Statistics

FACULTY OF ACTUARIAL SCIENCE & ECONOMETRICS

A & E REPORT

REPORT AE 10/80

FORMULAE FOR THE COVARIANCE STRUCTURE OF THE SAMPLED AUTOCOVARIANCES FROM SERIES GENERATED BY GENERAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE PROCESSES OF ORDER (p,d,q) , $d = 0$ or 1

O.D. Anderson and J.G. de Gooijer

University of Amsterdam

FACULTY OF ACTUARIAL SCIENCE & ECONOMETRICS

Report AE 10/80

FORMULAE FOR THE COVARIANCE STRUCTURE OF THE SAMPLED AUTOCOVARIANCES FROM SERIES GENERATED BY GENERAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE PROCESSES OF ORDER (p,d,q) , $d = 0$ or 1.

by O.D. Anderson and J.G. de Gooijer

Some keywords: autocovariances, exact moments, stationary and non-stationary processes, variances

Classification: AMS(MOS) subject classification scheme (1989): 62M10

Free copies can be obtained from:
Faculty of Actuarial Science and Econometrics
University of Amsterdam
Jodenbreestraat 23 - 1011 NH Amsterdam
The Netherlands telephone number: 020-5254217/12

University of Amsterdam

Formulae for the Covariance Structure of the Sampled Autocovariances from Series Generated by General Autoregressive Integrated Moving Average Processes of Order (p,d,q), d = 0 or 1.

By O.D. ANDERSON

Statistical Applications Limited

and

J.G. DE GOOIJER

University of Amsterdam

SUMMARY. We provide exact formulae for the variances and covariances of the sampled variances and autocovariances, given a time series realisation from any stationary or once integrated stationary mixed autoregressive moving average process, and indicate how these can then yield approximate first and second order moments for the associated sampled serial correlation behaviour.

1. INTRODUCTION

The general autoregressive integrated moving average process of order (p,d,q), which we shall abbreviate to ARIMA(p,d,q), is defined by a stochastic sequence, $\{Z_i\}$, satisfying

$$(1-\phi_1 B-\ldots-\phi_p B^p)(1-B)^d Z_i = (1-\theta_1 B-\ldots-\theta_q B^q)A_i \tag{1}$$

where $(\phi_1,\ldots,\phi_p)$ and $(\theta_1,\ldots,\theta_q)$ are two sets of real parameters, with the first subject to the stationarity condition, namely that the polynomial $1-\phi_1\zeta-\ldots-\phi_p\zeta^p$ in the complex variable ζ has no zero within or on the unit circle; and B is the backshift operator, such that B^j operating on any X_i, for instance A_i or Z_i, produces X_{i-j}. $\{A_i\}$ is a white noise sequence of independent but identically distributed normal zero-mean random variables, all with variance σ^2 say. Note that it is unnecessary to impose any further restrictions on the real θ-parameters; and for convenience, when p or q are unity, we drop those suffices from, respectively, ϕ_1 and θ_1.

In this paper, we will be concerned with the sampled variance and autocovariances $c_{k,d}^{(n)}$ $(k = 0,1,\ldots,n-1)$ for any series realisation $\{z_i : i = 1,\ldots,n\}$, of length n, which has been generated from a model (1) process. These are then given by

$$c_{k,d}^{(n)} = n^{-1} \sum_{i=1}^{n-k} (z_i - \bar{z})(z_{i+k} - \bar{z}) \qquad (k = 0,1,\ldots,n-1) \qquad (2)$$

where $\bar{z} = n^{-1}(z_1 + \ldots + z_n)$ is the mean of the observed series. When no confusion is likely to arise, we suppress the suffix d in $c_{k,d}^{(n)}$.

