Biometrika (1986), **73**, 1, *pp.* 217–21
Printed in Great Britain

Fractional time series modelling

BY W. K. LI
Department of Statistics, University of Hong Kong, Hong Kong

AND A. I. MCLEOD
Department of Statistical and Actuarial Sciences, University of Western Ontario, London, Canada N6A 5B9

SUMMARY

Aspects of model building using fractionally differenced autoregressive-moving average processes are discussed. An algorithm for approximate maximum likelihood estimation is outlined and the large-sample distribution of the maximum likelihood estimates is derived. The large-sample distribution of the residual autocorrelations is also derived and a modified portmanteau test statistic is obtained for checking model adequacy.

Some key words: Fractional differencing; Long-memory time series; Maximum likelihood; Parsimony; Portmanteau test; Residual autocorrelation.

1. INTRODUCTION

The fractional autoregressive-moving average model of order (p, q) for a stationary time series Z_t $(t = 1, 2, \ldots)$ with mean μ may be written, using operator notation,

$$\phi(B)\,\Delta(B)\,(Z_t - \mu) = \theta(B)\,a_t, \tag{1}$$

where

$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p, \quad \theta(B) = 1 - \theta_1 B - \ldots - \theta_q B^q,$$

$$\Delta(B) = (1-B)^d = \sum_{i=0}^{\infty} \binom{d}{i} (-1)^i B^i$$

and a_t is a sequence of independent normal variables with mean zero and variance σ^2. This general model was introduced by Hosking (1981). Granger & Joyeux (1980) discussed the $(0, 0)$ model. As shown by Granger & Joyeux and Hosking, for stationarity and invertibility, $|d| < \frac{1}{2}$ and all roots of $\phi(B)\,\theta(B) = 0$ are outside the unit circle. It is also assumed that $\phi(B)$ and $\theta(B)$ contain no common factor in order to insure unique identifiability of the parameters.

Hosking (1981) showed that $\gamma_l = O(l^{2d-1})$ as $l \to \infty$, where $\gamma_l = \text{cov}\,(Z_t, Z_{t-l})$. Thus for $0 < d < \frac{1}{2}$, $\Sigma \gamma_l$ diverges and so the time series may be said to have a long memory component. Previously, various researchers in stochastic hydrology, see for example McLeod & Hipel (1978), have discussed a similar type of model for the modelling of annual geophysical time series. It is also of interest that Granger (1980) showed that long memory time series may also arise with economic data.

Previously Whittle (1954) and Hannan (1970, p. 245) have suggested that a large number of parameters may sometimes be needed in the usual autoregressive-moving average model. In these circumstances the fractional autoregressive-moving average model may possibly produce a more parsimonious fit.

2. Approximate maximum likelihood

The fractional model of order (p, q) may be approximated by the model

$$\phi(B)\,\Delta_r(B)\,(Z_t-\mu) = \theta(B)\,a_t, \tag{2}$$

where

$$\Delta_r(B) = \sum_{i=0}^{r}\binom{d}{i}(-1)^i B^i.$$

It follows from the Kakeya–Enström Theorem (Henrici, 1974, p. 462) that $\Delta_r(B) = 0$ has all roots outside the unit circle. Hence (2) is stationary for any $r \geqslant 0$. Moreover, for r large enough and d not larger than $\frac{1}{2}$ the difference in models (1) and (2) can be made negligible.

The backforecasting method of Box & Jenkins (1976, Ch. 7) for autoregressive-moving average estimation can be conveniently adapted using the algorithm of McLeod & Sales (1983). The backward and forward equations may be written

$$\Delta_r(B)\,\tilde{Z}_t = b_t, \quad \phi(B)\,b_t = \theta(B)\,a_t, \quad \Delta_r(F)\,\tilde{Z}_t = c_t, \quad \phi(F)\,c_t = \theta(F)\,e_t,$$

where $\tilde{Z}_t = Z_t-\mu$, $F = B^{-1}$ and e_t is a sequence of independent normal variables with mean zero and variance σ^2. Hosking (1981, 1984) and the unpublished University of Western Ontario Ph.D. dissertation by W. K. Li contain further details on the simulation and estimation of fractional autoregressive-moving average models.

3. Asymptotic distribution

The asymptotic distribution of the maximum likelihood estimator is given in Theorem 1 below. Let $\lambda = (\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q,d)$ denote the true parameter values. Let $\hat{\lambda}$ and λ denote the maximum likelihood estimate and arbitrary admissible parameter values respectively. It is necessary to assume that μ is known. Without loss of generality it is further assumed that $\mu = 0$ and $\sigma^2 = 1$.

