T E C H N I C A L

R E P O R T

0643

THE ASYMPTOTIC AND EXACT FISHER INFORMATION MATRICES OF A VECTOR ARMA PROCESS

KLEIN A., MELARD G. and A. SAIDI

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

The asymptotic and exact Fisher information matrices of a vector ARMA process

André Klein[a], Guy Mélard[b] and Abdessamad Saidi[c]
[a]Department of Quantitative Economics,
University of Amsterdam, Roetersstraat 11,
1018 WB Amsterdam, The Netherlands
(phone +31-20-5254245, fax: +31-205254349,
e-mail: A.A.B.Klein@uva.nl).
[b] ECARES CP 114, Université Libre de Bruxelles
Avenue Franklin Roosevelt 50, B-1050 Bruxelles, Belgium (gmelard@ulb.ac.be).
[c] Département de mathématiques et de statistique,
Université de Montréal, CP 6128, Succursale A,
Montréal, Québec, H3C 3J7, Canada (saidiabd@dms.umontreal.ca).

Abstract

The exact Fisher information matrix of a Gaussian vector autoregressive-moving average (VARMA) process has been considered for a time series of length N in relation with the exact maximum likelihood estimation method. In this paper it is shown that the Gaussian exact Fisher information matrix converges to the asymptotic Fisher information matrix when N goes to infinity.

AMS classification: 62M10, 62F12, 62B10

Keywords: Exact Fisher information matrix, Asymptotic Fisher information matrix, VARMA process.

[1]

[1] G. Mélard has benefited from an IAP-network in Statistics, contract P5/24, financially supported by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (OSTC).

1 Introduction

The exact Fisher information matrix of a Gaussian vector autoregressive-moving average (VARMA) process has been considered for a time series of length N in relation with the exact maximum likelihood estimation method. In this paper it is shown that the Gaussian exact Fisher information matrix converges to the asymptotic Fisher information matrix when N goes to infinity.

Several recent papers have discussed either the asymptotic Fisher information matrix (e.g. Godolphin and Bane, 2005) or the exact Fisher information matrix (e.g. Terceiro, 2000) but we have seen no indication of the result mentioned in the previous paragraph. Only Zadrozny (1989, 1992) mentions the two information matrices, exact and asymptotic, but we could not see a convergence between the two expressions. On the contrary, the asymptotic Fisher information is defined as the limit of the exact Fisher information.

Consider $\{y_t, t \in \mathbb{Z}\}$, $\mathbb{Z}$ the set of integers, a Gaussian vector autoregressive-moving average (VARMA) process of order (p, q) in dimension n, which satisfies the vector difference equation

$$\sum_{j=0}^{p} \alpha_j \; y_{t-j} = \sum_{j=0}^{q} \beta_j \; \varepsilon_{t-j}, \; t \in \mathbb{Z} \tag{1}$$

where $\{\varepsilon_t, t \in \mathbb{Z}\}$ is the innovation process, a sequence of independent zero mean n-dimensional random variables each having positive definite covariance matrix Σ, and where α_j, $\beta_j \in \mathbb{R}^{n \times n}$ are the parameter matrices, and $\alpha_0 \equiv \beta_0 \equiv I_n$.

We use L to denote the backward shift operator on $\mathbb{Z}$, equation, for example $L\, y_t = y_{t-1}$, then (1) can be written as

$$\alpha(L) \; y_t = \beta(L) \; \varepsilon_t \tag{2}$$

where

$$\alpha(z) = \sum_{j=0}^{p} \alpha_j \; z^j, \qquad \beta(z) = \sum_{j=0}^{q} \beta_j \; z^j$$

are the associated matrix polynomial. We further assume the eigenvalues of the matrix polynomials $\alpha(z)$ and $\beta(z)$ to be outside the unit circle so the elements of $\alpha^{-1}(z)$ and $\beta^{-1}(z)$ can be written as power series in z. These eigenvalues are obtained by solving the scalar polynomials $\det(\alpha(z)) = 0$ and $\det(\beta(z)) = 0$, where $\det(X)$ is the determinant of X. We assume that the matrix polynomials $\alpha(z)$ and $\beta(z)$ have no common eigenvalues so that non-singularity of Fisher's information matrix is guaranteed (e. g. Klein *et al.*, 2005).

Let $\{y_t, t = 1, ..., N\}$ be a time series generated by the VARMA process (2) and let the set of parameters $\vartheta = (\vartheta_1, \cdots, \vartheta_\ell)^\top$, where $\top$ denotes transposition and $\ell = n^2(p+q)$. The following definition of the parameter vector ϑ is introduced: $\vartheta = \text{vec}\left\{\alpha_1, \alpha_2, ..., \alpha_p, \beta_1, \beta_2, ..., \beta_q\right\}$, where vec X as usual stands for the vector resulting from stacking the columns of a matrix X on top of each other.

