Biometrika (1969), 56, 3, p. 595
Printed in Great Britain

On the inverse of the covariance matrix of a first order moving average

BY PAUL SHAMAN
Carnegie-Mellon University

SUMMARY

Three distinct expressions are noted for the elements of the inverse of the covariance matrix $\mathbf{\Sigma}_T$ of T successive observations from a stationary first order moving average process. An observation of Arato (1961) shows that the problem of finding the exact inverse can be reduced to the problem of finding the covariance determinant $|\mathbf{\Sigma}_T|$. Approximate expressions for $\mathbf{\Sigma}_T^{-1}$ are then obtained by approximating various forms of the determinant. The resulting approximate inverses are all the same and coincide with the expression usually employed.

1. INTRODUCTION

The stationary first order moving average process $\{x_t\}$ with mean zero is defined by

$$x_t = \epsilon_t + \beta\epsilon_{t-1} \quad (t=0, \pm 1, \ldots), \tag{1}$$

where $\{\epsilon_t\}$ is a set of uncorrelated random variables with mean zero and variance v. We assume $|\beta| < 1$. The process $\{x_t\}$ has covariance sequence $\sigma(0) = v(1+\beta^2)$, $\sigma(1) = v\beta$, $\sigma(r) = 0$ $(r>1)$ and $\sigma(-r) = \sigma(r)$. It is also convenient to set $\sigma^2 = v(1+\beta^2)$ and $\rho = \beta/(1+\beta^2)$, so that $\sigma(0) = \sigma^2$ and $\sigma(1) = \sigma^2\rho$. Then $|\rho| < \frac{1}{2}$,

$$\beta = \frac{1-\sqrt{(1-4\rho^2)}}{2\rho} = \left\{\frac{1+\sqrt{(1-4\rho^2)}}{2\rho}\right\}^{-1}, \tag{2}$$

$$v = \tfrac{1}{2}\sigma^2\{1+\sqrt{(1-4\rho^2)}\}. \tag{3}$$

The stationary normal first order moving average is a model which is commonly studied in time series analysis. Problems of statistical inference arise when an observed time series of finite length is available. Estimation of β has been discussed by Whittle (1953), Durbin (1959) and Walker (1961). Walker (1967) has considered the problem of testing the hypothesis that the observations arise from an autoregressive scheme of order one versus the alternative of a moving average of order one, and vice versa.

Even when $\{x_t\}$ is normal, exact statistical inference for the model (1) is particularly difficult because the dimensionality of the minimal sufficient set of statistics is equal to the number of observations in the time series. One proof is by Arato (1961).

When $x_1, \ldots, x_T$ satisfy (1) the covariance matrix of $\mathbf{x} = \{x_1, \ldots, x_T\}'$ is

$$\mathbf{\Sigma}_T = v\{(1+\beta^2)\,\mathbf{I}_T + \beta\mathbf{A}_T\},$$

where $\mathbf{I}_T$ is the $T\times T$ identity matrix and $\mathbf{A}_T$ is $T\times T$. Here $\mathbf{A}_T$ has ones on the diagonals one element removed from the main diagonal, and has zeros elsewhere. An approximate expression for $\mathbf{\Sigma}_T^{-1}$ which is commonly used is $\mathbf{B}_T = \{b_{ij}\} = \{b_{|i-j|}\}$, with

$$\begin{aligned} b_{ij} &= \frac{1}{v(1-\beta^2)}(-\beta)^{|j-i|} \\ &= \frac{(-1)^{|j-i|}}{\sigma^2\sqrt{(1-4\rho^2)}}\left\{\frac{1-\sqrt{(1-4\rho^2)}}{2\rho}\right\}^{|j-i|} \quad (i,j=1,\ldots,T); \end{aligned} \tag{4}$$

see, for example, Whittle (1954, §§ 2·2, 2·5) and Hannan (1960, p. 47). The matrix $\mathbf{B}_T$ is the exact inverse of v times a matrix with $1, 1+\beta^2, \ldots, 1+\beta^2, 1$ down the diagonal, β's above and below the diagonal and zeros everywhere else. The first order stationary autoregressive process $y_t+\beta y_{t-1} = \epsilon_t$ has covariance generating function $v\{(1+\beta z)(1+\beta z^{-1})\}^{-1}$, and the generating function for the sequence $\{b_r\}$ $(r=0, \pm 1, \ldots)$ is $\{v(1+\beta z)(1+\beta z^{-1})\}^{-1}$. That is, the exact inverse of $\mathbf{B}_T$ is proportional to the inverse of the covariance matrix of $\mathbf{y} = \{y_1, \ldots, y_T\}'$ generated by the first order stationary autoregression. Under normality an approximate log likelihood for $\mathbf{x}$ is (Walker, 1967)

$$L(v,\beta) = -\tfrac{1}{2}T\log(2\pi v) - \tfrac{1}{2}T\sum_{r=-(T-1)}^{T-1} b_r C_r,$$

where

$$C_r = \frac{1}{T}\sum_{t=1}^{T-r} x_t x_{t+r} \quad (r \geqslant 0)$$

and $C_{-r} = C_r$. The approximate inverse $\mathbf{B}_T$ is obtained by ignoring the end effects due to the presence of a finite series.