Then it may be shown (Box & Jenkins, 1976, p. 237), $\partial a_t/\partial\phi_j = u_{t-j}$ and $\partial a_t/\partial\theta_j = v_{t-j}$, where u_t and v_t are the auxiliary autoregressions $\phi(B)\,u_t = -a_t$ and $\theta(B)\,v_t = a_t$. Furthermore, $d_t = \partial a_t/\partial d = (\log\nabla)\,a_t$, where $\nabla = 1-B$ and $\log\nabla = -B-\frac{1}{2}B^2-\frac{1}{3}B^3-\ldots$.

Since $\operatorname{var}(d_t) = \Sigma\,1/l^2 = \frac{1}{6}\pi^2$, it follows that d_t is stationary. Using the technique of Box & Jenkins (1976, p. 240) we can easily show that the Fisher large-sample information matrix per observation, I, has (i, j)th entry given by $E\{(\partial a_t/\partial\lambda_i)\,(\partial a_t/\partial\lambda_j)\}$. Hence

$$I = \left[\begin{array}{c|c} I_{p,q} & J \\ \hline J' & \pi^2/6 \end{array}\right],$$

where $J = [\gamma_{ud}(0),\ldots,\gamma_{ud}(p-1),\ \gamma_{vd}(0),\ldots,\gamma_{vd}(q-1)]$ and $I_{p,q}$ is the usual information matrix of autoregressive-moving average process on $(\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q)$. It may be shown that, for $l \geqslant 0$,

$$\gamma_{ud}(l) = \sum_{i=0}^{\infty}\phi_i'/(l+i+1), \quad \gamma_{vd}(l) = \sum_{i=0}^{\infty}\theta_i'/(l+i+1),$$

where $\phi^{-1}(B) = \Sigma\,\phi_i' B^i$ and $\theta^{-1}(B) = \Sigma\,\theta_i' B^i$. Note that the information on d is independent of the parameter value. In general, $I_{p,q}$ may be obtained using the method of McLeod (1984).

Theorem 1. *The asymptotic distribution of* $\sqrt{n}(\hat{\lambda}-\lambda)$ *is normal with mean zero and covariance matrix* I^{-1}.

Proof. The log likelihood is approximately $L = -\frac{1}{2}S(\lambda)$, where

$$S(\lambda) = \sum_{t=1-Q}^{n} \dot{a}_t^2,$$

for some large enough Q. Following Pierce (1971, eqn (4.2.9)),

$$n^{-1}\,\partial L/\partial\lambda_i \to 0, \quad n^{-1}\,\partial^2 L/(\partial\lambda_i\,\partial\lambda_j) \to I_{ij},$$

in probability and

$$n^{-1}\,\partial^3 L/(\partial\lambda_i\,\partial\lambda_j\,\partial\lambda_k) = O_p(1).$$

Hence, by the standard Taylor series technique, $\hat{\lambda}$ is consistent. Similarly (Pierce, 1971), asymptotic normality may be established. □

The above result holds when μ is known. When the mean μ is estimated by the maximum likelihood method the situation is more complicated and a similar result has not yet been obtained. Alternatively, the series may be centred by the sample mean, but then, as pointed out by the referee, the approximation is of order $n^{d-\frac{1}{2}}$ rather than the usual $n^{-\frac{1}{2}}$. However, the simulation experiments reported in the next paragraph suggested that the theorem still provides a good approximation.

Bias and the mean squared error of the proposed method of estimation in § 2 has been studied by simulation. The (0, 0) model was considered with $\mu = 0$. One hundred replications of length 200 were generated for each of the cases $d = 0{\cdot}1$, $0{\cdot}2$, $0{\cdot}3$ and $0{\cdot}4$, with truncation point $r = 50$. The series were generated using an exact generation technique (Hosking, 1984). The results are summarized in Table 1. The cases μ known and μ estimated by the sample mean are also compared. If the mean μ is known the procedure is very effective. The bias is negligible and the mean squared error is very close to the theoretical value $6/(\pi^2 \times 200) = 0{\cdot}00304$. When the series is centred by the sample mean a small negative bias is introduced and the mean squared error increases. Nevertheless the increases appear to be quite acceptable. Note that, as found by Hosking (1984), all the mean squared errors are far smaller than those reported using other methods, for example, that of Janacek (1982).

Table 1. *Empirical bias and mean squared error of* $\hat{d}$ *in* 100 *replications of length* 200

	Mean known		Centred by $\bar{Z}$	
d	Bias	Mean squared error	Bias	Mean squared error
0·1	−0·0024	0·00309	−0·0199	0·00386
0·2	0·0026	0·00337	−0·0187	0·00400
0·3	0·0059	0·00389	−0·0210	0·00476
0·4	0·0115	0·00343	−0·0255	0·00458

4. Residual autocorrelation

Box & Jenkins (1976, Ch. 8) pointed out that it is important to check the assumption of independence of the a_t series by using the residual autocorrelation function,

$$\hat{r}_a(l) = \Sigma\,\hat{a}_t\,\hat{a}_{t-l}/\Sigma\,\hat{a}_t^2,$$

to test for significant autocorrelation. Note that the residuals are obtained from the algorithm of § 2. It is assumed that μ is known.

