Biometrika (1987), **74**, 1, *pp.* 103–13
Printed in Great Britain

Contemporaneous bivariate time series

BY F. CAMACHO, A. I. McLEOD
Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Ontario, Canada N6A 5B9
AND K. W. HIPEL
Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

SUMMARY

Bivariate autoregressive-moving average time series with diagonal parameter matrices for the autoregressive and moving average components exhibit only contemporaneous or instantaneous correlation. In practice, different lengths of each series may be available. An efficient maximum likelihood algorithm for parameter estimation is derived. The statistical efficiency of this new procedure is compared with that of the standard multivariate and univariate procedures which utilize only part of the available data.

Some key words: Autoregressive-moving average model; Contemporaneous Granger causality; Multiple time series; Statistical efficiency.

1. INTRODUCTION

Multiple time series which exhibit only instantaneous or contemporaneous correlation have been studied by several authors (Nelson, 1976; Pierce, 1977; Moriarty & Salomon, 1980; Risager, 1980, 1981; Umashankar & Ledolter, 1983; Cipra, 1984). The contemporaneous autoregressive-moving average model is defined as

$$\phi_h(B)(Z_t^{(h)}-\mu_h)=\theta_h(B)a_t^{(h)}\quad (h=1,\ldots,k),$$

where

$$\phi_h(B)=1-\phi_1^{(h)}B-\ldots-\phi_{p_h}^{(h)}B^{p_h},\quad \theta_h(B)=1-\theta_1^{(h)}B-\ldots-\theta_{q_h}^{(h)}B^{q_h},$$

B is the backshift operator, p_h and q_h are, respectively, the autoregressive and moving averages orders for the series $Z_t^{(h)}$, for $h=1,\ldots,k$, $\mu=(\mu_1,\ldots,\mu_k)'$ is the vector of the means and vectors $a_t=(a_t^{(1)},\ldots,a_t^{(k)})'$ form a sequence of independent normal random vectors with mean vector zero and covariance matrix $\Delta=((\sigma_{gh}))$. If $\sigma_{gh}=0$ for $g\neq h$, the model collapses to a set of k independent univariate models as defined by Box & Jenkins (1976). On the other hand, this model can also be considered as a particular case of the general multivariate model (Tiao & Box, 1981; Jenkins & Alavi, 1981), where the autoregressive and moving average parameter matrices are constrained to be diagonal.

The contemporaneous model has been successfully employed to model and forecast many actual time series. Umashankar & Ledolter (1983), Moriarity & Salomon (1980) and Nelson (1976) apply this model to increase the efficiency of the estimated parameters and to improve the accuracy of the forecasts. Camacho, McLeod & Hipel (1985) give applications of the contemporaneous model to stochastic hydrology. Risager (1980) fitted a bivariate contemporaneous model to mean annual ice core measurements for which data were available for the years 1861–1974 and 1869–1975. The contemporaneous model also describes the case when only contemporaneous Granger causality is present among the series (Granger, 1969; Pierce & Haugh, 1977, 1979). As illustrated by Risager's example, it is quite common to have time series with unequal sample sizes. The practice in this circumstance has been to eliminate the additional information available in the longer series so that all the series end up with an equal number of observations.

In this paper, the bivariate contemporaneous model is discussed when $m+N$ observations $\{Z_t^{(1)}\}$, for $t=1-m,\ldots,0,1,\ldots,N$, and N observations $\{Z_t^{(2)}\}$, for $t=1,\ldots,N$, are available. In § 2 expressions for the likelihood function are given. In § 3, some simplifications are considered to obtain an efficient algorithm to calculate the likelihood function. In § 4, the distribution of the estimates is given and the possible gain in efficiency in the estimation of the parameters is considered.

It is possible to extend the results to the case of three or more series, one of which has m additional observations. However, the more general case of k-series, each one of them having a different sample size is more difficult to handle since the exact likelihood function can be very complicated.

2. THE LIKELIHOOD FUNCTION

The likelihood function for the general multivariate model has been given by Hillmer & Tiao (1979) and Nicholls & Hall (1979). However, this likelihood function was derived under the assumption of equal sample sizes for all the series and hence is not applicable in this case, where it is assumed that the time series have unequal numbers of observations.

In order to find expressions for the likelihood function of the bivariate model, the following notation is introduced: let $\{Z_t^{(1)}\}$, for $t=1-m,\ldots,0,1,\ldots,N$, denote $m+N$ observations of $Z_t^{(1)}$ and let $\{Z_t^{(2)}\}$, for $t=1,\ldots,N$, denote N observations of $Z_t^{(2)}$, where $Z_t=(Z_t^{(1)}, Z_t^{(2)})'$ follows the bivariate model given in § 1. It is assumed that for $h=1,2$ the polynomials $\phi_h(B)$ and $\theta_h(B)$ have their roots outside the unit circle and the pairs $\{\phi_h(B), \theta_h(B)\}$ do not have common factors. These assumptions assure stationarity, invertibility and identifiability of the model. It is also assumed, without the loss of generality, that $p_h=p$, $q_h=q$ $(h=1,2)$ and that $\mu=0$. Let (β, Δ) denote the parameters of the model, where

$$\beta=(\beta_1', \beta_2')', \quad \beta_h=(\phi_1^{(h)},\ldots,\phi_p^{(h)}, \theta_1^{(h)},\ldots,\theta_q^{(h)}) \quad (h=1,2).$$