Arato (1961) has observed that $\mathbf{\Sigma}_T = \{\sigma_{ij}\}$ has the exact inverse $\mathbf{\Sigma}_T^{-1} = \{\sigma^{ij}\}$ given by

$$\sigma^{ij} = (-\sigma^2\rho)^{j-i}\,|\mathbf{\Sigma}_{i-1}|\,|\mathbf{\Sigma}_{T-j}|/|\mathbf{\Sigma}_T| \quad (j \geqslant i). \tag{5}$$

The minor in the cofactor of σ_{ij} $(j \geqslant i)$ is evaluated by using Laplace's expansion.

In § 2 we give σ^{ij} in three distinct forms. These exact expressions are then approximated in § 3. The resulting approximate inverses are all the same and agree with (4).

2. The exact inverse

The determinant $|\mathbf{\Sigma}_i|$ satisfies the second order difference equation

$$|\mathbf{\Sigma}_i| = v(1+\beta^2)\,|\mathbf{\Sigma}_{i-1}| - v^2\beta^2|\mathbf{\Sigma}_{i-2}| \quad (i=2,3,\ldots), \tag{6}$$

where $|\mathbf{\Sigma}_0| = 1$ and $|\mathbf{\Sigma}_1| = v(1+\beta^2)$. The roots of the associated polynomial equation $x^2 - v(1+\beta^2)x + v^2\beta^2 = 0$ are $x_1 = v$ and $x_2 = v\beta^2$, and the well-known result

$$|\mathbf{\Sigma}_i| = \frac{v^i(1-\beta^{2i+2})}{1-\beta^2} = v^i(1+\beta^2+\ldots+\beta^{2i}) \quad (i=0,1,\ldots) \tag{7}$$

follows. Then $\mathbf{\Sigma}_T^{-1} = \{\sigma^{ij}\}$ is, by (5),

$$\begin{aligned} \sigma^{ij} &= \frac{(-\beta)^{j-i}}{v(1-\beta^2)}\cdot\frac{(1-\beta^{2i})\{1-\beta^{2(T-j+1)}\}}{1-\beta^{2(T+1)}} \\ &= (-\beta)^{j-i}\frac{\{1+\beta^2+\ldots+\beta^{2(i-1)}\}\{1+\beta^2+\ldots+\beta^{2(T-j)}\}}{v(1+\beta^2+\ldots+\beta^{2T})} \quad (j \geqslant i). \end{aligned} \tag{8}$$

This result is also given by Whittle (1963, p. 75, Exercise 3). From (2), (3) and (7),

$$|\mathbf{\Sigma}_i| = \frac{\sigma^{2i}}{\sqrt{(1-4\rho^2)}}\left[\left\{\frac{1+\sqrt{(1-4\rho^2)}}{2}\right\}^{i+1} - \left\{\frac{1-\sqrt{(1-4\rho^2)}}{2}\right\}^{i+1}\right]$$
$$= \frac{\sigma^{2i}\rho^{i+1}}{\sqrt{(1-4\rho^2)}}\left[\left\{\frac{1-\sqrt{(1-4\rho^2)}}{2\rho}\right\}^{-(i+1)} - \left\{\frac{1+\sqrt{(1-4\rho^2)}}{2\rho}\right\}^{-(i+1)}\right] \quad (i=0,1,\ldots). \tag{9}$$

Expansion of the numerator of the first line of (9) yields

$$|\mathbf{\Sigma}_i| = \frac{\sigma^{2i}}{\sqrt{(1-4\rho^2)}}\cdot\frac{1}{2^{i+1}}\sum_{j=0}^{i+1}\binom{i+1}{j}(1-4\rho^2)^{\frac{1}{2}j}\{1-(-1)^j\}$$
$$= \frac{\sigma^{2i}}{2^i}\sum_{j=0}^{[\frac{1}{2}i]}\binom{i+1}{2j+1}(1-4\rho^2)^j$$
$$= \frac{\sigma^{2i}}{2^i}\sum_{k=0}^{[\frac{1}{2}i]}(-4\rho^2)^k\sum_{j=k}^{[\frac{1}{2}i]}\binom{i+1}{2j+1}\binom{j}{k}$$
$$= \sigma^{2i}\sum_{k=0}^{[\frac{1}{2}i]}(-1)^k\rho^{2k}\binom{i-k}{k} \quad (i=0,1,\ldots). \tag{10}$$