Anderson (1980a, equations 22 and 39) has already provided formulae for $E\left[c_k^{(n)}\right]$, in the special cases of ARIMA(p,d,q) processes with d=0 or 1, which are restated as results (5) below. Here we intend to extend the theory (as suggested by Anderson 1980b) to obtain $\mathrm{Var}\left[c_k^{(n)}\right]$ and $\mathrm{Cov}\left[c_k^{(n)}, c_h^{(n)}\right]$ for these models.

The motivation for our study is primarily to demonstrate that, given an ARIMA(p,1,q) model or an ARMA(p,q) which approaches being such, in the sense that it has an autoregressive operator factor $(1-\phi B)$ with ϕ less than but (depending on n) not much less than unity, then the serial correlations, $\{r_k^{(n)} = c_k^{(n)}/c_o^{(n)}\}$, are generally closely characterised by $\{E_k^{(n)} = E\left[c_k^{(n)}\right]/E\left[c_o^{(n)}\right]\}$. To do this we would show that, for these models, $E_k^{(n)}$ provides a good approximation for $E\left[r_k^{(n)}\right]$ and, in addition, the variability of the $r_k^{(n)}$ is also low.

2. BACKGROUND THEORY

First define an infinite set of parameters ψ_j by

$$\psi_j \equiv 0 \qquad (j < 0)$$

$$\psi_0+\psi_1 B+\psi_2 B^2+\ldots \equiv (1-\phi_1 B-\ldots-\phi_p B^p)^{-1}(1-\theta_1 B-\ldots-\theta_q B^q)$$

and then an $n \times \infty$ matrix $\underset{\sim}{\Psi}$ as $(\Psi_{r,s} \equiv \psi_{s-r})$. Consider $n \times n$ matrices as follows: the identity matrix $\underset{\sim}{I}$; a matrix of unities $\underset{\sim}{U}$; $\underset{\sim}{V} = \underset{\sim}{I} - n^{-1}\underset{\sim}{U}$; $\underset{\sim}{\Lambda}_k$ a null matrix except for ones everywhere on the kth superdiagonal; and $\underset{\sim}{F}_d$, defined as $\underset{\sim}{I}$ when d=0 but obtained from $\underset{\sim}{U}$ by putting all lower triangular elements to zero when d=1. Finally, define

$$\underset{\sim}{M}_{k,d} = n^{-1}\sigma^2\{\underset{\sim}{\Psi}'\underset{\sim}{V}\underset{\sim}{F}_d'(\underset{\sim}{\Lambda}_k+\underset{\sim}{\Lambda}_k')\underset{\sim}{F}_d\underset{\sim}{V}\underset{\sim}{\Psi}\}/2 \qquad (d=0,1).$$

Then it is simple to show for model (1), with d = 0 or 1, that the r-th cumulant of $c_{k,d}^{(n)}$ exists and is given by $2^{r-1}(r-1)!\,\mathrm{tr}(\underset{\sim}{M}_{k,d}^r)$. For instance, see Lancaster (1954). So that, in particular,

$$E\left[c_{k,d}^{(n)}\right] = \mathrm{tr}(\underset{\sim}{M}_{k,d}) \qquad (3)$$

$$\mathrm{Var}\left[c_{k,d}^{(n)}\right] = 2\mathrm{tr}(\underset{\sim}{M}_{k,d}^2). \qquad (4)$$

Now Anderson (1980a) has evaluated (3) explicitly and obtained results which can in fact be rewritten as

$$E\left[c_{k,d}^{(n)}\right] = n^{-3}\{n^2(n-k)F_d(k)+2nG_{d+1}(k)+2kF_{d+1}(n)\} \qquad (d=0,1) \qquad (5)$$

where

$$F_0(k) = \gamma_k$$

$$F_1(k) = -\frac{k}{2}\gamma_o - \sum_{j=1}^{k-1}(k-j)\gamma_j$$

$$F_2(k) = \frac{k(k^2-1)}{12}\gamma_o + \frac{1}{6}\sum_{j=1}^{k-2}(k-j-1)(k-j)(k-j+1)\gamma_j$$

$$G_{d+1}(k) = F_{d+1}(n-k)-F_{d+1}(k)$$

and $\{\gamma_k\}$ are the theoretical autocovariances for the ARMA(p,q) process, obtained from the original model (1) be deleting $(1-B)^d$.

