The Annals of Statistics
1980, Vol. 8, No. 5, 1093–1099

AN INVARIANCE PROPERTY FOR THE MAXIMUM LIKELIHOOD ESTIMATOR OF THE PARAMETERS OF A GAUSSIAN MOVING AVERAGE PROCESS

BY E. J. GODOLPHIN

Royal Holloway College, University of London

It is shown that the estimation procedure of Walker leads to estimates of the parameters of a Gaussian moving average process which are asymptotically equivalent to the maximum likelihood estimates proposed by Whittle and represented by Godolphin.

1. Introduction. Several authors have considered the problem of estimating the parameter vector $B = (\beta_1, \cdots, \beta_q)'$ in a Gaussian moving average $\{X_t\}$ of order q:

$$X_t = \varepsilon_t + \beta_1\varepsilon_{t-1} + \cdots + \beta_q\varepsilon_{t-q}, \tag{1.1}$$

where $\{\varepsilon_t\}(t = 0, \pm 1, \pm 2, \cdots)$ comprises a sequence of independent and normally distributed random variables with expectation zero and a common variance σ^2, and it is assumed that the zeros of the polynomial $\beta(z) = 1 + \beta_1 z + \cdots + \beta_q z^q$ lie strictly outside the unit circle. The earliest reference to the problem appears to be that of Whittle (1953) who established that the maximum likelihood estimator is consistent, efficient and asymptotically normal. Recently Godolphin (1977) has obtained a direct iterative formula which yields the maximum likelihood estimator.

A number of alternative estimation procedures have been proposed in the literature and that due to Walker (1961) is of particular interest since it places the main emphasis on estimating not B directly but $\rho = (\rho_1, \cdots, \rho_q)'$, where, with $\beta_0 = 1$,

$$\rho_k = \Sigma_{i=0}^{q-k} \beta_i\beta_{i+k}/\Sigma_{i=0}^{q} \beta_i^2 \tag{1.2}$$

denotes the parent autocorrelation of lag $k(1 \leqslant k \leqslant q)$. This method is founded on a theorem of Bartlett (1946) which enables consideration of the likelihood estimator of B to be avoided completely, a feature which has the practical virtue of pseudoquadratic convergence which yields a substantial reduction in the number of iterations required for convergence.

Mention has been made however of a theoretical difficulty concerning the efficiency of Walker's method which does not seem to have been resolved in the literature. Walker (1961) pointed out that his estimate of B, which is derived from that of ρ by the Cramér-Wold factorization, has not been shown to be fully

Received March 1977; revised April 1979.

AMS 1970 *subject classifications.* 62M10, 62M99.

Key words and phrases. Autocorrelation function, invariance, maximum likelihood estimation, moving average process, stationary time series.

efficient except when $q = 1$ owing to the intricate nature of the transformed covariance matrix. Both Walker and Hannan (1969) conjectured that the estimator would be efficient, however, and some empirical evidence in support of this for the case $q = 2$ was given by Godolphin (1976).

In the present paper it is shown that this conjecture is correct for general q. Indeed the argument adopted here establishes the stronger result that, asymptotically, Walker's estimator and the maximum likelihood estimator of B actually coincide. It is shown that the iterative equation system for estimating ρ based upon Walker's procedure differs from that derived from the direct maximum likelihood procedure by a quantity which can be made as small as required in a neighbourhood of the solution, provided that the number of observations on the process (1.1) is sufficiently large. This result therefore takes the form of an extension of the invariance theorem for maximum likelihood in the two domains where the natural parametric vectors are B and ρ respectively, which conclusion appears to contradict a remark on Walker's procedure made by Anderson (1975). Moreover, since the transformation from ρ to B is nonsingular, it is possible to adopt the approach of Godolphin (1977, 1978) and obtain the maximum likelihood estimator of B directly from Walker's procedure without invoking the Cramér-Wold factorization, which conclusion is also relevant to the criticism of Walker's method made by Hannan (1969).

In Section 2 Walker's procedure and the maximum likelihood procedure in the form due to Godolphin (1977) are described. The main result is derived in Section 3 and its implications briefly discussed in Section 4.