The last equality in (10) is a consequence of

$$\frac{1}{2^{i-2k}}\sum_{j=k}^{[\frac{1}{2}i]}\binom{i+1}{2j+1}\binom{j}{k} = \binom{i-k}{k},$$

which may be shown, although somewhat laboriously, by differentiating k times and then setting $x = 1$ in both sides of

$$\sum_{j=0}^{[\frac{1}{2}i]}\binom{i+1}{2j+1}x^j = \{(1+x^{\frac{1}{2}})^{i+1} - (1-x^{\frac{1}{2}})^{i+1}\}/(2x^{\frac{1}{2}}).$$

Another expression for $\mathbf{\Sigma}_T^{-1} = \{\sigma^{ij}\}$ is, therefore,

$$\sigma^{ij} = \frac{(-\rho)^{j-i}}{\sigma^2}\cdot\frac{\sum_{k=0}^{[\frac{1}{2}(i-1)]}\binom{i-1-k}{k}(-1)^k\rho^{2k}\sum_{l=0}^{[\frac{1}{2}(T-j)]}\binom{T-j-l}{l}(-1)^l\rho^{2l}}{\sum_{k=0}^{[\frac{1}{2}T]}\binom{T-k}{k}(-1)^k\rho^{2k}} \quad (j\geqslant i). \tag{11}$$

The expressions for $\mathbf{\Sigma}_T^{-1}$ given by (8) and (11) were obtained via a difference equation approach. Next we find $|\mathbf{\Sigma}_i|$ by using the characteristic roots of $\mathbf{\Sigma}_i$.

As noted above $\mathbf{\Sigma}_i = \sigma^2(\mathbf{I}_i + \rho\mathbf{A}_i)$ and $\mathbf{A}_i$ has roots

$$2\cos\left(\frac{\pi s}{i+1}\right) \quad (s=1,\ldots,i);$$

see, for example, Grenander & Szegö (1958, §5.3). Therefore,

$$|\mathbf{\Sigma}_i| = \sigma^{2i}\prod_{s=1}^{i}\left\{1+2\rho\cos\left(\frac{\pi s}{i+1}\right)\right\} \quad (i=0,1,\ldots) \tag{12}$$

and $\mathbf{\Sigma}_T^{-1} = \{\sigma^{ij}\}$ is given by

$$\sigma^{ij} = \frac{(-\rho)^{j-i}}{\sigma^2}\cdot\frac{\prod_{s=1}^{i-1}\left\{1+2\rho\cos\left(\frac{\pi s}{i}\right)\right\}\prod_{t=1}^{T-j}\left\{1+2\rho\cos\left(\frac{\pi t}{T-j+1}\right)\right\}}{\prod_{s=1}^{T}\left\{1+2\rho\cos\left(\frac{\pi s}{T+1}\right)\right\}} \quad (j\geqslant i). \tag{13}$$

3. THE APPROXIMATE INVERSE

In this section we approximate several expressions for $|\boldsymbol{\Sigma}_i|$, namely (7), (9) and (12). In each case substitution into the right-hand side of (5) yields $\mathbf{B}_T$ in (4). The methods of approximation employed here have been used in studies of serial correlation. A comprehensive review of the theory of serial correlation in a theoretical framework is presented by Anderson (1970, Chapter 6).

Let $y_1, \ldots, y_T$ be independently distributed, each as $N(0,1)$. Then the moment generating function of $Q = 2\sum_{t=1}^{T-1} y_t y_{t+1}$ is

$$\phi_T(\theta) = E(e^{\theta Q}) = |\mathbf{I}_T - 2\theta \mathbf{A}_T|^{-\frac{1}{2}} = \prod_{s=1}^{T}\left\{1 - 4\theta\cos\left(\frac{\pi s}{T+1}\right)\right\}^{-\frac{1}{2}},$$

and we see that $|\boldsymbol{\Sigma}_T| = \sigma^{2T}\phi_T^{-2}(-\frac{1}{2}\rho)$.

