On Fisher's information matrix of an ARMA process and Sylvester's resultant matrix

Klein, A.; Spreij, P.

1993

Faculty of Economics and Econometrics

On Fisher's information matrix of an ARMA process and Sylvester's resultant matrix

André Klein
Peter Spreij

Research Memorandum 1993-33

July 1993

vrije Universiteit *amsterdam*

On Fisher's information matrix of an ARMA process and Sylvester's resultant matrix

André Klein
Department of Economic Statistics
University of Amsterdam
Roetersstraat 11
1018 WB Amsterdam

Peter Spreij
Department of Econometrics
Free University
De Boelelaan 1105
1081 HV Amsterdam

Abstract

We consider two relations between Fisher's information matrix of a stationary ARMA (autoregressive moving average) process and Sylvester's resultant matrix. One is based on the Wald test statistic for testing common roots of the AR and MA polynomials of an ARMA process, and the other one is established by using the structure of Fisher's information matrix. It turns out that the latter is also a resultant.

1 Introduction

The Cramér-Rao bound is of paramount importance for evaluating the performance of (stationary) autoregressive moving average (ARMA) models, where the focus is on the error covariance matrix of the estimated parameters. See Cramér [3] and Rao [7]. For computing the Cramér-Rao bound the inverse of Fisher's information matrix is needed. The latter is singular in the presence of common roots of the AR and the MA polynomial.

The purpose of the present paper is to study the link between Fisher's information matrix and Sylvester's resultant matrix. From this investigation we conclude that Fisher's information matrix of a stationary ARMA model is also a resultant.

The link between statistical considerations and algebraic results is of independent interest since one is based on the Wald test statistic for testing common roots (see Klein [6]), and the other one is deduced from the structure of the Fisher information matrix of a stationary ARMA process.

In Barnett [1] a relationship between Sylvester's resultant matrix and the companion matrix of a polynomial is given. Kalman [4] has investigated the concept of observability and controllability in function of Sylvester's resultant matrix. Similar results can be found in Barnett [2] which contains discussions on these topics and a number of further references. Furthermore in Söderström & Stoica [8] (page 162 ff.) a discussion on overparametrization in terms of the transfer function of a system can be found.

We consider the Wald test for testing common roots of two polynomials where the unknown vector valued parameter θ belongs to $\Theta \subset \mathbb{R}^s$ and satisfies the restrictions

$$g_1(\theta) = \ldots = g_r(\theta) = 0.$$

$\hat{\theta}_n$ denotes the unrestricted maximum likelihood estimator of θ based on n observations. For a large number of observations n we assume $\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow N(0, F(\theta)^{-1})$ in distribution, where $F(\theta)$ is the Fisher information matrix and $(g_1(\hat{\theta}_n), \ldots, g_r(\hat{\theta}_n)) = g(\hat{\theta}_n)$ will be considered at the proximity to the zero vector.

The Wald test is based on the distance between $g(\hat{\theta}_n)$ and $g(\theta)$. If the true parameters satisfy the restrictions, then $\sqrt{n}g(\hat{\theta}_n) = G^T(\theta)\sqrt{n}(\hat{\theta}_n - \theta) + o_P(1)$, where $G(\theta)$ is the $s \times r$ matrix with ij-th element given by $\partial g_i(\theta)/\partial\theta_j$.

Therefore $\sqrt{n}g(\hat{\theta}_n)$ is under the null hypothesis that the restrictions are satisfied, asymptotically normal with zero mean and covariance matrix equal to

$$V(\theta) = G^T(\theta)F^{-1}(\theta)G(\theta),$$

provided of course that this matrix is nonsingular. A Wald test statistic is then

$$ng^T(\hat{\theta}_n)V(\theta)^{-1}g(\hat{\theta}_n). \tag{1.1}$$

If $g(\theta) = 0$, the test statistic from equation (1.1) is distributed approximately as $\chi^2(r)$. See Klein [6].

If the restriction $g(\theta)$ is chosen to express the difference between the roots of two polynomials involved in the model, then $g(\theta) = 0$ in the case of equal roots. Under this condition $F(\theta)$ becomes singular and hence the Wald test statistic is not well defined anymore. In Klein [6] a solution is proposed for the Wald test, when this situation occurs.