2. Estimation procedures in the ρ and B domains.

2.1 *Preliminary remarks.* It is assumed that a realization

$$X_1, X_2, \cdots, X_n \tag{2.1}$$

of the moving average process (1.1) is available, where n is large. The sample serial correlation of lag k, r_k, is defined by

$$r_k = \{n\Sigma_{t=1}^{n-k}X_tX_{t+k}\}/\{(n-k)\Sigma_{t=1}^{n}X_t^2\} \quad (k = 1, \cdots, m), \tag{2.2}$$

where m is suitably chosen. Other definitions of the sample serial correlation of lag k are commonly employed (see e.g., the comments of Anderson & Walker, 1964), but they have the same asymptotic distribution as r_k and, furthermore, (2.2) seems to be more suitable on the grounds of bias reduction which is perhaps the main reason why this definition was adopted both by Walker and by the author in the derivation and application of their estimation procedures.

It follows from the work of Whittle (1953), Durbin (1959), Walker (1961), the present author (Godolphin, 1977) and others that data reduction is effectively possible for the estimation problem since near-efficient estimates can be obtained from a small set of sample serial correlations. In typical applications $q < m \ll n$, for example, Durbin and Walker both chose $m = 5$ when applying their procedures with $q = 1$ and $n = 100$ whilst the author has considered $m = 30$ for $q = 2$ and $n = 198$ when the solution is close to the boundary of the invertibility region. Of course it makes little practical difference to values of the estimates if m is permitted to become close to $n - 1$, its maximum possible value, since all coefficients of r_k for large k are negligible (in the case of both estimation procedures to be described below) as must occur with any convergent series which, in the present case, is given below by the representation (2.6) or by representation (2.10). (Strictly speaking $m = n - 1$ and a case has to be made for taking m small and thus discarding statistics which contain information about B.) In deriving asymptotic results in what follows it is assumed that m can be taken as large as required to facilitate the argument.

For simplicity the statistics (2.2) are summarized by the vectors

$$R_1 = (r_1, r_2, \cdots, r_q)' \quad \text{and} \quad R_2 = (r_{q+1}, r_{q+2}, \cdots, r_m)'. \tag{2.3}$$

2.2 *Estimation in the ρ domain.* The asymptotic covariance matrix $W = (w_{ij})$ of $n^{\frac{1}{2}}(R_1' - \rho', R_2')'$ is due to Bartlett (1946) who showed that

$$w_{ij} = \phi_{j-i} + \phi_{j+i} + 2(\rho_i\rho_j\phi_0 - \rho_i\phi_j - \rho_j\phi_i) \quad (i, j = 1, \cdots, m) \tag{2.4}$$

with

$$\phi_k = \phi_{-k} = \Sigma_{h=-\infty}^{\infty}\rho_h\rho_{h+k}. \tag{2.5}$$

It is helpful to partition W into four components W_{ij} $(i,j = 1, 2)$ where W_{11}, W_{22} have dimensions $q \times q$, $(m - q) \times (m - q)$ and are the covariance matrices of $n^{\frac{1}{2}}R_1, n^{\frac{1}{2}}R_2$ respectively whilst $W_{21} = W_{12}'$; then the maximum likelihood estimator of ρ is obtained from the asymptotic distribution of $n^{\frac{1}{2}}(R_1' - \rho', R_2')'$. It is given (Walker, 1961) by the regression equation

$$\rho = R_1 - W_{12}W_{22}^{-1}R_2. \tag{2.6}$$

It may be noted that the right side of (2.6) depends upon ρ through (2.4) and (2.5), consequently equation (2.6) is to be regarded as valid primarily as a final iteration.

2.3 *Estimation in the B domain.* Under Gaussian assumptions, a direct representation for the maximum likelihood estimator of B is given (Godolphin, 1977) by the iterative equation

$$B = H_1R_1 + H_2R_2, \tag{2.7}$$

where $H_1 = (h_{ij})$ is an everywhere nonsingular matrix whose elements are rational functions of B chosen to satisfy $B = H_1\rho$ whilst the coefficients $(H_2)_{ij} = h_{i,q+j}$ are obtained from the simple recurrence relation

$$x_k = x_{-k}(1 \leqslant k \leqslant q), \qquad x_k + \Sigma_{i=1}^{2q}\beta_i^*x_{k-i} = 0(k \geqslant q + 1) \tag{2.8}$$

where β_i^* is the coefficient of z^i in the expansion of

$$\beta(z)^2 = (1 + \beta_1 z + \cdots \beta_q z^q)^2 = 1 + \Sigma_{h=1}^{2q}\beta_h^* z^h. \tag{2.9}$$

Here $h_{ik} = x_k(k = 1, 2, \cdots)$ where $x_0 = -2\beta_i$ for each $i = 1, \cdots, q$.

