Biometrika (1977), **64**, 3 *pp.* 531–34
Printed in Great Britain

Improved Box–Jenkins estimators

BY A. I. McLEOD
Department of Statistics, University of Waterloo, Ontario

SUMMARY

An easily implemented modification of the estimation procedure given by Box & Jenkins (1970) for the autoregressive-moving average time series model is suggested and some new results on covariance determinants are given. The proposed modification provides a closer approximation to the exact maximum likelihood estimators. Simulation experiments which demonstrate the effectiveness of this modification are presented.

Some key words: Autoregressive-moving average time series; Covariance determinant; Time series estimation.

1. INTRODUCTION

Consider the autoregressive-moving average time series model of order (p, q),

$$\phi(B)\, w_t = \theta(B)\, a_t \quad (t = 1, \ldots, n),$$

where $\phi(B) = 1-\phi_1 B - \ldots - \phi_p B^p$, $\theta(B) = 1-\theta_1 B - \ldots - \theta_q B^q$, B is the backshift operator and a_t is a sequence of independent, normally distributed disturbances with mean zero and variance σ^2. This model is considered admissible if and only if it is stationary and invertible, or, equivalently, $\phi(z)\,\theta(z) = 0$ has all roots outside the unit circle. Then the likelihood of $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \sigma^2$ can be shown to be (Box & Jenkins, 1970, p. 273)

$$L(\Phi, \Theta, \sigma^2) \propto \sigma^{-n} |M_n^{(p,q)}(\Phi, \Theta)|^{\frac{1}{2}} \exp\left\{-\frac{S(\Phi, \Theta)}{2\sigma^2}\right\},$$

where $\Phi = (\phi_1, \ldots, \phi_p)$, $\Theta = (\theta_1, \ldots, \theta_q)$, $S(\Phi, \Theta)$ is the unconditional sum of squares function of Box & Jenkins (1970) and

$$M_n^{(p,q)}(\Phi, \Theta) = \sigma^2\{\Gamma_n^{(p,q)}(\Phi, \Theta, \sigma^2)\}^{-1},$$

where $\Gamma_n^{(p,q)}(\Phi, \Theta, \sigma^2)$ is the covariance matrix of the series, that is the $n \times n$ matrix

$$\Gamma_n^{(p,q)}(\Phi, \Theta, \sigma^2)$$

has (i,j)th element γ_{i-j}, where $\gamma_k = E(w_t w_{t-k})$. The unconditional sum of squares function,

$$S(\Phi, \Theta) = \sum_{t=-\infty}^{n} \hat{a}_t^2,$$

where $\hat{a}_t = E(a_t | w_1, \ldots, w_n; \Phi; \Theta)$, is readily calculated using the back forecasting algorithm of Box & Jenkins (1970, pp. 215–20).

In general, the quantity $|M_n^{(p,q)}(\Phi, \Theta)|$ cannot be readily calculated and so it is suggested that we replace it by $m_{p,q}(\Phi, \Theta)$, where $m_{p,q}(\Phi, \Theta) = \lim |M_n^{(p,q)}(\Phi, \Theta)|$ as $n \to \infty$. This limit clearly exists and is nonzero since the sequence $|M_n|^{-1}$ is bounded and nondecreasing. The approximate likelihood is then

$$L_m(\Phi, \Theta, \sigma^2) \propto \sigma^{-n}\{m_{p,q}(\Phi, \Theta)\}^{\frac{1}{2}} \exp\left\{-\frac{S(\Phi, \Theta)}{2\sigma^2}\right\}. \tag{1}$$

For any covariance stationary time series, w_t, it is easily shown that

$$|\Gamma_n| = \prod_{k=0}^{n-1} \sigma^2(k),$$

where $|\Gamma_n|$ is the covariance determinant of n consecutive observations, $\sigma^2(0) = \gamma_0$ and $\sigma^2(k)$ ($k = 1, 2, \ldots$) is the mean squared error of prediction of w_t using the minimum mean squared error linear predictor $\phi_{k,1} w_{t-1} + \ldots + \phi_{k,k} w_{t-k}$. For autoregressive-moving average models it can be shown that $\sigma^2(k) = \sigma^2 + O(r^k)$, where $0 < r < 1$. It follows that

$$|M_n^{(p,q)}(\Phi, \Theta)| = m_{p.q}(\Phi, \Theta) + O(r^n) \tag{2}$$

and that the relative error of the approximation in (1) to the exact likelihood is $O(r^n)$. For autoregressive models the relative error of this approximation will be zero since

$$m_{p,0}(\Phi) = |M_n^{(p,0)}(\Phi)|$$

provided that $n \geqslant p$.