3. NEW RESULTS

Similarly, on evaluation, (4) yields a result analogous to (5), namely (6) below. For convenience, consider

$$F(k,\{\alpha_j\}) = \sum_{j=0}^{k-1} f_j(k)\alpha_j$$

such that $F(k,\{\gamma_j\}) \equiv -F_1(k)$. Then

$$\begin{aligned}\mathrm{Var}\left[c_{k,d}^{(n)}\right] = 2n^{-6}\Big[n^4 F(n-k,\{F_d^2(j)+F_d(|k-j|)F_d(k+j)\}) \\ -n^2\{kH_d(k,n,n)+nH_d(k,n-k,n)-nH_d(k,k,0)\} \\ +2n^2\{F_{d+1}(n)F_{d+1}(n-2k)+G_{d+1}^2(k)\} \\ +4kF_{d+1}(n)\{kF_{d+1}(n)+2nG_{d+1}(k)\}\Big] \qquad (d=0,1) \qquad (6)\end{aligned}$$

where

$$H_d(k,a,b) = P_d(k,a,k-n+a)+Q_d(k,b,a)$$

$$P_d(f,g,h) = \sum_{s=0}^{n-1-f}\{\sum_{u=-s}^{g-1-s} F_d(u) \sum_{v=-h-s}^{g-1-h-s} F_d(v)\}$$

$$Q_d(f,g,h) = \sum_{s=0}^{n-1-f}\{\sum_{u=1-g+s}^{h-g+s} F_d(u) \sum_{v=n-f-g-s}^{h-1+n-f-g-s} F_d(v)\}$$

and note that $P_d(k,n,k) \equiv Q_d(k,n,n)$.

For instance, when d=0, writing (6) out in full in terms of the theoretical autocovariances, we get

$$
\begin{aligned}
\mathrm{Var}\left[c_{k,o}^{(n)}\right] = n^{-6}\Big[& n^4\Big\{(n-k)(\gamma_o^2+\gamma_k^2)+2\sum_{j=1}^{n-k-1}(n-k-j)(\gamma_j^2+\gamma_{k-j}\gamma_{k+j})\Big\} \\
& -2n^2\sum_{s=0}^{n-k-1}\Big\{2k\sum_{i=-s}^{n-1-s}\gamma_i\sum_{j=-k-s}^{n-k-1-s}\gamma_j+n\Big(\Big\{\sum_{j=-s}^{n-k-1-s}\gamma_j\Big\}^2+\sum_{i=k-s}^{n-1-s}\gamma_i\sum_{j=-k-s}^{n-2k-1-s}\gamma_j-\sum_{i=-s}^{k-1-s}\gamma_i\sum_{j=n-2k-s}^{n-k-1-s}\gamma_j-\sum_{i=1+s}^{k+s}\gamma_i\sum_{j=n-k-s}^{n-1-s}\gamma_j\Big)\Big\} \\
& +n^2\Big\{\Big(n\gamma_o+2\sum_{j=1}^{n-1}(n-j)\gamma_j\Big)\Big((n-2k)\gamma_o+2\sum_{j=1}^{n-2k-1}(n-2k-j)\gamma_j\Big)+\Big((n-2k)\gamma_o+2\sum_{j=1}^{n-k-1}(n-k-j)\gamma_j-2\sum_{j=1}^{k-1}(k-j)\gamma_j\Big)^2\Big\} \\
& +2k\Big(n\gamma_o+2\sum_{j=1}^{n-1}(n-j)\gamma_j\Big)\Big\{k\Big(n\gamma_o+2\sum_{j=1}^{n-1}(n-j)\gamma_j\Big)+2n\Big((n-2k)\gamma_o+2\sum_{j=1}^{n-k-1}(n-k-j)\gamma_j-2\sum_{j=1}^{k-1}(k-j)\gamma_j\Big)\Big\}\Big]
\end{aligned}
\tag{7}
$$