Let a represent the innovations of the process where $a=(a_1', a_2')'$, $a_1=(a_{10}', a_{11}')'$, $a_{10}=(a_{1-m}^{(1)},\ldots,a_0^{(1)})'$, $a_{11}=(a_1^{(1)},\ldots,a_N^{(1)})'$, $a_2=(a_1^{(2)},\ldots,a_N^{(2)})'$. Let $Z=(Z_1', Z_2')'$, where $Z_1=(Z_{10}', Z_{11}')'$, $Z_{10}=(Z_{1-m}^{(1)},\ldots,Z_0^{(1)})'$, $Z_{11}=(Z_1^{(1)},\ldots,Z_N^{(1)})'$, $Z_2=(Z_1^{(2)},\ldots,Z_N^{(2)})'$. Finally, let $e=(e_1, e_2)$, where $e_1=(a_{-(m-1+q)}^{(1)},\ldots,a_{-m}^{(1)}, Z_{-(m-1+p)}^{(1)},\ldots,Z_{-m}^{(1)})'$, $e_2=(a_{1-q}^{(2)},\ldots,a_0^{(2)}, Z_{1-p}^{(2)},\ldots,Z_0^{(2)})'$. So e represents the initial values of the process. It is easy to see that for suitable matrices H_h, F_h and G_h $(h=1,2)$, the relationship between Z, a and e can be written as $HZ=Fa+Ge$, where H, F and G are block diagonal matrices with blocks H_1, H_2, F_1, F_2, G_1 and G_2. After multiplying by F^{-1} and rearranging terms, we have that

$$(e', a')=(O', K')'Z+(1_{2(p+q)}, L')'e=\Xi Z+\Lambda e, \quad (1)$$

where $1_{2(p+q)}$ is the identity matrix of order $2(p+q)$ and K and L are block diagonal matrices with blocks $F_1^{-1}H_1$, $F_2^{-1}H_2$, $F_1^{-1}G_1$ and $F_2^{-1}G_2$, respectively.

From (1), it can be seen that the Jacobian of the linear transformation from $(e', a')'$ to $(e', Z')'$ is one, so that the joint distribution of $(e', Z')'$ is multivariate normal with probability density function given by

$$L(e, Z) = (2\pi)^{-\frac{1}{2}\{2N+m+2(p+q)\}}|\Omega|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}S(Z, e)\}, \tag{2}$$

where

$$S(Z, e) = (\Xi Z + \Lambda e)'\Omega^{-1}(\Xi Z + \Lambda e) = S(Z, \hat{e}) + (e - \hat{e})'\Lambda'\Omega^{-1}\Lambda(e - \hat{e}),$$

$$\hat{e} = -(\Lambda'\Omega^{-1}\Lambda)^{-1}\Lambda'\Omega^{-1}\Xi Z.$$

On integrating out e from (3) the distribution of Z is obtained and is given by

$$L(Z) = (2\pi)^{-\frac{1}{2}(2n+M)}|\Omega|^{-\frac{1}{2}}|\Lambda'\Omega^{-1}\Lambda|^{-\frac{1}{2}} \exp\{-\tfrac{1}{2}S(Z, e)\}. \tag{3}$$

After the data Z is available, (3) yields the likelihood function of β and Δ, $L(\beta, \Delta)$.

3. Calculation of the likelihood

3·1. *Introduction*

It is interesting that the form of the likelihood function $L(\beta, \Delta)$ given by (3) is similar to the likelihood of a multivariate autoregressive-moving average model as given by Hillmer & Tiao (1979) or Nicholls & Hall (1979). Hall & Nicholls (1980) gave an algorithm to evaluate the likelihood function of the general multivariate model. Although their approach could be employed in the contemporaneous case, it would not be computationally efficient. In fact, due to the structure of the model, some simplifications can be made to obtain a more efficient procedure. In this section some explicit expressions for the terms of the likelihood function as well as some simplifications are given.

3·2. *Calculation of the sum of squares $S(Z, \hat{e})$*

In this section a method to calculate $S(Z, \hat{e})$ which corresponds to the unconditional sum of squares of Box & Jenkins (1976) is presented. Now $S(Z, \hat{e})$ is given by

$$S(Z, \hat{e}) = (\Xi Z + \Lambda\hat{e})\Omega^{-1}(\Xi Z + \Lambda\hat{e}), \tag{4}$$

where $\hat{e}$ is the vector of estimated initial values and Ω is the covariance matrix of $(e', a')' = (e_1', e_2', a_{10}', a_{11}', a_2')'$. The inverse covariance matrix, Ω^{-1}, is block diagonal with blocks

$$\Gamma^{-1} = \begin{bmatrix} \Omega_{e_1e_1} & \Omega_{e_1e_2} & 0 \\ \Omega_{e_2e_1} & \Omega_{e_2e_2} & \Omega_{e_2a_{10}} \\ 0 & \Omega_{a_{10}e_2} & \sigma_{11}1_m \end{bmatrix}^{-1}$$

and $\Delta^{-1}\otimes 1_N$, where $\Omega_{uv} = E(uv')$ $(u, v = e_1, e_2, a_{10})$. It can be shown that $\Gamma^{-1} = J_1'P^{-1}J_1 + J_2$, where $J_1 = [1_{2(p+q)} | -\sigma_{11}^{-1}\Omega_{ea_{10}}]$, J_2 is block diagonal with blocks 0 and $\sigma_{11}^{-1}1_m$, and

$$P = \begin{bmatrix} \Omega_{e_1e_1} & \Omega_{e_1e_2} \\ \Omega_{e_2e_1} & \Omega_{e_2e_2} - \Omega_{e_2a_{10}}\Omega_{a_{10}e_2}/\sigma_{11} \end{bmatrix}.$$

For a moving average process, $\Omega_{e_2a_{10}}\Omega_{a_{10}e_2} = \sigma_{11}\rho^2\Omega_{e_2e_2}$, where ρ is the lag zero correlation coefficient between $a_t^{(1)}$ and $a_t^{(2)}$. For the univariate process it can be shown that $\Omega_{e_2a_{10}}\Omega_{a_{10}e_2} = \sigma_{11}\rho^2\Omega_{e_2e_2} + O(\delta^{2(m-p)})$, where $|\delta| < 1$. Hence $\Omega_{e_2e_2} - \Omega_{e_2a_{10}}\Omega_{a_{10}e_2}/\sigma_{11} \simeq (1-\rho^2)\Omega_{e_2e_2}$. Now, let $U = (\Xi Z + \Lambda\hat{e}) = (\hat{e}', \hat{a}')$, where $\hat{a}_t^{(h)}$ corresponds to the estimated value of the innovations of series $Z_t^{(h)}$ $(h=1, 2)$ using the data $\{Z_t^{(1)}\}$ $(t=1-m, \ldots, N)$ and $\{Z_t^{(2)}\}$ $(t=1, \ldots, N)$ and the vector of starting values $\hat{e}$. The values of $\hat{a}_t^{(h)}$ can be obtained recursively using (1). For example,

$$\hat{a}_t^{(1)} = Z_t^{(1)} - \phi_1^{(1)} Z_{t-1}^{(1)} - \ldots - \phi_p^{(1)} Z_{t-p}^{(1)} + \theta_t^{(1)} \hat{a}_{t-1}^{(1)} + \ldots + \theta_q^{(1)} \hat{a}_{t-q}^{(1)},$$

$t=1-m, \ldots, N$, with starting values given by $\hat{e}_1$.

