion in which the sample estimate of Σ_k is replaced by its theoretical value, we obtain $k = 4$ when $T = 100$, with

$$\Sigma_4 = \begin{bmatrix} 1.02 & 0.51 \\ 0.51 & 1.27 \end{bmatrix}.$$

For $T = 100$, realizations of the above ARMA (1, 1) process of length $T + 5$ were generated, with the ε_t normally distributed with mean zero and covariance matrix Σ. Autoregressive models of orders k, for $k = 2$, 3, 4 and 5, were fit to the first T observations of the realizations using (2.4). For $h = 1$ to 5, h step ahead predictions of the future values at times $T + 1, \ldots, T + 5$ were formed based on the fitted autoregressive models. These predictions were compared to their corresponding actual values and squared prediction errors were computed. The averages of these squared prediction errors, based on 2500 realizations, are given in Table I under the heading "Observed." Also given in Table I are the diagonal elements of the theoretical prediction mean square error matrices based on the asymptotic approximation (4.1). Comparing the observed average squared prediction errors with the approximation (4.1), we find reasonably good agreement, and note that in particular the approximation $(1 + kr/T)\,\Sigma(h)$ is clearly to be preferred over the "unadjusted" value $\Sigma(h)$.

TABLE I

Mean Square Errors (Diagonal Elements of $\Sigma_k(h)$) of Predicting by Autoregressive Model Fitting for Bivariate ARMA (1, 1) Example ($T = 100$, 2500 Replications)

Lead	AR order	Observed		Theoretical $(1+2k/100)\,\Sigma(h)$		Theoretical with known parameters $\Sigma(h)$	
$h=1$	$k=2$	1.13	1.43	1.04	1.30	1.00	1.25
	3	1.09	1.40	1.06	1.33		
	4	1.11	1.39	1.08	1.35		
	5	1.12	1.42	1.10	1.38		
$h=2$	$k=2$	3.86	1.56	3.74	1.42	3.60	1.36
	3	3.94	1.56	3.82	1.44		
	4	4.01	1.55	3.89	1.47		
	5	4.05	1.57	3.96	1.50		
$h=3$	$k=2$	7.31	2.65	7.11	2.57	6.83	2.47
	3	7.48	2.64	7.24	2.62		
	4	7.53	2.72	7.38	2.67		
	5	7.63	2.75	7.52	2.72		
$h=4$	$k=2$	10.24	4.77	9.89	4.59	9.50	4.41
	3	10.50	4.83	10.08	4.68		
	4	10.49	4.81	10.27	4.76		
	5	10.67	4.87	10.46	4.85		
$h=5$	$k=2$	11.94	6.74	11.55	6.62	11.11	6.37
	3	12.20	6.83	11.78	6.75		
	4	12.20	6.82	12.00	6.88		
	5	12.45	6.96	12.22	7.00		

REFERENCES

[1] AKAIKE, H. (1969). Power spectrum estimation through autoregressive model fitting. *Ann. Inst. Statist. Math.* **21** 407–419.

[2] AKAIKE, H. (1970). Statistical predictor identification. *Ann. Inst. Statist. Math.* **22** 203–217.

[3] AKAIKE, H. (1971). Autoregressive model fitting for control. *Ann. Inst. Statist. Math.* **23** 163–180.

[4] ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

[5] BAILLIE, R. T. (1979). Asymptotic prediction mean squared error for vector autoregressive models. *Biometrika* **66** 675–678.

[6] BERK, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* **2** 489–502.

[7] BHANSALI, R. J. (1978). Linear prediction by autoregressive model fitting in the time domain. *Ann. Statist.* **6** 224–231.

[8] BLOOMFIELD, P. (1972). On the error of prediction of a time series. *Biometrika* **59** 501–507.

[9] CHUNG, K. L. (1974). *A Course in Probability Theory*, 2nd ed. Academic Press, New York.

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

[11] HANNAN, E. J. (1970). *Multiple Time Series.* Wiley, New York.

[12] MANN, H. B., AND WALD, A. (1943). On stochastic limit and order relationships. *Ann. Math. Statist.* **14** 217–226.

[13] PARZEN, E. (1969). Multiple time series modeling. In *Multivariate Analysis* II (P. R. Krishnaiah, Ed.), pp. 389–409. Academic Press, New York.

[14] PARZEN, E. (1974). Some recent advances in time series modeling. *IEEE Trans. Automat. Control* **AC-19** 723–730.

[15] REINSEL, G. C. (1980). Asymptotic properties of prediction errors for the multivariate autoregressive model using estimated parameters. *J. Roy. Statist. Soc. Ser. B* **42** 328–333.

[16] SCHMIDT, P. (1974). The asymptotic distribution of forecasts in the dynamic simulation of an econometric model. *Econometrica* **42** 303–309.

[17] SCOTT, D. J. (1973). Central limit theorems for martingales using a Skorokhod representation approach. *Adv. Appl. Probab.* **5** 119–137.

[18] SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

[19] SHIBATA, R. (1977). Convergence of least squares estimates of autoregressive parameters. *Austral. J. Statist.* **19** 226–235.

[20] SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **8** 147–164.

[21] WIENER, N., AND MASANI, P. (1958). The prediction theory of multivariate stochastic processes, II. *Acta Math.* **99** 93–137.

[22] YAMAMOTO, T. (1976). Asymptotic mean square prediction error for an autoregressive model with estimated coefficients. *Appl. Statist.* **25** 123–127.

[23] YAMAMOTO, T. (1981). Prediction of multivariate autoregressive-moving average models. *Biometrika* **68** 485–492.