and, in particular, when k=0 this gives

$$\mathrm{Var}\left[c_{o,o}^{(n)}\right] = 2n^{-4}\left[n^2\{n\gamma_o^2+2\sum_{1}^{n-1}(n-j)\gamma_j^2\}-2n\sum_{s=0}^{n-1}\left(\sum_{-s}^{n-1-s}\gamma_j\right)^2\right.$$

$$\left.+\{n\gamma_o+2\sum_{1}^{n-1}(n-j)\gamma_j\}^2\right] . \qquad (8)$$

So, for the case of white noise say, we get

$$\mathrm{Var}\left[c_{o,o}^{(n)}\right] = 2\sigma^4(n-1)/n^2 \qquad (9)$$

which is seen to be correct by direct calculation.

From (4), using the definition of $\underset{\sim}{M}_{k,d}$, it immediately follows that the sum of the coefficients of $\gamma_u\gamma_v$ in $\mathrm{Var}\left[c_{k,d}^{(n)}\right]$ should be zero, which can be used as a simple check on (7).

Finally, if we consider an ARMA(p+1,q), with autocovariances $\{\gamma_j^*(\alpha)\}$ say, formed by putting in an extra factor, $(1-\alpha B)$, on the autoregressive side of the original ARMA(p,q), and if we then express these $\gamma_j^*(\alpha)$'s in terms of α and the γ_j's of the ARMA(p,q), we find that in the limit as $\alpha\to1$, (7) then yields the same expression as (6), when written out in full for d=1. Of course, it is this observation which encourages up to write the results for both d=0 and d=1 in the single combined form (6).

Again, it is simple to show that (4) can be generalised to give

$$\mathrm{Cov}\left[c_{k,d}^{(n)},c_{h,d}^{(n)}\right] = 2\mathrm{tr}(\underset{\sim}{M}_{k;d}\underset{\sim}{M}_{h,d}) \qquad (10)$$

which could be straightforwardly evaluated, if required. However, we restrict our attention here to the case h=0, when one gets

$$\mathrm{Cov}\left[c_{k,d}^{(n)},c_{o,d}^{(n)}\right] = 2n^{-5}\left[n^3\{(n-k)\sum_{j=0}^{k-1}F_d(j)F_d(k-j)+2F(n-k),\{F_d(j)F_d(k+j)\})\}\right.$$

$$-n^2\{P_d(k,n,k)+P_d(o,n-k,-k)-P_d(o,k,k-n)\}-nkP_d(o,n,o)$$

$$+4F_{d+1}(n)\{kF_{d+1}(n)+nF_{d+1}(n-k)-nF_{d+1}(k)\}\Big] \quad (d=0,1). \quad (11)$$

For d=0, writing this out in full, we find

$$\mathrm{Cov}\Big[c_{k,o}^{(n)},c_{o,o}^{(n)}\Big] = 2n^{-5}\Big[n^3\{(n-k)(\sum_{j=0}^{k-1}\gamma_j\gamma_{k-j}+\gamma_o\gamma_k)+2\sum_{j=1}^{n-k-1}(n-k-j)\gamma_j\gamma_{k+j}\}$$

$$-n^2\sum_{s=0}^{n-k-1}(\sum_{i=-s}^{n-1-s}\gamma_i\sum_{j=-k-s}^{n-k-1-s}\gamma_j)-n\sum_{s=0}^{n-1}\{n(\sum_{i=-s}^{n-k-1-s}\gamma_i\sum_{j=k-s}^{n-1-s}\gamma_j-\sum_{i=-s}^{k-1-s}\gamma_i\sum_{j=n-k-s}^{n-1-s}\gamma_j)+k(\sum_{j=-s}^{n-1-s}\gamma_j)^2\}$$