Hence $S(Z, \hat{e})$ can then be calculated as

$$S(Z, \hat{e}) = U'\Omega^{-1}U = (\hat{e}', \hat{a}_{10}') J_1' P^{-1} J_1 (\hat{e}', \hat{a}_{10}')' + \sum_{t=1-m}^{0} (a_t^{(1)})^2/\sigma_{11} + \sum_{t=1}^{N} \hat{a}_t' \Delta^{-1} \hat{a}_t,$$

where $\hat{a}_t = (\hat{a}_t^{(1)}, a_t^{(2)})'$. For large values of m the first term of this expression can be expressed as $\zeta' P^{-1} \zeta$, where $\zeta = (\hat{e}_1', \hat{\xi}')$, $\xi = (\varepsilon_{1-q}^{(2)}, \ldots, \varepsilon_0^{(2)}, X_{1-p}, \ldots, X_0)$, $\varepsilon_t^{(2)} = a_t^{(2)} - \sigma_{11}^{-1}\sigma_{12} a_t^{(1)}$, $X_t = Z_t^{(2)} - \sigma_{11}^{-1}\sigma_{12}\phi_2(B) a_t^{(1)}$. Thus

$$S(Z, \hat{e}) = \zeta' P^{-1} \zeta + \sum_{t=1-m}^{0} (a_t^{(1)})^2/\sigma_{11} + \sum_{t=1}^{N} \hat{a}_t \Delta^{-1} \hat{a}_t.$$

3·3. *Calculation of the starting values*

Even though the vector of initial values $\hat{e}$ can be calculated as $\hat{e} = -(\Lambda'\Omega^{-1}\Lambda)^{-1}\Lambda'\Omega^{-1}\Xi Z$ it would be useful to have an alternative algorithm to obtain the vector $\hat{e}$ which could be more efficient, in particular, when dealing with seasonal or large-order models. In this section the backforecasting method of Box & Jenkins (1976, Ch. 7) for univariate models is extended. The backward and forward representations of the model do not have, in general, the same parameters as is the case for univariate models (Whittle, 1963). Furthermore, in the contemporaneous model the parameter matrices of the forward representation are in general nondiagonal.

In order to apply the backforecasting technique to the contemporaneous model consider the modified Cholesky decomposition of the covariance matrix Δ. Let $\Delta = LVL'$, where L is the lower triangular matrix with ones on the diagonal and off-diagonal entry σ_{21}/σ_{11} and V is the diagonal matrix given by $V = \text{diag}\,(\sigma_{11}, \sigma_{22} - \sigma_{12}^2/\sigma_{11}) = \text{diag}\,(\sigma_{12}, \sigma_{2:1})$. Then the model can be expressed as

$$\begin{bmatrix} \phi_1(B) Z_t^{(1)} \\ \phi_2(B) Z_t^{(2)} \end{bmatrix} = \begin{bmatrix} \theta_1(B) & 0 \\ \sigma_{12}\theta_2(B)/\sigma_{11} & \theta_2(B) \end{bmatrix} \begin{bmatrix} \varepsilon_t^{(1)} \\ \varepsilon_t^{(2)} \end{bmatrix},$$

where $\varepsilon_t = L^{-1} a_t = (a_t^{(1)}, a_t^{(2)} - \sigma_{11}^{-1}\sigma_{12} a_t^{(1)})'$. Let

$$X_t = Z_t^{(2)} - (\sigma_{12}/\sigma_{11})\{\theta_2(B)/\phi_2(B)\} a_t^{(1)} = Z_t^{(2)} - \sigma_{11}^{-1}\sigma_{12}\psi_2(B) a_t^{(1)}.$$

Then the model can be written as two independent series, $Z_t^{(1)}$ and X_t, where $\phi_2(B) X_t = \theta_2(B)\varepsilon_t^{(2)}$, $\varepsilon_t^{(2)} = a_t^{(2)} - \sigma_{12} a_t^{(1)}/\sigma_{11}$ and $\text{var}\,(\varepsilon_t^{(2)}) = \sigma_{2:1}$.

The iterative backforecasting algorithm of McLeod & Sales (1983) can be applied to each one of these models to obtain the estimated innovations $\hat{a}_1$ and $\hat{\varepsilon}_2$ and the initial values $\hat{e}_1$ and $\hat{\xi}$, say, of the series $Z_t^{(1)}$ and X_t. The initial values for $Z_t^{(2)}$ are easily obtained from X_t as $Z_t^{(2)} = X_t + \sigma_{11}^{-1}\sigma_{12}\psi_2(B)\hat{a}_t^{(1)}$ and the innovations for the series $Z_t^{(2)}$ are obtained from $\hat{a}_t^{(2)} = \hat{\varepsilon}_t^{(2)} + \sigma_{11}^{-1}\sigma_{12}\hat{a}_t^{(1)} a_t$. Although X_t is not directly observable, it can be calculated from $X_t = Z_t - \sigma_{11}^{-1}\sigma_{12} Y_t$, where $\phi_2(B) Y_t = \theta_2(B)\hat{a}_t^{(1)}$.

3·4. *Calculation of the covariance determinant*

The calculation of the term $|\Omega||\Lambda'\Omega^{-1}\Lambda|$ of (3) is now considered. The inclusion of this term in the likelihood function improves the small properties of the parameter estimators, particularly in models with moving average operators having roots close to the unit circle (Hillmer & Tiao, 1979; McLeod, 1977).

The joint distribution of $\{Z_t^{(1)}\}$ $(t=1-m,\ldots,N)$ and $\{Z_t^{(2)}\}$ $(t=1,\ldots,N)$ is normal so that the likelihood function can also be expressed as

$$L(\beta,\Delta\,|\,Z)=(2\pi)^{-\frac{1}{2}(2N+m)}|\Gamma|^{-\frac{1}{2}}\exp\left(-\tfrac{1}{2}Z'\Gamma^{-1}Z\right),$$

where Γ is the variance covariance matrix of Z. Comparing this expression with (3) it follows that $|\Omega|\,|\Lambda'\Omega^{-1}\Lambda|=|\Gamma|$. The calculation of this determinant may be quite laborious so that it would be desirable to obtain an adequate approximation which is computationally attractive.