2 Preliminary algebraic results

Consider the following two scalar polynomials in the variable z.

$$\begin{aligned} A(z) &= z^p + a_1 z^{p-1} + \ldots + a_p && (2.1) \\ C(z) &= z^q + c_1 z^{q-1} + \ldots + c_q && (2.2) \end{aligned}$$

The Sylvester resultant matrix of A and C is defined as the $(p+q) \times (p+q)$ matrix

$$S(a,c) = \begin{array}{l} q\left\{ \begin{array}{l} \\ \\ \\ \end{array} \right. \\ \\ p\left\{ \begin{array}{l} \\ \\ \\ \end{array} \right. \end{array} \left[\begin{array}{cccccccc} 1 & a_1 & \cdots & \cdots & a_p & & & 0 \\ & \ddots & \ddots & & & & \ddots & \\ 0 & & & 1 & a_1 & \cdots & \cdots & a_p \\ \hdashline 1 & c_1 & \cdots & \cdots & c_q & & & 0 \\ & \ddots & \ddots & & & & \ddots & \\ 0 & & & 1 & c_1 & \cdots & \cdots & c_q \end{array} \right] \tag{2.3}$$

In the presence of common roots of A and C the matrix $S(a,c)$ becomes singular. Moreover it is known that

$$\det S(a,c) = \prod_{i=1}^{p} \prod_{j=1}^{q} (\gamma_j - \alpha_i) \tag{2.4}$$

where the α_i and the γ_j are the roots of A and C respectively.
If one takes in the previous section the coefficients of A and C as the parameter θ, then an appropriate choice for the restriction function $g(\theta)$ in the Wald test is

$$g(\theta) = \prod_{i=1}^{p} \prod_{j=1}^{q} (\gamma_j - \alpha_i) \tag{2.5}$$

It follows then that the statistic of equation (1.1), under the restriction that (2.5) is zero, has asymptotically a $\chi^2(1)$ distribution.
A first relationship involving $F(\theta)$ and $S(\theta)$ with $\theta = (a,c)$, when equation (2.5) is substituted in equation (1.1) yields

$$n\{[\frac{\partial}{\partial\theta} \log\det S(\hat{\theta}_n)]^T F^{-1}(\theta) [\frac{\partial}{\partial\theta} \log\det S(\hat{\theta}_n)]\}^{-1} \rightarrow \chi^2(1) \tag{2.6}$$

in distribution as n tends to infinity.

3 Main result

In this section we show that $F(\theta)$ is a resultant.
First we specify Fisher's information matrix of an ARMA(p,q) process. Let A and C be the same monic polynomials as in the previous section. Define $A^*(z) = z^p A(z^{-1})$ and $C^*(z) = z^q C(z^{-1})$. Consider then the stationary ARMA process y that satisfies

$$A^*(L)y = C^*(L)\varepsilon \tag{3.1}$$

with L the lag operator and ε a white noise sequence. Assume also that C has no zeros on the unit circle. As can be found in Klein & Mélard [5] the Fisher information matrix $F(\theta)$ then has the following block decomposition

$$F(\theta) = \begin{bmatrix} F_{aa} & F_{ac} \\ F_{ac}^T & F_{cc} \end{bmatrix} \tag{3.2}$$

where the matrices appearing here have the following elements

$$F_{aa}^{jk} = \frac{1}{2\pi i} \oint_{|z|=1} \frac{z^{j-k+p-1}}{A(z)A^*(z)} dz, (j, k = 1, \ldots, p) \tag{3.3}$$

$$F_{ac}^{jk} = \frac{-1}{2\pi i} \oint_{|z|=1} \frac{z^{j-k+q-1}}{C(z)A^*(z)} dz, (j = 1, \ldots, p, k = 1, \ldots, q) \tag{3.4}$$

$$F_{cc}^{jk} = \frac{1}{2\pi i} \oint_{|z|=1} \frac{z^{j-k+q-1}}{C(z)C^*(z)} dz, (j, k = 1, \ldots, q) \tag{3.5}$$

The main theorem is now the following

Theorem 3.1 *The Fisher information matrix of an ARMA(p,q) process with polynomials $A^*(z)$ and $C^*(z)$ of order p, q repectively becomes singular iff these polynomials have at least one common root.*