THEOREM 2. *For any fixed* $M \geqslant 1$, $\sqrt{n}\{\hat{r}_a(1), \ldots, \hat{r}_a(M)\}$ *is asymptotically normal with mean zero and covariance matrix* $1_M - XI^{-1}X'$, *where* 1_M *is the* $M \times M$ *identity matrix and*

$$X = (-\phi'_{i-j} \vdots \theta'_{i-j} \vdots K)_M,$$

where $K' = \{1, \frac{1}{2}, \ldots, 1/(M-1)\}$.

The proof of this theorem is a straightforward extension of the method used by McLeod (1978).

It may also be shown that $1_M - XI^{-1}X'$ is approximately idempotent with rank $M-p-q-1$ for M large enough and hence the usual modified portmanteau test statistic Q_M (Ljung & Box, 1978) is approximately χ^2-distributed with $M-p-q-1$ degrees of freedom. It is also useful to test for significant autocorrelation in the residuals by using estimated standard deviations obtained from Theorem 2. For example, in the $(0,0)$ model $\text{var}\{\hat{r}_a(l)\} = \{1-6/(\pi l)^2\}/n$.

The small-sample behaviour of $\text{var}\{\hat{r}_a(1)\}$ and Q_{20} for the $(0,0)$ model was examined by simulation. Using only the first 50 terms of $\Delta(B)$, 500 simulations of series of length $n = 250$ were generated for $d = 0{\cdot}0$, $0{\cdot}1$, $0{\cdot}2$, $0{\cdot}3$ and $0{\cdot}4$. The parameter d was estimated using the method of § 2. The number of rejections using the portmanteau test at the 0·05 level ($Q_{20} > 30{\cdot}14$), the mean of Q_{20} and its standard deviation are shown in Table 2. The number of rejections in each case does not differ significantly from 25. The mean of Q_{20} is always quite close to its asymptotic value of 19. The observed standard deviations of $\hat{r}_a(1)$ are all very close to 0·0396 as predicted.

Table 2. *Number of times* $Q_{20} > 30{\cdot}14$, *empirical average value of* Q_{20} *and empirical standard deviation of* $\hat{r}_a(1)$ *in* 500 *simulation of* $(0,0)$ *model with* $n = 250$

d	$Q_{20} > 30{\cdot}14$	Average Q_{20}	St dev $\{\hat{r}_a(1)\}$
0·0	25	$19{\cdot}12 \pm 0{\cdot}28$	0·0394
0·1	23	$18{\cdot}92 \pm 0{\cdot}28$	0·0389
0·2	19	$18{\cdot}91 \pm 0{\cdot}29$	0·0384
0·3	23	$18{\cdot}37 \pm 0{\cdot}29$	0·0394
0·4	27	$18{\cdot}91 \pm 0{\cdot}29$	0·0412

ACKNOWLEDGEMENTS

The authors would like to thank a referee and Carlos Jiménez for helpful comments.

REFERENCES

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

GRANGER, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *J. Econometrics* **14**, 227–8.

GRANGER, C. W. J. & JOYEUX, R. (1980). An introduction to long-memory time series models and fractional differencing. *J. Time Series Anal.* **1**, 15–29.

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

HENRICI, P. (1974). *Applied and Computational Complex Analysis*, **1**. New York: Wiley.

HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165–76.

HOSKING, J. R. M. (1984). Modelling persistence in hydrological time series using fractional differencing. *Water Resources Res.* 20, 1898–908.

JANACEK, G. J. (1982). Determining the degree for time series via the log spectrum. *J. Time Series Anal.* **3**, 177–83.

LJUNG, G. M. & BOX, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303.

MCLEOD, A. I. (1978). On the distribution of residual autocorrelations in Box–Jenkins models. *J. R. Statist. Soc.* B **40**, 296–302.

MCLEOD, A. I. (1984). Duality and other properties of multiplicative seasonal autoregressive moving-average models. *Biometrika* **80**, 207–11.

MCLEOD, A. I. & HIPEL, K. W. (1978). Preservation of the rescaled adjusted range. *Water Resources Res.* **14**, 491–518.

MCLEOD, A. I. & SALES, P. R. HOLANDA (1983). An algorithm for approximate likelihood calculations of ARMA and seasonal ARMA models. *Appl. Statist.* **32**, 211–23.

PIERCE, D. A. (1971). Least squares estimation in the regression model with autoregressive-moving average errors. *Biometrika* **58**, 299–312.

WHITTLE, P. (1954). The statistical analysis of a seiche record. *J. Marine Res.* **13**, 76–100.

[*Received December* 1983. *Revised July* 1985]