Now Γ can be represented as a partitioned matrix with (i,j)th entry in the (g,h)th partition given by $\sigma_{gh}\Gamma_{gh}(i-j)$, where

$$\sigma_{gh}\Gamma_{gh}(i-j)=E(Z_{t-i}^{(g)}Z_{t-j}^{(h)})=\sigma_{gh}\Sigma\,\psi_r^{(g)}\psi_{r+(i-j)}^{(h)},\quad \psi_h(B)=\theta_h(B)/\phi_h(B),\quad \psi_{-r}^{(h)}=0$$

for $r>0$. Let A_g $(g=1,2)$ be $(m+N)\times R$ and $N\times R$ matrices with $R>N+m$ and with (i,j)th entry, $\psi_{j-i}^{(g)}$. Then the (g,h)th partition of Γ is approximated by $\sigma_{gh}A_gA_h'$. The error of this approximation is $O(\lambda^{R-N-m})$ where $|\lambda|<1$ and λ corresponds to the largest root of the polynomial $\phi_1(B)\phi_2(B)=0$. Now using a well-known result for the determinant of a partitioned matrix it follows that

$$|\Gamma|=|\sigma_{11}\Gamma_{11}|\,|\sigma_{22}\Gamma_{22}-(\sigma_{12}^2/\sigma_{11})\Gamma_{21}\Gamma_{11}\Gamma_{12})|.$$

Hence using the approximation to $|\Gamma|$ it follows that

$$\sigma_{22}\Gamma_{22}-(\sigma_{12}/\sigma_{11})\Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}=\sigma_{22}A_2\{1_R-\rho^2A_1'(A_1A_1')^{-1}A_1\}A_2'+O(\lambda^{R-N-M}),$$

where $\rho^2=\sigma_{21}^2/(\sigma_{11}\sigma_{22})$. It can be shown that the determinant of this last expression is given by

$$\begin{aligned}|\sigma_{22}A_2\{1_R-\rho^2A_1'(A_1A_1')^{-1}A_1\}A_2'|&=\sigma_{22}^N(1-\rho^2)^{N-1}|A_2A_2'|(1-\rho^2a)\\&\simeq\sigma_{22}^N(1-\rho^2)^N|\Gamma_{22}|,\end{aligned}$$

where $0<a<1$.

The determinant of Γ can then be approximated by

$$|\Gamma|=\sigma_{11}^{N+m}\sigma_{22}^N(1-\rho^2)^N|\Gamma_{11}|\,|\Gamma_{22}|=|\Delta|^N\sigma_{11}^m|\Gamma_{11}|\,|\Gamma_{22}|.$$

The error introduced in the above approximation is negligible for moderate to large values of $m+N$. It can be shown that the exact expression of $|\Gamma|$ for the contemporaneous bivariate first-order autoregression is given by

$$|\Gamma|=|\Gamma_{11}|\,|\Gamma_{22}|\sigma_{11}^{N+m}\sigma_{22}^N(1-\rho^2)^{N-1}\left\{1-\frac{\rho^2\phi_1^2(\phi_1-\phi_2)(\phi_1-\phi_2^{2m-1})}{(1-\rho^2)(1-\phi_1\phi_2)^2}\right\}.$$

Taking logarithms, the logarithms of the last factor is $O(1)$ whereas the logarithm of the rest of the expression is $O(N+m)$.

3·5. *Algorithm to calculate the likelihood function*

The algorithm given in this section calculates the approximate likelihood function of the contemporaneous model when the series have different sample sizes, using the simplifications discussed in the previous sections. Then it can be shown that apart from an arbitrary constant, the logarithm of the concentrated likelihood function can be written as

$$\log L(\beta) = -\{\tfrac{1}{2}(N+m)\} \log \{S_{m1}/(N+m)\} - (\tfrac{1}{2}N) \log (S_{m2}/N), \qquad (5)$$

where S_{m1} and S_{m2} are the modified sum of squares defined by

$$S_{m1} = S_1\{M_{(N+m)}(\beta_1, \Delta)\}^{-1/(N+m)}, \quad S_{m2} = S_2\{M_N(\beta_2, \Delta)\}^{-1/N},$$

$$S_1 = (\hat{a}^{(1)}_{-m-Q})^2 + \ldots + (\hat{a}^{(1)}_N)^2, \quad S_2 = (\hat{\varepsilon}^{(2)}_{1-Q})^2 + \ldots + (\hat{\varepsilon}^{(2)}_N)^2,$$

$\varepsilon^{(2)}_t$ is an auxiliary series given by

$$\phi_2(B)[Z^{(2)}_t - \sigma_{12}\sigma_{11}^{-1}\{\theta_2(B)/\phi_2(B)\}a^{(1)}_t] = \theta_2(B)\varepsilon^{(2)}_t$$

and $M_n(\beta, \sigma)$ represents a term involving the determinant of the covariance matrix of a univariate process with parameters β, σ and n observations (McLeod, 1977).

For given parameter values β, $\log L(\beta)$ of (5) can be calculated using the following algorithm:

1. Using $\{Z^{(1)}_t\}$, for $t = 1-m, \ldots, N$, and $\beta_1 = (\phi^{(1)}_1, \ldots, \phi^{(1)}_p, \theta^{(1)}_1, \ldots, \theta^{(1)}_q)$ obtain the residuals $a^{(1)}_t$ and S_{m1} using the algorithm given by McLeod & Sales (1983).
2. Using $\{Z^{(2)}_t\}$, for $t = 1, \ldots, N$, and $\beta_2 = (\phi^{(2)}_1, \ldots, \phi^{(2)}_p, \theta^{(2)}_1, \ldots, \theta^{(2)}_q)$ obtain the residuals $[a^{(2)}_t]$.
3. Calculate initial estimated values for σ_{11} and σ_{12} using the residuals $\hat{a}^{(1)}_t$ and $[a^{(2)}_t]$.
4. Calculate the auxiliary series Y_t given by $\theta_2(B)\hat{a}^{(1)}_t = \phi_2(B)Y_t$.
5. Using $\{Z^{(2)}_t - (\sigma_{11}^{-1}\sigma_{21})Y_t\}$, for $t = 1, \ldots, N$ and β_2 obtain the auxiliary residuals $\varepsilon^{(2)}_t$ and S_{m2}.
6. Calculate $\log L(\beta)$ using (5).