Godolphin (1977) has shown that the equation system (2.7) can be expressed in the equivalent form

$$(2.10) \qquad \rho = R_1 + AR_2$$

where the elements $(A)_{ij} = a_{i,q+j}$ satisfy, for each $i = 1, \cdots, q$, the recurrence relation (2.8) with

$$(2.11) \quad x_0 = -2\rho_i, \qquad x_k = \delta_{ik}(1 \leqslant k \leqslant q) \qquad \text{and} \quad x_k = a_{ik}(k \geqslant q+1)$$

where $\delta_{ik} = 1(i = k), = 0(i \neq k)$.

It is possible to summarize the results (2.10) and (2.11) in the following way. Let J be a $q \times (q-1)$ matrix with unity occupying the positions $(i, q-i)(i = 1, \cdots, q-1)$ and zero in all other positions—the "mirror image" of I with first column removed. Define the $q \times 2q$ matrix K by

$$(2.12) \qquad K = [J, -2\rho, I]$$

and denote the elements of K also by a_{ij} for $1 \leqslant i \leqslant q$ and $1 - q \leqslant j \leqslant q$. For each $i = 1, \cdots, q$, it follows that any $2q$ consecutive elements in the ith. row of the $q \times (m + q)$ matrix $[K, A]$ satisfy the recurrence relation (2.8): therefore

$$(2.13) \qquad a_{ik} + \Sigma_{j=1}^{2q} \beta_j^* a_{i,k-j} = 0 \qquad (k \geqslant q+1)$$

where β_j^* is defined by (2.9).

3. Asymptotic equivalence of the two methods. In this section it is shown that Walker's procedure and the maximum likelihood estimator of B lead to estimates which differ by an asymptotically negligible quantity. This is achieved by demonstrating that the iterative equation systems (2.6) and (2.10) yield solutions for estimating ρ which are asymptotically equivalent. For this purpose it will be shown that every element of the matrix $A + W_{12}W_{22}^{-1}$ can be made as small as required when m is suitably large. A preliminary result is required.

Let a_{ij} be defined as in (2.13) and γ_k denote the coefficient of z^k in the expansion of

$$(3.1) \qquad \left(1 + \Sigma_{i=1}^q \beta_i z^i\right)^2 \left(1 + \Sigma_{i=1}^q \beta_i z^{-i}\right)^2 \equiv \Sigma_{k=-2q}^{2q} \gamma_k z^k.$$

We first show that for all $j \geqslant 1 - q$ and for each $i = 1, \cdots, q$

$$(3.2) \qquad \Sigma_{k=0}^{4q} a_{i,j+k} \gamma_{k-2q} = 0.$$

However, (3.2) is a consequence of the definitions (2.9) and (3.1) which yield immediately

$$\gamma_k = \gamma_{-k} = \Sigma_{h=0}^{2q-k} \beta_h^* \beta_{h+k}^* \qquad 0 \leqslant k \leqslant 2q$$

and therefore by (2.13)

$$\Sigma_{k=0}^{4q} a_{i,j+k} \gamma_{k-2q} = \Sigma_{h=0}^{2q} \beta_h^* \Sigma_{k=0}^{2q} \beta_k^* a_{i,j+2q+h-k} = 0.$$

This result is of particular relevance to the problem since Wold (1949) demonstrated that the term ϕ_k defined by (2.5) is given by the coefficient of z^k in the expansion of $\{\beta(z)\beta(z^{-1})\}^2/(1+\beta_1^2+\cdots+\beta_q^2)^2$. It follows that

$$\Sigma_{k=0}^{4q} a_{i,j+k}\phi_{k-2q} = 0 \qquad (i = 1,\cdots,q; j > -q). \tag{3.3}$$