The proof of this theorem is an immediate consequence of the following lemma together with equation (2.4). First we have to introduce some auxilary notation. Write for each positive integer k $u_k(z) = [1, z, \ldots, z^{k-1}]^T$, $u_k^*(z) = [z^{k-1}, \ldots, 1]^T$ $= z^{k-1}u_k(z^{-1})$and let $K(z) = A(z)A^*(z)C(z)C^*(z)$. Define moreover

$$P(\theta) = \frac{1}{2\pi i} \oint_{|z|=1} \frac{u_{p+q}(z)u_{p+q}^*(z)^T}{K(z)} dz \tag{3.6}$$

Lemma 3.2 *The following factorization holds.*

$$F(\theta) = S(-c, a)P(\theta)S(-c, a)^T \tag{3.7}$$

PROOF: A simple computation shows that we can write $F(\theta)$ in matrix form as

$$F(\theta) = \frac{1}{2\pi i} \oint_{|z|=1} \frac{1}{K(z)} \begin{bmatrix} -C^*(z)u_p(z) \\ A^*(z)u_q(z) \end{bmatrix} \begin{bmatrix} -C(z)u_p^*(z)^T & A(z)u_q^*(z)^T \end{bmatrix} dz \tag{3.8}$$

It also straightforward to verify that the following identities hold.

$$S(-c,a)u_{p+q}(z) = \begin{bmatrix} -C^*(z)u_p(z) \\ A^*(z)u_q(z) \end{bmatrix} \tag{3.9}$$

$$S(-c,a)u^*_{p+q}(z) = \begin{bmatrix} -C(z)u^*_p(z) \\ A(z)u^*_q(z) \end{bmatrix} \tag{3.10}$$

Hence equation (3.7) follows now immediately from equations (3.6), (3.8), (3.9) and (3.10). □

PROOF of theorem 3.1: Clearly the matrix $F(\theta)$ becomes singular if A and C have at least one common root in view of equation (2.4). In order to prove the converse, we only have to prove that $P(\theta)$ is strictly positive definite. This is also straight forward:
Rewrite $P(\theta)$ as

$$P(\theta) = \frac{1}{2\pi i}\oint_{|z|=1} \frac{u_{p+q}(z)u_{p+q}(z^{-1})^T}{A(z)A(z^{-1})C(z)C(z^{-1})}z^{-1}dz$$

Take now $z = e^{i\phi}$, then we get

$$P(\theta) = \frac{1}{2\pi}\int_0^{2\pi} \frac{u_{p+q}(e^{i\phi})u_{p+q}(e^{-i\phi})^T}{A(e^{i\phi})A(e^{-i\phi})C(e^{i\phi})C(e^{-i\phi})}d\phi$$

which in turn can be rewritten as

$$P(\theta) = \frac{1}{2\pi}\int_0^{2\pi} \frac{u_{p+q}(e^{i\phi})}{A(e^{i\phi})C(e^{i\phi})}\overline{\frac{u_{p+q}(e^{i\phi})}{A(e^{i\phi})C(e^{i\phi})}}^T d\phi$$

Let now $x \in \mathbb{R}^{p+q}$ such that $x^T P(\theta) = 0$. Then it follows that $x^T \frac{u_{p+q}(e^{i\phi})}{A(e^{i\phi})C(e^{i\phi})} = 0$ for almost all ϕ. But this is clearly only possible if $x = 0$. So $P(\theta) > 0$. □

The matrix $P(\theta)$ can be calculated by means of Cauchy's integral formula in the presence of common roots as follows. Let δ be a common root of A and C that appears as a zero of AC of order $l \geq 2$. Then

$$P(\theta) = \frac{1}{2\pi i}\oint_{|z|=1} \frac{f(z)}{(z-\delta)^l}dz$$

with

$$f(z) = \frac{u_{p+q}(z)u_{p+q}(z^{-1})^T}{A(z)A(z^{-1})C(z)C(z^{-1})}\frac{(z-\delta)^l}{z},$$

which is analytic in a disk of radius ρ around δ for sufficiently small ρ. Cauchy's theorem states that $P(\theta)$ is the sum of residuals, of which in particular the residual in δ can be computed as

$$f^{(l-1)}(\delta) = \frac{(l-1)!}{2\pi i}\oint_{|z-\delta|=\rho} \frac{f(z)}{(z-\delta)^l}dz. \tag{3.11}$$

It then follows that the more common roots A and C have, the less residuals are needed for the computation of $P(\theta)$.