4. LARGE-SAMPLE PROPERTIES OF THE ESTIMATORS

4·1. *Distribution of $\tilde{\beta}$*

In order to obtain the asymptotic distribution of the parameter of $\tilde{\beta}$, the maximum likelihood estimate for the model, it is observed that for large values of $m+N$ the log likelihood function can be approximated by (Hillmer & Tiao, 1979)

$$l(\beta, \Delta) = -\tfrac{1}{2}m \log \sigma_{11} - \tfrac{1}{2} N \log |\Delta| - \frac{1}{2\sigma_{11}} \sum_{t=1-m}^{0} \{a^{(1)}_t\}^2 + \tfrac{1}{2} \sum_{t=1}^{N} a'_t\Delta^{-1}a_t. \qquad (6)$$

THEOREM 1. *The asymptotic distribution of $\sqrt{N}(\tilde{\beta} - \beta)$ is multivariate with mean zero and covariance matrix*

$$V_{\tilde{\beta}} = \begin{bmatrix} (\sigma_{11} + \sigma_{11}^{-1}m_N)I_{11} & \sigma^{12}I_{12} \\ \sigma^{21}I_{21} & \sigma^{22}I_{22} \end{bmatrix}^{-1},$$

where $\Delta^{-1} = (\sigma^{ij})$, $m_N = \lim m/N$ and

$$I_{gh}\begin{bmatrix} \gamma_{V_gV_h}(i-j) & \gamma_{V_gU_h}(i-j) \\ \gamma_{U_gV_h}(i-j) & \gamma_{U_gU_h}(i-j) \end{bmatrix}$$

with $\gamma_{U_hV_h}(i-j) = E(U^{(h)}_{t-i}V^{(h)}_{t-j})$; $U^{(h)}_t$, $V^{(h)}_t$ are auxiliary series defined as $\phi_h(B)V^{(h)}_t = -a^{(h)}_t$, $\theta_h(B)U^{(h)}_t = a^{(h)}_t$. The other terms in the matrix are given by similar expressions. In practical applications it can be assumed that $m_N = m/N$.

Proof. Under the assumptions of normality, stationarity and invertibility the log likelihood satisfies the usual regularity conditions. It follows from taking a Taylor expansion of $\partial l/\partial\beta$ about β, the true value, and evaluating at $\tilde{\beta}$ that

$$0=\frac{\partial l}{\partial\beta}+\frac{\partial^2 l}{\partial\beta\,\partial\beta^{\mathrm{T}}}(\tilde{\beta}-\beta)+O_p(1).$$

Further, from the derivations given in the Appendix, it follows that

$$(\tilde{\beta}-\beta)=N^{-1}V_{\tilde{\beta}}\partial l/\partial\beta+O_p(1/N).$$

Apart from terms $O_p(1/N)$, linear combinations of $(\tilde{\beta}-\beta)$ are an average of martingale differences. Normality then follows from the martingale central limit theorem (Billingsley, 1961). □

It is of practical interest to compare the asymptotic distribution of $\tilde{\beta}$ with the asymptotic distribution of the maximum likelihood estimator of β, $\hat{\beta}$ say, obtained using only the N pairs of $\{(Z_t^{(1)}, Z_t^{(2)})\}'$ $(t=1,\ldots,N)$. F. Camacho in his Ph.D. thesis showed that $\surd N(\hat{\beta}-\beta)$ is asymptotically normally distributed with mean zero and covariance matrix

$$V_{\hat{\beta}}=\begin{bmatrix}\sigma^{11}I_{11} & \sigma^{12}I_{12}\\ \sigma^{21}I_{21} & \sigma^{22}I_{22}\end{bmatrix}^{-1}$$

with I_{gh} given by Theorem 1. It follows that $V_{\tilde{\beta}}^{-1}-V_{\hat{\beta}}^{-1}$ is a positive-semidefinite matrix and therefore the estimator $\tilde{\beta}$ has smaller variance than the estimator $\hat{\beta}$.

4·2. *An illustrative example*

The bivariate first-order autoregression is used as an example to examine the possible gain in efficiency of the estimators obtained using all the available information, $\tilde{\beta}$, compared to the estimators obtained using only part of the data, $\hat{\beta}$.

The model is given by $Z_t^{(h)}=\phi_h Z_{t-1}^{(h)}+a_t^{(h)}$ $(h=1,2)$, where $a_t=\{a_t^{(1)}, a_t^{(2)}\}'$ is a series of independent normal vectors with mean zero and covariance matrix Δ.

It can be shown that $V_{\hat{\beta}}$ is given by

$$V_{\hat{\beta}}=\frac{(1-\rho^2)}{N(1-\rho_\phi^2)}\begin{bmatrix}(1-\phi_1^2) & \dfrac{\rho^2(1-\phi_1^2)(1-\phi_2^2)}{(1-\phi_1\phi_2)}\\ \dfrac{\rho^2(1-\phi_1^2)(1-\phi_2^2)}{(1-\phi_1\phi_2)} & (1-\phi_2^2)\end{bmatrix},$$

where $\rho_\phi^2=a\rho^4$, $a=(1-\phi_1^2)(1-\phi_2^2)/(1-\phi_1\phi_2)^2$. Now $V_{\tilde{\beta}}$ is given by

$$V_{\tilde{\beta}}=\frac{(1-\rho^2)}{N\{1+m_N(1-\rho^2)-\rho_\phi^2\}}\begin{bmatrix}(1-\phi_1^2) & \dfrac{\rho^2(1-\phi_1^2)(1-\phi_2^2)}{(1-\phi_1\phi_2)}\\ \dfrac{\rho^2(1-\phi_1^2)(1-\phi_2^2)}{(1-\phi_1\phi_2)} & \{1+m_N(1-\rho^2)\}(1-\phi_2^2)\end{bmatrix}.$$

First, the effect of an unequal sample estimation approach on the estimation of ϕ_2 is considered. The efficiency of $\hat{\phi}_2$ relative to $\tilde{\phi}_2$ is given by

$$(1-a\rho^4)\{1+m_N(1-\rho^2)\}/\{1+m_N(1-\rho^2)-a\rho^4\}$$

which converges to $(1-a\rho^4)$ as m_N increases. If $\phi_1=\phi_2$, for example, then $a=1$ and the efficiency tends to $1-\rho^4$. If a tends to zero the values of the efficiency converges to 1 so that no gain is expected in the estimation of ϕ_2.