2 State space form and Fisher information matrices

Although it is not strictly needed, the exact Fisher information of a Gaussian process is often introduced using a state space representation (e.g. Hannan and Deitsler, 1988) using a vector of the state variables $x_t \in \mathbb{R}^m$, $t \in \mathbb{N}$. Among other possibilities, using a specific basis in the state space, the following state space structure is considered

$$x_{t+1} = \phi x_t + F \varepsilon_t \tag{3}$$

$$y_t = H x_t + \varepsilon_t, \tag{4}$$

where $\varepsilon_t \in \mathbb{R}^n$ is a Gaussian white noise process with $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}\left(\varepsilon_t \varepsilon_t^\top\right) = \Sigma > 0$, and

$$\phi = \begin{pmatrix} \alpha_1 & I_n & & 0_n \\ \alpha_2 & 0_n & \ddots & \\ \vdots & & \ddots & I_n \\ \alpha_h & 0_n & \cdots & 0_n \end{pmatrix}, \; F = \begin{pmatrix} \alpha_1 - \beta_1 \\ \alpha_2 - \beta_2 \\ \vdots \\ \alpha_h - \beta_h \end{pmatrix}, \quad \text{and } H^\top = \begin{pmatrix} I_n \\ 0_n \\ \vdots \\ 0_n \end{pmatrix}, \tag{5}$$

and $h = \max(p, q)$, $\alpha_i = 0_n$, $i > p$, $\beta_i = 0_n$, $i > q$, and consequently $m = hn$.

In Klein and Neudecker (2000) an appropriate representation at the vector-matrix level for the exact Fisher information matrix $\mathcal{J}(\vartheta)$ is set forth. It is based on the multivariate version of minus the logarithm of the likelihood of the system described by (3) and (4) and is given by

$$l(\vartheta) = -\log L(\vartheta) = \sum_{t=1}^{N} \left\{ \frac{n}{2} \log 2\pi + \frac{1}{2} \log \det B_t + \frac{1}{2} \widetilde{y}_t^\top B_t^{-1} \widetilde{y}_t \right\}.$$

where $\widetilde{y}$ and B_t are defined below. The exact information matrix is then

$$\mathcal{J}_N(\vartheta) = \mathbb{E} \frac{\partial^2 l(\vartheta)}{\partial \vartheta \partial \vartheta^\top},$$

to obtain

$$\mathcal{J}_N(\vartheta) = \sum_{t=1}^{N} \left[\frac{1}{2} \left(\frac{\partial \mathrm{vec} B_t}{\partial \vartheta^\top} \right)^\top (B_t \quad B_t)^{-1} \left(\frac{\partial \mathrm{vec} B_t}{\partial \vartheta^\top} \right) + \mathbb{E} \left\{ \left(\frac{\partial \widetilde{y}_t}{\partial \vartheta^\top} \right)^\top B_t^{-1} \left(\frac{\partial \widetilde{y}_t}{\partial \vartheta^\top} \right) \right\} \right]. \tag{6}$$

The operator   represents the Kronecker product of two matrices. The sample innovation $\widetilde{y}_t$ and its covariance matrix $B_t = \mathbb{E}[\widetilde{y}_t \widetilde{y}_t^\top]$ are obtained through the Kalman filter equations, see e.g. [1]

$$\begin{aligned} \widehat{y}_{t|t-1} &= H \hat{x}_{t|t-1} \\ \widetilde{y}_t &= y_t - \widehat{y}_{t|t-1} \\ \hat{x}_{t+1|t} &= (\phi - K_t H)\, \hat{x}_{t|t-1} + K_t y_t \\ K_t &= \left(\phi P_t H^\top\right) \left(H P_t H^\top\right)^{-1} \\ P_{t+1} &= \phi P_t \phi^\top + Q - \left(\phi P_t H^\top\right) \left(H P_t H^\top\right)^{-1} \left(H P_t \phi^\top\right). \end{aligned}$$

Note that Mélard and Klein (1994), Zadrozny and Mittnik (1994), Terceiro (1990, 2000) have introduced the exact Fisher information matrix for VARMA processes or more general dynamic systems. They have displayed it at a scalar level, and sometimes used the more efficient alternative Chandrasekhar recurrences instead of Kalman filter recurrences.

The asymptotic Fisher information matrix is more standard in statistics, econometrics and statistical signal processing, e.g. Whittle (1953), Friedlander (1984). For very general statistical models, under some regularity conditions, it is proved that the asymptotic covariance matrix of a maximum likelihood estimator, or any asymptotically equivalent estimator, is the inverse of the asymptotic Fisher information matrix. For the model defined by (1), it is given by

$$\mathcal{F}(\vartheta) = \mathbb{E} \left\{ \left(\frac{\partial \varepsilon}{\partial \vartheta} \right)^\top \Sigma^{-1} \left(\frac{\partial \varepsilon}{\partial \vartheta} \right) \right\}, \tag{7}$$

where we have omitted subscript t because the expectation does not depend on t. In Klein (2000), equivalence of (7) with a vector-matrix level version of Whittle (1953) is proven.