Let us approximate (9). The second line involves the factor

$$\begin{aligned}
&\left\{\frac{1-\sqrt{(1-4\rho^2)}}{2\rho}\right\}^{-(i+1)} - \left\{\frac{1+\sqrt{(1-4\rho^2)}}{2\rho}\right\}^{-(i+1)} \\
&= \left\{\frac{1-\sqrt{(1-4\rho^2)}}{2\rho}\right\}^{-(i+1)}\left[1-\rho^{2(i+1)}\left\{\frac{1+\sqrt{(1-4\rho^2)}}{2}\right\}^{-2(i+1)}\right] \\
&= \left\{\frac{1-\sqrt{(1-4\rho^2)}}{2\rho}\right\}^{-(i+1)}\left[1-\rho^{2(i+1)}\left\{1+2(i+1)\sum_{k=1}^{\infty}\frac{\Gamma(2i+2+2k)\rho^{2k}}{\Gamma(2i+3+k)k!}\right\}\right]. \qquad (14)
\end{aligned}$$

Dixon (1944) has used this summation formula, which is due to Laplace. If ρ^2 is small, the second summand inside the brackets on the last line of (14) is small. We, therefore, approximate $|\boldsymbol{\Sigma}_i|$ by

$$\frac{\sigma^{2i}}{\sqrt{(1-4\rho^2)}}\left\{\frac{1-\sqrt{(1-4\rho^2)}}{2\rho^2}\right\}^{-(i+1)} \quad (i=0,1,\ldots) \qquad (15)$$

and use of (5) and (15) gives the matrix $\mathbf{B}_T$. This method of approximating $|\boldsymbol{\Sigma}_i|$ corresponds to truncation of a moment generating function.

Now let us consider (7). When β^2 is small $|\boldsymbol{\Sigma}_i|$ in this form is approximately (Durbin, 1959)

$$v^i/(1-\beta^2) \quad (i=0,1,\ldots). \qquad (16)$$

But (15) and (16) are equivalent. This second method of approximation may be interpreted as the result of ignoring the end effects due to the presence of an observed time series of finite length. Now we approximate (12), which is

$$|\boldsymbol{\Sigma}_i| = \sigma^{2i}\exp\left[\sum_{s=1}^{i}\log\left\{1+2\rho\cos\left(\frac{\pi s}{i+1}\right)\right\}\right] \quad (i=0,1,\ldots).$$

We smooth the roots by replacing the discrete set $\cos\{\pi s/(i+1)\}$ $(s=1,\ldots,i)$ by a continuous variable. That is,

$$\frac{\pi}{i+1}\left[\sum_{s=1}^{i}\log\left\{1+2\rho\cos\left(\frac{\pi s}{i+1}\right)\right\}+\tfrac{1}{2}\log(1+2\rho)+\tfrac{1}{2}\log(1-2\rho)\right]$$

is a Riemann sum approximating the integral

$$\int_0^{\pi}\log(1+2\rho\cos u)\,du = \pi\log\{1+\sqrt{(1-4\rho^2)}\}/2 \quad (|\rho|<\tfrac{1}{2}).$$

Then $|\mathbf{\Sigma}_i|$ is approximately

$$\sigma^{2i}\exp\left\{\left(\frac{i+1}{\pi}\right)\pi\log\left\{\frac{1+\sqrt{(1-4\rho^2)}}{2}\right\}-\tfrac{1}{2}\log(1-4\rho^2)\right\}$$
$$=\frac{\sigma^{2i}}{\sqrt{(1-4\rho^2)}}\left\{\frac{1+\sqrt{(1-4\rho^2)}}{2}\right\}^{i+1}\quad (i=0,1,\ldots), \qquad (17)$$

which is (15) again. This method of smoothing roots was used by Koopmans (1942) to obtain an approximate density for the first order circular serial correlation, and Dixon (1944) used the method to truncate the moment generating function of the first order circular serial correlation. Rubin (1945) showed that the results of Koopmans and Dixon produced the same approximate density. We have shown that for the problem of approximating $\mathbf{\Sigma}_T^{-1}$ for a first order moving average, three methods are equivalent. These correspond to truncation of a generating function, neglecting the end effects due to presence of a finite time series and smoothing roots.

Whittle (1951, Chapter 4) has also studied the problem of approximating $\mathbf{\Sigma}_T^{-1}$ by modifying the end elements of the time series.