To study the possible gain in efficiency in the estimation of ϕ_1 two cases are considered. In the first case it is assumed that $\hat{m}_N < \rho^2(1-a\rho^2)/(1-\rho^2)$ so that the joint estimator of ϕ_1 and ϕ_2, $\hat{\phi}_1$ say, using the N common pairs of observations of the series has smaller asymptotic variance than the univariate estimator, $\bar{\phi}_1$, obtained using only the $m+N$ observations of the Z_{1t}. The relative efficiency of $\hat{\phi}_1$ with respect to $\tilde{\phi}_1$ in this case is given by $(1-a\rho^4)/\{1+m_N(1-\rho^2)-a\rho^4\}$.

In the second case, $m_N > \rho^2(1-a\rho^2)/(1-\rho^2)$, the asymptotic variances of $\bar{\phi}_1$ and $\tilde{\phi}_1$ are compared. The relative efficiency of $\bar{\phi}_1$ with respect to $\tilde{\phi}_1$ is given by

$$\{1+m_N(1-\rho^2)-\rho^2\}/\{1+m_N(1-\rho^2)-a\rho^4\}.$$

When ρ^2 is large, the gain in efficiency can be quite substantial. For example, if $\rho=0\cdot9$ and $m_N=0\cdot6$ then the relative efficiencies for the alternative estimator of $\hat{\phi}_1$ for $a=0\cdot0$, $0\cdot2$, $0\cdot4$, $0\cdot6$, $0\cdot8$ and $1\cdot0$ are, respectively, $0\cdot27$, $0\cdot31$, $0\cdot36$, $0\cdot42$, $0\cdot52$ and $0\cdot66$.

4·3. *Distribution of $\tilde{\mu}$ and $\tilde{\Delta}$*

In this section the asymptotic distribution of $\tilde{\mu}$ and $\tilde{\Delta}$ obtained by maximizing the likelihood function (3) is given. Throughout this section $\tilde{\Delta}$ denotes the vector $\tilde{\Delta}=(\tilde{\sigma}_{11}, \tilde{\sigma}_{21}, \tilde{\sigma}_{12}, \tilde{\sigma}_{22})$. The following theorem gives the distribution of $\tilde{\mu}$.

THEOREM 2. *The asymptotic distribution of $\sqrt{N}(\tilde{\mu}-\mu)$ is multivariate normal with mean zero and covariance matrix*

$$V_\mu = 1/(1+m_N)\begin{bmatrix} \sigma_{11}/C_1^2 & \sigma_{11}/(C_1C_2) \\ \sigma_{11}/(C_1C_2) & \{1+m_N(1-\rho^2)\}\sigma_{22}/C_2^2 \end{bmatrix},$$

where $m_N=\lim m/N$ and $C_h=\phi_h(1)/\theta_h(1)$ $(h=1,2)$. Furthermore, it is independent of the asymptotic distribution of $\sqrt{N}(\tilde{\beta}-\beta)$.

Proof. The proof follows the same lines as the proof of Theorem 1 and observing from the Appendix that, as $N\to\infty$,

$$-\lim N^{-1}E(\partial^2 l/\partial\mu\partial\mu') = \begin{bmatrix} C_1 & 0 \\ 0 & C_1 \end{bmatrix}\begin{bmatrix} \sigma^{11}+m_N\sigma_{11} & \sigma^{12} \\ \sigma^{21} & \sigma^{22} \end{bmatrix}\begin{bmatrix} C_1 & 0 \\ 0 & C_2 \end{bmatrix},$$

$$-\lim N^{-1}E(\partial^2 l/\partial\beta\partial\mu)=0, \quad -\lim N^{-1}E(\partial^2 l/\partial\Delta\partial\mu)=0.$$

This implies that the Fisher information matrix of (β, μ) is block diagonal, which shows that $\tilde{\beta}$ and $\tilde{\mu}$ are asymptotically independent. Normality is shown as in Theorem 1. □

It is interesting to compare the asymptotic variance of $\tilde{\mu}_2$ with that of $\hat{\mu}_2$ obtained using only the N pairs of common observations. The relative efficiency of $\hat{\mu}_2$ with respect to $\tilde{\mu}_2$ is given by $1-m_N\rho^2/(1+m_N)$.

This shows that, in general, there is a gain in efficiency in the estimation of μ_2 when all the available information is used. On the other hand, the variance of $\tilde{\mu}_1$ is the same as the asymptotic variance of $\bar{\mu}_1$ obtained using the $N+m$ observations of $Z_t^{(1)}$, so that no gain in efficiency is expected in the estimation of μ_1. The following theorem gives the distribution of $\tilde{\Delta}$.

THEOREM 3. *The asymptotic distribution of $\sqrt{N}(\tilde{\Delta}-\Delta)$ is normal with mean zero and covariance matrix, $V_{\tilde{\Delta}}$, given by $V_{\tilde{\Delta}}=\Delta\otimes\Delta(1_4+P)Q$, where the first row of Q is*

$$\left((1+m_N)^{-1}, -\frac{m_N\sigma_{12}}{\sigma_{11}(1+m_N)}, -\frac{m_N\sigma_{12}}{\sigma_{11}(1+m_N)}, -\frac{m_N\sigma_{12}^2}{\sigma_{11}^2(1+m_N)}\right),$$

the second, third and fourth rows are the same as those of 1_4 and P is a permutation matrix such that $P^2=1_4$, $P(\Delta^{-1}\otimes\Delta^{-1})=(\Delta^{-1}\otimes\Delta^{-1})P$. Furthermore, the distribution is independent of $\tilde{\beta}$ and $\tilde{\mu}$.

