Asymptotic Prediction Mean Squared Error for Vector Autoregressive Models

Richard T. Baillie


Stable URL:
http://links.jstor.org/sici?sici=0006-3444%28197912%2966%3A3%3C675%3AAPMSEP%3E2.0.CO%3B2-B

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

Biometrika is published by Biometrika Trust. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/bio.html.

Biometrika
©1979 Biometrika Trust

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2002 JSTOR
Asymptotic prediction mean squared error for vector autoregressive models

BY RICHARD T. BAILLIE
University of Aston Management Centre, Birmingham

SUMMARY

This paper derives the asymptotic mean squared error of multistep prediction for the general vector autoregressive process. For one-step-ahead prediction the result is a surprisingly simple generalization of the result for the scalar autoregressive process. Results for multistep prediction are also derived for the regression model with autoregressive errors, where the set of exogenous variables follows a vector autoregressive process.

Some key words: Autoregressive model; Multivariate time series model; Prediction mean squared error; Regression model with autoregressive errors; Time series.

1. THE VECTOR AUTOREGRESSIVE PROCESS

This section considers the general vector autoregressive process of order q and dimension g. The process can be written in terms of the lag operator L, as

\[ A(L) x_t = \eta_t, \]  

where

\[ A(L) = 1 - \sum_{j=1}^{q} A_j L^j, \quad x_t' = (x_{1t} \ldots x_{gt}), \quad \eta_t' = (\eta_{1t} \ldots \eta_{gt}), \]

\[ E(\eta_t) = 0, \quad E(\eta_t \eta_t') = \begin{cases} \sum (r = 0), \\ 0 (r \neq 0). \end{cases} \]

Also \( \Sigma \) has typical element \( \sigma_{ij}^2 \), and the \( A_j \) are \( g \times g \) coefficient matrices.
It is convenient to collapse the qth order model (1.1) to a first order system by means of the companion form representation

\[
\begin{bmatrix}
  x_t \\
  x_{t-1} \\
  \vdots \\
  x_{t-q+1}
\end{bmatrix} =
\begin{bmatrix}
  A_1 & A_2 & \ldots & A_q \\
  I & 0 & \ldots & 0 \\
  0 & I & \ldots & 0 \\
  \vdots & \vdots & \ddots & \vdots \\
  0 & 0 & \ldots & 0
\end{bmatrix}
\begin{bmatrix}
  x_{t-1} \\
  x_{t-2} \\
  \vdots \\
  x_{t-q}
\end{bmatrix} +
\begin{bmatrix}
  \eta_t \\
  0 \\
  \vdots \\
  0
\end{bmatrix},
\]

which is written as \( X_t = AX_{t-1} + a_t \), where \( X_t \) and \( a_t \) are \( gq \times 1 \) and \( A \) is \( gq \times gq \).

For model (1.1) to be stationary it is necessary that the roots of \( A(L) \) lie outside the unit circle. At time \( n+l \) the model can be expressed as

\[ x_{n+l} = N' \sum_{j=0}^{l-1} A^j a_{n+l-j} + N'A^l X_n, \]

where \( N' = (I_g, 0) \) and is of dimension \( g \times gq \). The minimum mean squared error predictor made at time \( n \), \( l \) periods ahead is then

\[ x_{n,l} = N'A^l X_n. \]
Miscellanea

In practice the $g^2q$ structural parameters contained in $A$ will have to be estimated by an approximate maximum likelihood estimator as given by Tunnicliffe Wilson (1973). By using $R$, the row stacking operator, the $g^2q$ dimensional parameter vector $\gamma$, is given as $\gamma' = R(N'A)$. Anderson (1971, pp. 198–205) has shown that the maximum likelihood estimator of $\gamma$, based on $n$ observations, has the multivariate normal distribution

$$\sqrt{n} (\hat{\gamma} - \gamma) \sim N(0, n^{-1} \Sigma \otimes \Gamma^{-1}),$$

(1.3)

where $\Gamma = (X_n X_n')$.

Once the model has been estimated (1.2) can be modified to form the practical predictor

$$\hat{x}_{n,l} = N'A_l X_n$$

(1.4)

and the asymptotic mean squared error of $l$ step prediction is then defined as

$$\text{AMSE} (\hat{x}_{n,l}) = N' \sum_{j=0}^{l-1} A^j N \Sigma N'A'^j + E((\hat{x}_{n,l} - x_{n,l}) (\hat{x}_{n,l} - x_{n,l})'),$$

(1.5)

where the last term is due to parameter estimation and is $O(n^{-1})$.

Upon taking a first order Taylor series approximation on (1.4) we can show that

$$\hat{x}_{n,l} - x_{n,l} = -(\hat{\gamma} - \gamma)' H(l)(I \otimes X_n),$$

where

$$H(l) = \frac{\partial R(N'A_l)}{\partial \gamma} = F' \sum_{j=0}^{l-1} A'^j \otimes A^{l-1-j} F$$

and $F = (I \otimes 0)$ and is of dimension $g^2q \times g^2q$. The parametric expression for $H(l)$ is a simple extension of the results obtained by Schmidt (1974) and Yamamoto (1976). From (1.5) it can be shown that

$$\text{AMSE} (\hat{x}_{n,l}) = N' \sum_{j=0}^{l-1} A^j N \Sigma N'A'^j N + E((I \otimes X_n)' H(l)'(\hat{\gamma} - \gamma)(\hat{\gamma} - \gamma)' H(l)(I \otimes X_n)).$$

(1.6)

To obtain an expression to order $O(n^{-1})$ for this last term we note that $(\hat{\gamma} - \gamma)$ and $X_n$ are asymptotically independent to order $o(n^{-1})$. Justification for this statement is provided, for example, by Schmidt (1977, p. 998) and Phillips (1979).