$$+\{n\gamma_o+2\sum_{j=1}^{n-1}(n-j)\gamma_j\}\{n(n-k)\gamma_o+2k\sum_{j=1}^{n-1}(n-j)\gamma_j+2n\sum_{j=1}^{n-k-1}(n-k-j)\gamma_j-2n\sum_{j=1}^{k-1}(k-j)\gamma_j\}\Big] \quad (12)$$

and it is easy to see that, for k=0, this reduces to (8) - as indeed it should do. Again, a check on (12) is obtained by showing that the sum of the $\gamma_u\gamma_v$ coefficients there is zero.

4. CONCLUSION

We now have sufficient to evaluate the following approximations given in Anderson (1979, equations 10 and 13), namely

$$E\Big[r_k^{(n)}\Big] \simeq \frac{E\Big[c_k^{(n)}\Big]}{E\Big[c_o^{(n)}\Big]}\{1-\frac{E\Big[c_k^{(n)}c_o^{(n)}\Big]}{E\Big[c_k^{(n)}\Big]E\Big[c_o^{(n)}\Big]}+\frac{E\Big[c_o^{(n)2}\Big]}{E^2\Big[c_o^{(n)}\Big]}\} \quad (13)$$

$$\mathrm{Var}\Big[r_k^{(n)}\Big] \simeq \frac{E^2\Big[c_k^{(n)}\Big]}{E^2\Big[c_o^{(n)}\Big]}\{\frac{E\Big[c_k^{(n)2}\Big]}{E^2\Big[c_k^{(n)}\Big]}-\frac{2E\Big[c_k^{(n)}c_o^{(n)}\Big]}{E\Big[c_k^{(n)}\Big]E\Big[c_o^{(n)}\Big]}+\frac{E\Big[c_o^{(n)2}\Big]}{E^2\Big[c_o^{(n)}\Big]}$$

$$-(\frac{E\Big[c_k^{(n)}c_o^{(n)}\Big]}{E\Big[c_k^{(n)}\Big]E\Big[c_o^{(n)}\Big]}-\frac{E\Big[c_o^{(n)2}\Big]}{E^2\Big[c_o^{(n)}\Big]})^2\}. \quad (14)$$

In particular, we are able to access when $E_k^{(n)}$ is a sufficiently close approximation to $E\Big[r_k^{(n)}\Big]$ and when the sampling variability of observed $r_k^{(n)}$ will be negligible. As an example, for length-20 series realisations from the AR(1) model with ϕ-parameter = .8, we have asymptotic $\rho_1 = .8$ asymptotic $\{\text{variance}\}^{\frac{1}{2}} = .13 + O(.13)$

exact $E\left[r_1^{(20)}\right] = .5685$ exact $\{\mathrm{Var}\left[r_1^{(20)}\right]\}^{\frac{1}{2}} = .1831$

approx. $E\left[r_1^{(20)}\right] = .5575$ approx. $\{\mathrm{Var}\left[r_1^{(20)}\right]\}^{\frac{1}{2}} = .1519$

$E_1^{(20)} = .5255$

where the exact values are obtained from numerical integration, following a method due to Imhof (1961), and the asymptotic variance is that of Bartlett (1946), which can be shown to still hold to $O(n^{-1})$ for our situation - see below. (However, it is only fair to point out that, in practice, the above approximations are more conveniently obtained computationally from the matrix trace expressions. But, whichever way they are calculated, in other than simple cases it will be necessary to use McLeod's (1975) algorithm for obtaining the γ's corresponding to the particular ARMA(p,q) part of the model.)

Also, if we evaluated $\mathrm{Cov}\left[c_k^{(n)}, c_h^{(n)}\right]$ from (10) above, we could get the approximation to $\mathrm{Cov}\left[r_k^{(n)}, r_h^{(n)}\right]$ suggested by Anderson (1979, equation 12).