Now let $2q \times 2q$ matrices Φ, Θ be defined as follows. Φ is a symmetric Toeplitz matrix with elements $(\Phi)_{ij} = \phi_{j-i}$ $(i,j = 1,\cdots,2q)$, whilst $\Theta = (\theta_{ij})$ is an upper triangular matrix with common diagonal elements given by

$$\begin{aligned} \theta_{i,i+j} &= \phi_{2q-j} \qquad i = 1,\cdots,2q; j = 0,\cdots,2q-1, \\ &= 0 \qquad i = 1,\cdots,2q; j < 0. \end{aligned} \tag{3.4}$$

For simplicity, but without affecting the general point, let m be an odd multiple of q, say $m = (2p+1)q$ where $p \geqslant 2$. Then it follows from (2.4) that W_{22} can be partitioned into p^2 blocks, each of size $2q \times 2q$, as follows:

$$W_{22} = \begin{bmatrix} \Phi & \Theta' & \cdots & 0 \\ \Theta & \Phi & \cdots & 0 \\ 0 & \Theta & \cdots & 0 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & 0 & \cdots & \Phi \end{bmatrix}$$

This matrix is given in a suitable form for the product AW_{22} to be examined, where A is defined in equation (2.11). Let $A = [A_1, A_2, A_3]$ where A_1, A_3 are both of size $q \times 4q$. Then it follows from (3.3) that $[K, A_1][\Theta, \Phi, \Theta']' = 0$. Consequently the leading $q \times 2q$ submatrix in AW_{22} is given by

$$A_1[\Phi, \Theta']' = -K\Theta'. \tag{3.5}$$

However, from (2.12) and (3.4) we have the interesting result

$$K\Theta' = W_{12}^{(2q)}, \tag{3.6}$$

where $W_{12}^{(2q)}$ denotes the leading $q \times 2q$ submatrix of W_{12} defined by (2.4). This can be seen e.g., by partitioning the right side of (3.6) into two components of order $q \times q$, the first of which can be expressed as $(J, -2\rho)\Theta'_{11} + \Theta'_{12}$ and the second simply as Θ'_{11}, where the matrix Θ has here been partitioned into four components of size $q \times q$ in an obvious way.

It evidently follows from (3.3) that premultiplying the $p-2$ central column blocks of W_{22} by A yields a $q \times 2q(p-2)$ matrix consisting entirely of zeros. The final $q \times 2q$ submatrix in AW_{22} is given by $A_3[\Theta, \Phi]'$. However it follows from (3.3) that

$$[A_3, A_4][\Theta, \Phi, \Theta']' = 0,$$

where A_4 is the $q \times 2q$ matrix

$$A_4 = (a_{i,(2p+1)q+j}) \tag{3.7}$$

whose elements are defined recursively by (2.11). Therefore, by invoking (3.5) and

(3.6), we have the expression

$$A + W_{12}W_{22}^{-1} = [0, \cdots, 0, A_4\Theta]W_{22}^{-1}. \tag{3.8}$$

To assess the magnitude of the right-hand side of (3.8), it is necessary to take account of two relevant features. Firstly, it can be seen from (3.7) that all elements of A_4 rapidly approach zero as p becomes large. This follows from (2.13) since the inverse zeros $z_1, \cdots, z_q$ of $\beta(z)$ are strictly less than unity in modulus and therefore

$$a_{i,j} = \Sigma_{i=1}^{q}(B_i + jC_i)z_i^j \sim 0 \text{ for } j \text{ large}$$

(where B_i, C_i are fixed real constants). Secondly, Walker (1961) has pointed out that W_{22} can be considered the covariance matrix of $2pq$ consecutive observations from an invertible moving average process of order $2pq$ with covariance generating function

$$\{\phi_0 + \Sigma_{i=1}^{2q}\phi_i(z^i + z^{-i})\} = \{1 + \Sigma_{i=1}^{q}\rho_i(z^i + z^{-i})\}^2,$$

so W_{22}^{-1} can be regarded as the covariance matrix for $2pq$ observations from a stationary autoregression with the same parameters (see e.g., Shaman, 1975). The largest element in W_{22}^{-1} is the variance of this autoregression which does not depend on p and therefore the elements of $[0, \cdots, 0, \Theta]W_{22}^{-1}$ are bounded above in modulus by a quantity which is also independent of p. Consequently the elements of (3.8) can be made as close to zero as desired by taking p suitably large, from which the result follows.