The more standard modified maximum likelihood approach, summarized by Hannan (1970, § VI. 6), drops the factor $|M_n^{(p,q)}(\Phi, \Theta)|$ altogether. The relative error of this approximation to the exact likelihood is $O(1)$. For small-sample estimation the more accurate approximation to the likelihood given by (1) can be expected to be useful.

2. General procedure

It was shown by Finch (1960) that

$$\log m_{p,q}(\Phi, \Theta) = -\frac{1}{\pi} \iint_{\|z\| \leqslant 1} \left\| \frac{\phi'(z)}{\phi(z)} - \frac{\theta'(z)}{\theta(z)} \right\|^2 d\mu. \tag{3}$$

It follows that

$$\log m_{p,q}(\Phi, \Theta) = \log m_{p,0}(\Phi) + \log m_{q,0}(\Theta) + h_{p,q}(\Phi, \Theta),$$

where

$$h_{p,q}(\Phi, \Theta) = \frac{2}{\pi} \iint_{\|z\| \leqslant 1} \frac{\phi'(z)\,\theta'(\bar{z})}{\phi(z)\,\theta(\bar{z})}\, d\mu.$$

Now consider the autoregressive model of order $p+q$ with autoregressive operator

$$\phi^*(B) = \phi(B)\,\theta(B).$$

Then it follows from (3) that

$$\log m_{p+q,0}(\Phi^*) = \log m_{p,0}(\Phi) + \log m_{q,0}(\Theta) - h_{p,q}(\Phi, \Theta),$$

where $\Phi^* = (\phi_1^*, \ldots, \phi_{p+q}^*)$, and hence

$$m_{p,q}(\Phi, \Theta) = \frac{m_{p,0}^2(\Phi)\, m_{q,0}^2(\Theta)}{m_{p+q,0}(\Phi^*)}. \tag{4}$$

If $q = 0$, it is well known (Box & Jenkins, 1970, p. 275) that for $n \geqslant p$,

$$|M_n^{(p,0)}(\Phi)| = |M_p^{(p,0)}(\Phi)|$$

and that (Pagano, 1973) the matrix $M_p^{(p,0)}(\Phi)$ has (i,j)th element

$$\sum_{k=0}^{\min(i,j)} (\phi_{i-k-1}\,\phi_{j-k-1} - \phi_{p+1+k-i}\,\phi_{p+1+k-j}),$$

where $\phi_0 = -1$. Hence $m_{p,0}(\Phi) = |M_p^{(p,0)}(\Phi)|$ and so to compute $m_{p,q}(\Phi, \Theta)$ all that is necessary is to calculate the determinant of three easily obtained positive-definite matrices of orders p, q and $p+q$.

It is convenient to work with the modified sum of squares function

$$S_m(\Phi, \Theta) = S(\Phi, \Theta)\{m_{p,q}(\Phi, \Theta)\}^{-1/n}.$$

Minimizing the modified sum of squares function $S_m(\Phi, \Theta)$ is equivalent to maximizing the maximized approximate likelihood

$$\max\{L_m(\Phi, \Theta, \sigma^2): \sigma^2 \text{ varying}\}.$$

Furthermore if the observations are standardized by the transformation,

$$z_t^{(m)} = z_t/\{m_{p,q}(\Phi, \Theta)\}^{1/(2n)},$$

then the unconditional sum of squares of the standardized data is the suggested modified sum of squares and the nonlinear least squares method proposed by Box & Jenkins (1970, pp. 208–42, 504–5) can be used. Alternatively, the modified sum of squares can be readily minimized using the algorithm of Powell (1964).