Von Neumann (1941) has given an expansion of the moment generating function of a quadratic form which can be applied here. We have, by expanding $\log(1+u)$ in a Taylor series,

$$|\mathbf{\Sigma}_i| = \sigma^{2i}\exp\left\{2\rho\sum_{s=1}^{i}\cos\left(\frac{\pi s}{i+1}\right)-\frac{(2\rho)^2}{2}\sum_{s=1}^{i}\cos^2\left(\frac{\pi s}{i+1}\right)\right.$$
$$\left.+\ldots+(-1)^{k-1}\frac{(2\rho)^k}{k}\sum_{s=1}^{i}\cos^k\left(\frac{\pi s}{i+1}\right)+\ldots\right\}.$$

Then (von Neumann, 1941)

$$\sum_{s=1}^{i}\cos^k\left(\frac{\pi s}{i+1}\right)=\frac{i+1}{2^k}\left\{\binom{k}{\tfrac{1}{2}k}+2\sum_{h=1}^{\left[\frac{k}{2(i+1)}\right]}\binom{k}{\tfrac{1}{2}k-h(i+1)}\right\}-1\quad (k=0,2,\ldots)$$

and the sums are 0 for k odd. Therefore,

$$|\mathbf{\Sigma}_i| = \sigma^{2i}\exp\left[-(i+1)\sum_{k=1}^{\infty}\frac{\rho^{2k}}{2k}\left\{\binom{2k}{k}+2\sum_{h=1}^{\left[\frac{k}{i+1}\right]}\binom{2k}{k-h(i+1)}\right\}+\sum_{k=1}^{\infty}\frac{(2\rho)^{2k}}{2k}\right]$$
$$=\frac{\sigma^{2i}}{\sqrt{(1-4\rho^2)}}\left\{\frac{1+\sqrt{(1-4\rho^2)}}{2}\right\}^{i+1}\exp\left\{-(i+1)\sum_{k=1}^{\infty}\frac{\rho^{2k}}{k}\sum_{h=1}^{\left[\frac{k}{i-1}\right]}\binom{2k}{k-h(i+1)}\right\}\qquad (18)$$

and we see that smoothing roots is the same as dropping the second factor on the second line of (18).

Let us note the accuracy of the approximation $\mathbf{B}_T$. Let $\mathbf{P}_T=\{p_{ij}\}=\mathbf{\Sigma}_T\mathbf{B}_T$. Then

$$\begin{aligned}
p_{11} &= p_{TT} = 1/(1-\beta^2),\\
p_{1j} &= (-\beta)^{j+1}/(1-\beta^2)\quad (j=2,\ldots,T),\\
p_{ii} &= 1\quad (i=2,\ldots,T-1),\\
p_{Tj} &= (-\beta)^{T+2-j}/(1-\beta^2)\quad (j=1,\ldots,T-1),\\
p_{ij} &= 0\quad (i\neq j,\, i\neq 1,T)
\end{aligned}\qquad (19)$$

and $\mathbf{B}_T\mathbf{\Sigma}_T$ is the transpose of $\mathbf{P}_T$.

I am grateful to the referee for pointing out the result of Whittle (1963). Most of this work was carried out at the Department of Statistics, Stanford University, under a grant from the U.S. Army Research Office.

REFERENCES

ANDERSON, T. W. (1970). *An Introduction to the Statistical Analysis of Time Series*. New York: Wiley.

ARATO, M. (1961). On the sufficient statistics for stationary Gaussian random processes. *Theor. Probability Appl.* **6**, 199–201.

DIXON, W. J. (1944). Further contributions to the problem of serial correlation. *Ann. Math. Statist.* **15**, 119–44.

DURBIN, J. (1959). Efficient estimation of parameters in moving average models. *Biometrika* **46**, 306–16.

GRENANDER U. & SZEGÖ, G. (1958). *Toeplitz Forms and Their Applications*. Berkeley: University of California Press.

HANNAN, E. J. (1960). *Time Series Analysis*. London: Methuen.

KOOPMANS, T. (1942). Serial correlation and quadratic forms in normal variables. *Ann. Math. Statist.* **13**, 14–33.

RUBIN, H. (1945). On the distribution of the serial correlation coefficient. *Ann. Math. Statist.* **16**, 211–15.

VON NEUMANN, J. (1941). Distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **12**, 367–83.

WALKER, A. M. (1961). Large-sample estimation of parameters for moving average models. *Biometrika* **48**, 343–57.

WALKER, A. M. (1967). Some tests of separate families of hypotheses in time series analysis. *Biometrika* **54**, 39–68.

WHITTLE, P. (1951). *Hypothesis Testing in Time Series Analysis*. Uppsala: Almqvist and Wiksell.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat.* **2**, 423–34.

WHITTLE, P. (1954). Appendix 2 to *A Study in the Analysis of Stationary Time Series*, by H. Wold. Uppsala: Almqvist and Wiksell.

WHITTLE, P. (1963). *Prediction and Regulation by Linear Least-Square Methods*. London: English Universities Press.

[*Received April* 1969. *Revised July* 1969]