For one-step-ahead prediction $H(1) = I$ and from (1.3), equation (1.6) becomes

$$\text{AMSE} (\hat{x}_{n,1}) = \Sigma + n^{-1} E(\Sigma \otimes X_n \Gamma^{-1} X_n).$$

On noting that the last term in the above is a scalar and applying the trace operator we can easily show that

$$\text{AMSE} (\hat{x}_{n,1}) = (1 + gq/n) \Sigma.$$  

(1.7)

When $g = 1$, (1.7) reduces to $\sigma^2(1 + q/n)$, a result previously obtained by Davisson (1965) and Bloomfield (1972).

In order to obtain an unconditional formula for the error of multistep prediction it is simplest to consider a single element say $x_{it}$ of $x_t$. We then define $e$ as a $g$ dimensional column vector with unit element in the $i$th position and zeros elsewhere. From (1.6)

$$\text{AMSE} (\hat{x}_{i,n}) = \epsilon' N' \sum_{j=0}^{l-1} A^j N \Sigma N'A'^j N\epsilon + n^{-1} \Omega(l),$$

where

$$\Omega(l) = \text{tr} \{H(l)'(\Sigma \otimes \Gamma^{-1}) H(l) K\}$$

and $K$ is a square matrix of dimension $g^2q$ containing square submatrices, all of which are of dimension $gq$ and are null matrices, except the $(i,i)$th submatrix which is $\Gamma$.

A fuller derivation of these results is given in my London Ph.D. thesis.
2. Vector autoregressive models generating exogenous variables

We now consider the regression model containing \( k \) exogenous variables, no lagged endogenous variables, but a \( p \)th order autoregressive error process. In the usual notation the model is

\[
y_t = x_t' \beta + \epsilon_t, \quad \phi(L) u_t = \epsilon_t,
\]

where \( \epsilon_t \) is a white noise process with zero mean and constant variance \( \sigma^2 \) and \( \beta \) is a \( k \times 1 \) vector of parameters. It is assumed that all the roots of \( \phi(L) \) lie outside the unit circle, so that \( u_t \) has the unique moving average representation

\[
u_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}.
\]

At time \( n+l \), the model can be represented as

\[
y_{n+l} = \sum_{j=0}^{l-1} \psi_j \epsilon_{n+l-j} + x_{n+l}' \beta + \zeta^*(l)' (Y_n - X_n' \beta),
\]

where \( \zeta^*(l) \) is a \( p \times 1 \) vector of prediction weights, \( Y_n \) is a vector of the last \( p \) observations on \( y \) and \( X_n' \) is a \( p \times k \) matrix of the last \( p \) observations on the \( k \) independent variables; a full description of this result is given by Baillie (1979) who also derives the prediction asymptotic mean squared error of (2.1) for the situation when the future values of the exogenous variables are either known or predicted from their own autoregressive models. It is now assumed that the exogenous variables follow the vector autoregressive process (1.1); hence the predictor of \( y_{n+l} \) based on maximum likelihood estimates of the parameters is given by

\[
\hat{\beta}_{n,l} = x_{n+l}' \hat{\beta} + \zeta^*(l)' (Y_n - X_n' \hat{\beta}), \quad \hat{x}_{n,l} = N' \hat{A}^l X_n.
\]

The joint maximum likelihood estimates of \( \beta \), \( \phi \) and \( \gamma \) are asymptotically mutually independent, say with covariance matrices \( \sigma^2 B^{-1}/n, \sigma^2 V^{-1}/n \) and (1.3) respectively. Then

\[
\text{AMSE} (\hat{\beta}_{n,l}) = \sigma^2 \sum_{j=0}^{l-1} \psi_j^2 + \sigma^2 n^{-1} \text{tr} \{ M(l) VM(l)' V^{-1} \}
\]

\[
+ \sigma^2 n^{-1} \{ X_n' A^l N B^{-1} N' A^l X_n + \zeta(l)' X_n' B^{-1} X_n' \zeta(l) - 2X_n' A^l N B^{-1} X_n' \zeta(l) \}
\]

\[
+ \epsilon^2 N' \sum_{j=0}^{l-1} A^j N \sum N' A^j N \beta e + n^{-1} \Omega(l).
\]

A parametric expression for \( M(l) \) and a proof closely related to the above is given in § 4 of Baillie (1979).

The last two terms in (2.3) are components due to the exogenous variables following a vector autoregressive process. The first term in (2.3) is due to the naturally occurring random innovations \( \epsilon_{n+1}, \ldots, \epsilon_{n+l} \), the second term is due to estimation of the \( \phi \) parameters, the third due solely to estimation of \( \beta \), whilst the fourth and fifth terms are generated by the joint effect of estimating the \( \beta \) and \( \phi \) parameters.

The author is extremely grateful to Professors J. Durbin and K. F. Wallis for their help and supervision.

References


[Received March 1979. Revised May 1979]