Table 1. *Estimates of airline data model*

	Modified sum of squares method	Unconditional sum of squares method
$\hat{\theta}_1$	0·4018	0·3959
$\hat{\Theta}_1$	0·5569	0·6135
$\hat{\sigma}^2$	0·00135	0·00134

The following example is instructive since it illustrates that the proposed modified sum of squares method involves only a slight increase in the computations and that the estimates do sometimes differ somewhat from the unconditional sum of squares method, even in apparently large samples. Box & Jenkins (1970, Chapter 9) suggest the multiplicative seasonal moving average model $w_t = (1-\theta_1 B)(1-\Theta_1 B^{12})\, a_t$, where w_t ($t = 1, \ldots, 131$) is the differenced and seasonally differenced series of some logged monthly airline passenger data. This model was fitted using the modified sum of squares and the unconditional sum of squares methods and the calculations took respectively 14·4 and 8·5 seconds of processor time on a Honeywell 6060 Computer System using initial values $\theta_1 = \Theta_1 = 0{\cdot}0$. The two estimates of Θ_1, given in Table 1, differ by about 80 % of the estimated standard error.

3. Simulation experiments

A simulation study was done to examine the effects of the proposed modification in small samples. For series of length $n = 30, 60, 100$ autoregressive-moving average models of orders (1, 0), (0, 1), (1, 1) and the multiplicative model of order $(0, 1)\,(0, 1)_4$ defined by

$$w_t = (1-\theta_1 B)(1-\Theta_1 B^4)\, a_t$$

were simulated and fitted using the unconditional sum of squares and the modified sum of squares methods. The relative efficiencies shown in Table 2, which were determined by the ratio of the empirical mean squared errors in one thousand simulations of the models, are illustrative. Based on Table 2 and simulation results for other parameter values, the following general conclusions were made:

(i) the modified method was uniformly more efficient;
(ii) the improvement was greatest in the shorter series;
(iii) the improvement was more significant in the moving average parameters.

It is also of interest that K. M. Kang, in an unpublished simulation study of the (0, 1) model, found that in small samples an exact maximum likelihood procedure gave significantly smaller mean squared errors than the unconditional sum of squares method.

Table 2. *Percentage relative efficiency of unmodified versus modified method*

	(1, 0)	(0, 1)	(1, 1)		$(0, 1)(0, 1)_4$	
n	$\phi_1 = \frac{1}{2}$	$\theta_1 = \frac{1}{2}$	$\phi_1 = \frac{1}{2}$	$\theta_1 = -\frac{1}{2}$	$\theta_1 = \frac{1}{2}$	$\Theta_1 = \frac{1}{2}$
30	93	82	91	85	80	50
60	97	91	96	91	88	59
100	99	96	96	95	95	77

4. Concluding remarks

The autoregressive-moving average model is said to be redundant if the equations $\phi(z) = 0$ and $\theta(z) = 0$ have at least one root in common. Assuming that the model is not redundant, it follows from (2) and (4) that

$$|M_n^{(p,q)}(\Phi, \Theta)| = \frac{|I_{p,q}(\Phi, \Theta)|}{|I_p(\Phi)|^2 \, |I_q(\Theta)|^2 \, |J_{p,q}(\Phi, \Theta)|^2} + O(r^n), \tag{5}$$

where $0 < r < 1$, $I_{p,q}(\Phi, \Theta)$ is the large-sample information matrix per observation of Φ and Θ in the autoregressive-moving average model, $I_p(\Phi)$ and $I_q(\Theta)$ are the $p \times p$ and the $q \times q$ submatrices corresponding to Φ and Θ alone and $J_{p,q}(\Phi, \Theta)$ is the $(p+q) \times (p+q)$ matrix obtained by adjoining the matrices with (i, j)th elements respectively θ_{i-j}, and ϕ_{i-j}, $\theta_i = 0$ unless $i = 0, \ldots, q$, $\theta_0 = -1$ and $\phi_i = 0$ unless $i = 0, \ldots, p$. Equation (5) provides a more accurate approximation to $|M_n^{(p,q)}(\Phi, \Theta)|$ than that used by Box & Jenkins (1970, pp. 257, 283).

I would like to thank the referees for suggestions which improved this paper and the Mathematics Faculty at the University of Waterloo for providing computing facilities.

References

Box, G. E. P. & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control.* San Francisco: Holden-Day.

Finch, P. D. (1960). On the covariance determinants of moving average and autoregressive models. *Biometrika* **47**, 423–6.

Hannan, E. J. (1970). *Multiple Time Series.* New York: Wiley.

Pagano, M. (1973). When is an autoregressive process stationary? *Commun. Statist.* **1**, 533–44.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer J.* **7**, 155–62.

[*Received July* 1976. *Revised February* 1977]