Proof. The proof is similar to the proof of Theorem 2. In particular, the covariance matrix $V_{\tilde{\Delta}}$ is given by the inverse of the information matrix $I=-\lim E(\partial^2 l/\partial\Delta\,\partial\Delta)/N$. Now

$$-\frac{1}{N}E\left(\frac{\partial^2 l}{\partial\sigma_{11}^2}\right)=\frac{\sigma^{11}\sigma^{12}}{2}\{1+m_N(1-\rho^2)^2\},$$

$$-\frac{1}{N}E\left(\frac{\partial^2 l}{\partial\sigma_{ij}\,\partial\sigma_{11}}\right)=\frac{\sigma^{1i}\sigma^{1j}}{2}\quad((i,j)\neq(1,1)),$$

$$-\frac{1}{N}E\left(\frac{\partial^2 l}{\partial\sigma_{ij}\,\partial\sigma_{rs}}\right)=\frac{1}{2}\left(\frac{\sigma^{si}\sigma^{jr}+\sigma^{hi}\sigma^{js}}{2}\right)\quad((i,j)\neq(1,1),(r,s)\neq(1,1)).$$

Therefore, the information matrix can be written as:

$$I_\Delta=\Delta^{-1}\otimes\Delta^{-1}(1+P)/4+m_N qq',$$

where P is a permutation matrix and $q'=(\sigma^{11}(1-\rho^2),\,0,\,0,\,0)$. □

The explicit expressions for the variances are

$$N\operatorname{var}(\tilde{\sigma}_{11})=2\sigma_{11}^2/\{1+m_N\},\tag{7}$$

$$N\operatorname{var}(\tilde{\sigma}_{22})=2\sigma_{22}^2\{1-m_N\rho^4/(1+m_N)\},\tag{8}$$

$$N\operatorname{var}(\tilde{\sigma}_{21})=\sigma_{11}\sigma_{22}\{1+\rho^2(1-m_N)/(1+m_N)\}.\tag{9}$$

It is interesting to observe that the asymptotic variance of $\tilde{\sigma}_{11}$ is the same as the asymptotic variance of $\bar{\sigma}_{11}$ obtained using the $m+N$ observations of $Z_t^{(1)}$. On the other hand, it can be seen from (8) and (9) that there is a gain in efficiency of the estimators $\tilde{\sigma}_{22}$ and $\tilde{\sigma}_{21}$ compared with the estimators $\hat{\sigma}_{22}$ and $\hat{\sigma}_{21}$ obtained using the N pairs of common observations.

4·4. *On the estimation of $\tilde{\beta}$ and $\tilde{\Delta}$*

As was mentioned before, for moderate or large sample size the estimators $\tilde{\beta}$ and $\tilde{\Delta}$ can be obtained by maximizing the approximate likelihood function given in (6). To obtain the estimator for β the nonlinear system of equations $\partial l/\partial\beta=0$ need to be solved.

An iterative procedure like that of Newton–Raphson is required to obtain $\tilde{\beta}$. In particular, $\bar{\beta}_1$ and $\bar{\beta}_2$, the estimators of β_1 and β_2 obtained using the $N+m$ observations of $Z_t^{(1)}$ and the N observations of $Z_t^{(2)}$, respectively, can be used as initial values for

$$\beta_{k+1}=\beta_k-V_{\tilde{\beta}}(\partial l/\partial\beta)/N,\tag{10}$$

where the last term is evaluated at $\beta=\bar{\beta}_k$. It can be shown that just one iteration of (10) with $\bar{\beta}=(\bar{\beta}_1,\bar{\beta}_2)$ as initial values produces asymptotically efficient estimators.

The estimator for Δ is obtained by solving the equation $\partial l/\partial\Delta=0$ which yields

$$-\tfrac{1}{2}N\Delta^{-1}-\frac{m}{2\sigma_{11}}\begin{bmatrix}1&0\\0&0\end{bmatrix}+\sum_{t=1-m}^{0}\frac{a_{1t}}{2\sigma_{11}}\begin{bmatrix}1&0\\0&0\end{bmatrix}+\tfrac{1}{2}\Delta^{-1}\left(\sum_{t=1}^{N}a_ta_t'\right)\Delta^{-1}=0.$$

This system of equations can be solved explicitly:

$$\tilde{\sigma}_{11}=(SS_{11}+S_{10})/(N+m),\tag{11}$$

$$\tilde{\sigma}_{j1} = \tilde{\sigma}_{1j} = SS_{1j} \bigg/ \left(N + m - \frac{S_{10}}{\tilde{\sigma}_{11}} \right) \quad (j > 1), \tag{12}$$

$$\tilde{\sigma}_{ij} = \frac{1}{N} \left\{ SS_{22} + \frac{\tilde{\sigma}_{ij}\tilde{\sigma}_{lj}}{\tilde{\sigma}_{11}} \left(\frac{S_{10}}{\tilde{\sigma}_{11}} - m \right) \right\} \quad (i, j > 1), \tag{13}$$

where $S_{10} = \Sigma\, a_{1t}^2$ and $(SS_{ij}) = \Sigma\, a_t a_t'$.

So, given $\bar{\beta} = (\bar{\beta}_1' \bar{\beta}_2')$, initial estimators for Δ can be obtained using (11) to (13), replacing a_{ht} for $\bar{a}_{ht}$, the residuals being obtained from the univariate estimation.

ACKNOWLEDGEMENTS

The authors would like to thank Dr B. Bragança Pereira and Dr S. Power for helpful comments on this work. Partial support was provided by National Science and Engineering Research Council of Canada Operating Grants awarded to each of the authors.

APPENDIX

Derivative evaluations

In this Appendix the first and second derivatives of the log likelihood function given by (6) are derived. The first derivatives of $l(\beta, \Delta)$ with respect to β are

$$\frac{\partial l}{\partial \beta_{1j}} = -\frac{1}{\sigma_{11}} \sum_{t=1-m}^{0} a_t^{(1)} W_{1t-j}^{(j)} - \sum_{t=1}^{N} a_t' \Delta^{-1} \begin{bmatrix} W_{1t-j}^{(j)} \\ 0 \end{bmatrix}, \quad \frac{\partial l}{\partial \beta_{2j}} = -\sum_{t=1}^{N} a_t' \Delta^{-1} \begin{bmatrix} 0 \\ W_{2t-j}^{(j)} \end{bmatrix},$$

where W stands for V or U depending on whether β is ϕ or θ and the auxiliary series V and U are defined by $\phi_h(B) V_t^{(h)} = -a_t^{(h)}$, $\theta_h(B) U_t^{(h)} = a_t^{(h)}$, for $h = 1, 2$. The second derivatives of $l(\beta, \Delta)$ with respect to β are