Finally, we note that asymptotically, as $n\to\infty$, (5), (7) and (12) give, respectively, for d=0

$$E\left[c_k^{(n)}\right] \to \gamma_k \tag{15}$$

$$\mathrm{Var}\left[nc_k^{(n)}\right] \to \sum_{-\infty}^{\infty}(\gamma_j^2+\gamma_{k-j}\gamma_{k+j}) \tag{16}$$

$$\mathrm{Cov}\left[nc_k^{(n)}, c_0^{(n)}\right] \to 2\left(\sum_{-\infty}^{\infty}\gamma_j\gamma_{k+j} + \sum_{0}^{k+1}\gamma_j\gamma_{k-j}\right) \tag{17}$$

whereas, for d=1, we find that

$$E\left[c_k^{(n)}/n\right] \to \frac{1}{6}\sum_{-\infty}^{\infty}\gamma_j \tag{18}$$

$$\mathrm{Var}\left[c_k^{(n)}/n^2\right] \to \frac{1}{45}\left(\sum_{-\infty}^{\infty}\gamma_j\right)^2 \tag{19}$$

$$\mathrm{Cov}\left[c_k^{(n)}/n, c_0^{(n)}/n\right] \to \frac{1}{45}\left(\sum_{-\infty}^{\infty}\gamma_j\right)^2. \tag{20}$$

Results (13) to (17) then yield, for d=0,

$$E\left[r_k^{(n)}\right] \to \rho_k$$

$$\mathrm{Var}\left[nr_k^{(n)}\right] \to \sum_{-\infty}^{\infty}(\rho_j^2+\rho_{k-j}\rho_{k+j}-4\rho_j\rho_k\rho_{k+j}+2\rho_j^2\rho_k^2)-4\sum_{j=0}^{k-1}\rho_j\rho_k\rho_{k-j} \tag{22}$$

whilst, for d=1, we get from (13), (14), (18), (19) and (20) that

$$E\left[r_k^{(n)}\right] \to 1 \tag{23}$$

$$\mathrm{Var}\left[r_k^{(n)}\right] \to 0. \tag{24}$$

(Limit (15) is of course well-known, (16) recovers a special case of Bartlett's relation 5 (1946), but (17) modifies that result in general by an additional $2\sum_{0}^{k-1}\gamma_j\gamma_{k-j}$. Limits (18) and (19) agree with Roy and Lefrançois (1978), whilst (22) retrieves Bartlett's formula 7 (1946).- apart from the final subtracted term - as remarked on above.)

REFERENCES

ANDERSON, O.D. (1979): Some sample autocovariance function results for a once integrated q-th order moving average process. *Statistica*, **39**, 287-299.

ANDERSON, O.D. (1980a): Formulae for the expected values of the sampled variance and covariances from series generated by general autoregressive integrated moving average processes of order (p,d,q). *Sankhyā*, to appear.

ANDERSON, O.D. (1980b): Serial Dependence Properties of Linear Processes. *J. Opl Res. Soc.*, to appear.

BARTLETT, M.S. (1946): On the theoretical specification and sampling properties of autocorrelated time series. Supplement *J. Roy. Statist. Soc. B*, **8**, 27-41. Correction (1948) **10**, 1 .

IMHOF, J.P. (1961): Computing the distribution of quadratic forms in normal variables. *Biometrika*, **48**, 419-426.

LANCASTER, H.O. (1954): Traces and cumulants of quadratic forms in normal variables. *J. Roy. Statist. Soc. B*, **16**, 247-254.

McLEOD, A.I. (1975): Derivation of the theoretical autocovariance function of autoregressive-moving average time series. *Applied Statistics*, **24**, 255-6. Correction (1977) **26**, 194.

ROY, R. & LEFRANCOIS, P. (1978): On the sample autocovariance and autocorrelation functions for an autoregressive integrated moving average process. Paper presented at IMS Special Topics Meeting on Time Series Analysis, Iowa, 1-3 May 1978.