Note that the term A_4 in (3.8) can be regarded as another example of a time-series "end-effect". This is alone responsible for preventing the two estimation procedures from being equivalent in small samples notwithstanding the mathematical elegance of the connecting equation (3.6).

4. Implications of the equivalence. The invariance property of maximum likelihood, which in the present context states that if $\hat{B}$ is the maximum likelihood estimator of B then $\hat{\rho} = \rho(\hat{B})$ is the maximum likelihood estimator of ρ, is known to apply even though the transformation (1.2) is not one-one and the members of the realization (2.1) are not independent; see for example Zehna (1966). The present result shows rather more than invariance in a straightforward asymptotic sense, however, because the solution of (2.7) is the limiting form of an estimator based on the exact distribution whilst Walker's estimator (2.6) is based upon the asymptotic distribution. Moreover (2.7) is derived on the assumption that the process is Gaussian whilst the derivation of (2.6) requires only that $\{\varepsilon_t\}$ consists of independent and identically distributed random variables with a finite variance (Anderson & Walker (1964)).

Walker's procedure has been criticized by e.g., Hannan (1969) because of its dependence on the Cramér-Wold factorization. This point has been discussed in some detail by the author (Godolphin, 1976). It is interesting, however, that because (2.7) and (2.10) are equivalent then Walker's method can be transformed to the iterative equation in B

$$B = H_1 R_1 - H_1 W_{12} W_{22}^{-1} R_2, \tag{4.1}$$

where H_1 is an everywhere nonsingular matrix chosen to satisfy $B = H_1\rho$. (4.1) is independent of the Cramér-Wold factorization and, for sufficiently large m, yields an efficient estimate of B. On purely practical grounds, however, (4.1) seems to be less useful than (2.7). But (2.6) could conceivably be preferred to (2.10) in certain cases (see Godolphin (1977)).

Another consequence of (3.6) and (3.8) is the following formula for the limiting value of the covariance matrix of Walker's estimator (2.6):

$$\lim_{m\to\infty}\left(W_{11} - W_{12}W_{22}^{-1}W_{12}'\right) = W_{11} + A^{(2q)}\Theta K', \tag{4.2}$$

where Θ and K are given by (3.4) and (2.12) respectively and $A^{(2q)}$ denotes the leading $q \times 2q$ submatrix of A whose elements are found from (2.13). The representation (4.2) seems to provide a useful alternative to the asymptotic formula given by Walker (1961).

REFERENCES

ANDERSON, T. W. (1975). Maximum likelihood estimation of parameters of autoregressive processes with moving average residuals and other covariance matrices with linear structure. *Ann. Statist.* **3** 1283–1304.

ANDERSON, T. W. and WALKER, A. M. (1964). On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. *Ann. Math. Statist.* **35** 1296–1303.

BARTLETT, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. *J. Roy. Statist. Soc.* B **8** 27–41.

DURBIN, J. (1959). Efficient estimation of parameters in moving average models. *Biometrika* **46** 306–316.

GODOLPHIN, E. J. (1976). On the Cramér-Wold factorization. *Biometrika* **63** 367–380.

GODOLPHIN, E. J. (1977). A direct representation for the maximum likelihood estimator of a Gaussian moving average process. *Biometrika* **64** 375–384.

GODOLPHIN, E. J. (1978). Modified maximum likelihood estimation of Gaussian moving averages using a pseudoquadratic convergence criterion. *Biometrika* **65** 203–206.

HANNAN, E. J. (1969). The estimation of mixed moving average autoregressive systems. *Biometrika* **56** 579–593.

SHAMAN, P. (1975). An approximate inverse for the covariance matrix of moving average and autoregressive processes. *Ann. Statist.* **3** 532–538.

WALKER, A. M. (1961). Large-sample estimation of parameters for moving average models. *Biometrika* **48** 343–357.

WHITTLE, P. (1953). Estimation and information in stationary time series. *Ark. Mat. Fys. Astr.* **2** 423–434.

WOLD, H. (1949). A large-sample test for moving averages. *J. Roy. Statist. Soc. B* **11** 297–305.

ZEHNA, P. W. (1966). Invariance of maximum likelihood estimation. *Ann. Math. Statist.* **37** 744.

ROYAL HOLLOWAY COLLEGE
UNIVERSITY OF LONDON
EGHAM HILL
EGHAM, SURREY TW20 OEX
ENGLAND