$$-\frac{\partial^2 l}{\partial \beta_{1i} \partial \beta_{1j}} = \frac{1}{\sigma_{11}} \sum_{t=1-m}^{0} \left(a_t^{(1)} \frac{\partial W_{1t-j}^{(j)}}{\partial \beta_{1i}} + W_{1t-i}^{(i)} W_{1t-j}^{(j)} \right)$$

$$+ \sum_{t=1}^{N} \left(a_t' \Delta^{-1} \begin{bmatrix} \dfrac{\partial W_{1t-j}^{(j)}}{\partial \beta_{1i}} \\ 0 \end{bmatrix} + \sigma^{11} W_{1t-i}^{(i)} W_{1t-j}^{(j)} \right)$$

$$= \left(\frac{m}{\sigma_{11}} + N\sigma^{11} \right) \gamma_{W_1}(i) \gamma_{W_2}(j)^{(i-j)} + O_p\{(m+N)^{\frac{1}{2}}\},$$

$$-\frac{\partial^2 l}{\partial \beta_{1i} \partial \beta_{1j}} = \frac{1}{\sigma_{11}} \sum_{t=1-m}^{0} \left(a_t^{(1)} \frac{\partial W_{1t-j}^{(j)}}{\partial \beta_{1i}} + W_{1t-i}^{(i)} W_{1t-j}^{(j)} \right)$$

$$-\frac{\partial^2 l}{\partial \beta_{2i} \partial \beta_{2j}} = \sum_{t=1}^{N} \left(a_t' \Delta^{-1} \begin{bmatrix} 0 \\ \dfrac{\partial W_{2t-j}^{(j)}}{\partial \beta_{2i}} \end{bmatrix} + W_{2t-i}^{(i)} W_{2t-j}^{(j)} \right) = N\sigma^{22} \gamma_{W_2}(i) \gamma_{W_2}(j)^{(i-j)} + O_p(\sqrt{N}).$$

It is easy to see that var $(\partial^2 l / \partial \beta_{1i} \partial \beta_{2j}) = O(N+m)$ and var $(\partial^2 l / \partial \beta_{gi} \partial \beta_{hj}) = O(N)$, for $g = 2$, $h = 1$, 2, justifying the second expression for each term. Similarly,

$$\frac{\partial l}{\partial \mu_2} = \frac{1}{\sigma_{11}} \sum_{t=1-m}^{0} a_t^{(1)} C_1 - \sum_{t=1}^{N} a_t' \Delta^{-1} \begin{bmatrix} C_1 \\ 0 \end{bmatrix}, \quad \frac{\partial l}{\partial \mu_2} = -\sum_{t=1}^{N} a_t' \Delta^{-1} \begin{bmatrix} 0 \\ C_2 \end{bmatrix},$$

$$\frac{\partial^2 l}{\partial \mu_1^2} = N C_1^2 (\sigma^{11} + m_N / \sigma_{11}), \quad \frac{\partial^2 l}{\partial \mu_2 \partial \mu_h} = -N\sigma^{2h} C_2 C_h,$$

where $C_h = \phi_h(1) / \theta_h(1)$ for $h = 1, 2$.

References

BILLINGSLEY, P. (1961). The Lindeberg-Lévy theorem for martingales. *Proc. Am. Math. Soc.* **12**, 788-92.

BOX, G. E. P. & JENKINS, G. M. (1976). *Time Series Analysis: Forecasting and Control*, 2nd ed. San Francisco: Holden Day.

CIPRA, T. (1984). Simple correlated ARMA processes. *Math. Oper. Statist., Ser. Statist.* **15**, 513-25.

CAMACHO, F., MCLEOD, A. I. & HIPEL, K. W. (1985). Contemporaneous autoregressive-moving average modeling in water resources. *Water Resources Bull.* **21**, 709-21.

GRANGER, C. W. J. (1969). Investigating causal relations by econometric models and cross spectral methods. *Econometrica* **37**, 424-38.

HALL, A. D. & NICHOLLS, D. F. (1980). The evaluation of exact minimum likelihood estimates for VARMA models. *J. Statist. Comput. Simul.* **10**, 251-62.

HILLMER, S. C. & TIAO, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models. *J. Am. Statist. Assoc.* **74**, 602-7.

JENKINS, G. M. & ALAVI, A. S. (1981). Some aspects of modelling and forecasting multivariate time series. *J. Time Series Anal.* **2**, 1-47.

MCLEOD, A. I. (1977). Improved Box-Jenkins estimators. *Biometrika* **64**, 531-4.

MCLEOD, A. I. & HOLANDA SALES, P. R. (1983). Algorithm AS 191: An algorithm for approximate likelihood calculation of ARMA and seasonal ARMA models. *Appl. Statist.* **32**, 211-23.

MORIARTY, M. & SALOMON, G. (1980). Estimation and forecast performance of a multivariate time series model of sales. *J. Market Res.* **17**, 558-64.

NELSON, C. R. (1976). Gains in efficiency from joint estimation of systems of autoregressive-moving average processes. *J. Econometrics* **4**, 331-48.

NICHOLLS, D. F. & HALL, A. D. (1979). The exact likelihood of multivariate autoregressive-moving average models. *Biometrika* **66**, 259-64.

PIERCE, D. A. (1977). Relationships and the lack thereof—between economic time series with special reference to money and interest rates. *J. Am. Statist. Assoc.* **72**, 11-26.

PIERCE, D. A. & HAUGH, L. D. (1979). The characterization of instantaneous causality: A comment. *J. Econometrics* **10**, 257-9.

PIERCE, D. A. & HAUGH, L. D. (1977). Causality in temporal systems: characterizations and survey. *J. Econometrics* **5**, 265-93.

RISAGER, F. (1980). Simple correlated autoregressive process. *Scand. J. Statist.* **7**, 49-60.

RISAGER, F. (1981). Model checking of simple correlated autoregressive processes. *Scand. J. Statist.* **8**, 137-53.

TIAO, G. C. & BOX, G. E. P. (1981). Modeling multiple time series with applications. *J. Am. Statist. Assoc.* **76**, 802-16.

UMASHANKAR, S. & LEDOLTER, J. (1983). Forecasting with diagonal multiple time series models: An extension of univariate models. *J. Market Res.* **20**, 58-63.

WHITTLE, P. (1963). On the fitting of multivariate autoregressions, and the approximate canonical factorization of the spectral density matrix. *Biometrika* **50**, 129-34.

[*Received August* 1985. *Revised May* 1986]