As a corollary to lemma 3.2 we mention the following. Consider an AR process of order m, with AR polynomial $\tilde{A}^*(z)$ of order m. According to equation (3.7) and the fact that the Sylvester matrix is now the m-dimensional unit matrix, the Fisher information matrix $\tilde{F}$ becomes in this case

$$\tilde{F} = \frac{1}{2\pi i} \oint_{|z|=1} \frac{u_m(z) u_m^*(z)^T}{\tilde{A}(z)\tilde{A}^*(z)} dz \tag{3.12}$$

Take now in particular $\tilde{A}^*(z) = A^*(z)C^*(z)$ (so $m = p+q$), then it follows again from equation (3.12) and the fact that now $\tilde{A}(z)\tilde{A}^*(z) = K(z)$, that one has $P(\theta) = \tilde{F}$ and hence equation (3.7) reads

$$F(\theta) = S(-c, a)\tilde{F}S(-c, a)^T \tag{3.13}$$

So equation (3.13) gives a relationship between the Fisher information matrix of an ARMA(p,q) process and that of an appropriate AR(p+q) process.

4 Conclusions

It has been shown that the Fisher information matrix of an ARMA process, which is essential in a fully statistical framework, embraces Sylvester's resultant matrix and that both matrices are resultants. By establishing an explicit expression containing both matrices a simplified way can be deduced for evaluating the Fisher information matrix of a high order (p+q say) AR process by computing the Fisher information matrix of a suitable ARMA(p,q) process.

References

[1] Barnett, S. (1971). *A new formulation of the the theorems of Hurwitz, Routh and Sturm*, J. Inst. Maths. Applics. 8, pp. 240 - 250.

[2] Barnett, S. (1973). *Matrices, polynomials and linear time invariant systems*, IEEE Trans. Automat. Control. AC-18, pp. 1 - 10.

[3] Cramér, H. (1951), Mathematical methods of statistics, Princeton NJ; Princeton University Press.

[4] Kalman, R.E. (1963). *Mathematical description of linear dynamical systems*, SIAM J. Contr. 1, pp. 152 - 192.

[5] Klein, A. & Mélard, G. (1990). *Fisher's information matrix for seasonal ARMA models*, J. Time Series, Vol. 11, pp. 231 - 237.

[6] Klein, A. (1993). *A Wald-test for testing common roots*, sumitted for publication.

[7] Rao, C.R. (1965). Linear statistical inference and it applications, New York; Wiley

[8] Söderström, T. & Stoica, P. (1989). System identification, Prentice hall.

1992-1	R.J. Boucherie N.M. van Dijk	Local Balance in Queueing Networks with Positive and Negative Customers
1992-2	R. van Zijp H. Visser	Mathematical Formalization and the Analysis of Cantillon Effects
1992-3	H.L.M. Kox	Towards International Instruments for Sustainable Development
1992-4	M. Boogaard R.J. Veldwijk	Automatic Relational Database Restructuring
1992-5	J.M. de Graaff R.J. Veldwijk M. Boogaard	Why Views Do Not Provide Logical Data Independence
1992-6	R.J. Veldwijk M. Boogaard E.R.K. Spoor	Assessing the Software Crisis: Why Information Systems are Beyond Control
1992-7	R.L.M. Peeters	Identification on a Manifold of Systems
1992-8	M. Miyazawa H.C. Tijms	Comparison of Two Approximations for the Loss Probability in Finite-Buffer Queues
1992-9	H. Houba	Non-Cooperative Bargaining in Infinitely Repeated Games with Binding Contracts
1992-10	J.C. van Ours G. Ridder	Job Competition by Educational Level
1992-11	L. Broersma P.H. Franses	A model for quarterly unemployment in Canada
1992-12	A.A.M. Boons F.A. Roozen	Symptoms of Dysfunctional Cost Information Systems
1992-13	S.J. Fischer	A Control Perspective on Information Technology
1992-14	J.A. Vijlbrief	Equity and Efficiency in Unemployment Insurance
1992-15	C.P.M. Wilderom J.B. Miner A. Pastor	Organizational Typology: Superficial Foursome of Organization Science?
1992-16	J.C. van Ours G. Ridder	Vacancy Durations: Search or Selection?
1992-17	K. Dzhaparidze P. Spreij	Spectral Characterization of the Optional Quadratic Variation Process
1992-18	J.A. Vijlbrief	Unemployment Insurance in the Netherlands, Sweden, The United Kingdom and Germany