3 Main result

Theorem 1 *For a VARMA model under the conditions given in Section 1, given parametrization (5) inserted in (3) and (4) and using the expressions (6) and (7), we have*

$$\lim_{N\to+\infty} N^{-1}\mathcal{J}_N(\vartheta) = \mathcal{F}(\vartheta).$$

Proof. In addition to the ARMA and state-space forms, there is, at least in the stationary case, a direct representation of y_t in terms of an infinite sequence of lagged ε_t. Indeed, (3) can be written as

$$(I_n - \phi L)x_t = F\varepsilon_{t-1}. \tag{8}$$

Clearly, (8) is a vector difference equation. Imposing the stability condition $\lambda_{\max}(\phi) < 1$, where $\lambda_{\max}(\phi)$ denotes the eigenvalues of the matrix ϕ of maximum modulus, one solution is given by

$$x_t = \sum_{j=1}^{\infty} \phi^{j-1} F\varepsilon_{t-j}. \tag{9}$$

Hence the corresponding y_t solution is given by:

$$y_t = \sum_{j=1}^{\infty} H\phi^{j-1} F\varepsilon_{t-j} + \varepsilon_t. \tag{10}$$

Under the stability assumption every solution of $(I - \phi L)x_t = 0$ satisfies $x_t \longrightarrow 0$ for $t \longrightarrow +\infty$. Therefore, every solution of (8) converges to (9) for $t \longrightarrow +\infty$ which implies, of course that every solution y_t of (3) and (4) converges to (10) for $t \longrightarrow +\infty$. The solution (10) is called the steady state solution. Since, y_t converges to the steady state solution given by (10) as $t \longrightarrow +\infty$, it follows that when $t \longrightarrow +\infty$ the sample innovation $\widetilde{y}_t$ converges to ε_t which implies also that B_t converges to Σ. Indeed from (10) we can easily check that $\widetilde{y}_t = y_t - \widehat{y}_{t|t-1} = \varepsilon_t$. Now, as $t \longrightarrow +\infty$, the first sum in (6) multiplied by N^{-1} converges to zero since B_t will be independent of ϑ as $t \longrightarrow +\infty$. The second term of (6) multiplied by N^{-1} converges to the right hand side of (7) which effectively does not depend on t. Using Toeplitz Lemma (e.g. Loève, 1977, p. 250) with weights N^{-1}, that shows that the exact Fisher information matrix converges to the asymptotic Fisher information matrix when $N \longrightarrow +\infty$. ■

References

[1] ANDERSON, B.D.O., MOORE, J.B.,1979. Optimal Filtering. Prentice-Hall, Englewood Cliffs, N.J..

[2] FRIEDLANDER, B., 1984. On the computation of the Cramér-Rao bound for ARMA parameter estimation, IEEE Trans. Acoust. Speech, Signal Processing 32, 721-727.

[3] GODOLPHIN, E.J., BANE, S.R., 2005. On the evaluation of the information matrix for multiplicative seasonal time series models, J. Time Series Analysis 27, 167-190.

[4] HANNAN, E.J. , DEISTLER, M., 1988. The Statistical Theory of Linear Systems. John Wiley and Sons, New York.

[5] KLEIN, A., 2000. A generalization of Whittle's formula for the information matrix of vector mixed time series, Linear Algebra and its Applications 321, 197-208.

[6] KLEIN, A., MÉLARD, G., 1994. Computation of the Fisher information matrix for SISO models. IEEE Transactions on Signal Proc.42, 684-688.

[7] KLEIN, A., MÉLARD, G., P. SPREIJ, 2005. On the resultant property of the Fisher information matrix of a vector ARMA process, Linear Algebra Appl. 403, 291-313.

[8] Klein, A., Neudecker, H. , 2000. A direct derivation of the exact Fisher information matrix of Gaussian vector state space models. Linear Algebra Appl. 321, 233-238.

[9] Loève, M. , 1977. Probability Theory I, 4th edition, Springer, New York.

[10] Mélard, G., Klein, A., 1994. On a fast algorithm for the exact information matrix of a Gaussian ARMA time series, IEEE Trans. Signal Processing 42, 2201-2203.

[11] Terceiro Lomba, J., 1990. Estimation of Dynamic Econometric Models with Errors in Variables, Springer-Verlag, Berlin.

[12] Terceiro, J. , 2000. Comments on "Kalman-filtering methods for computing information matrices for time-invariant, periodic, and generally time-varying VARMA models and samples", Computers and Mathematics with Applications 40, 405-411.

[13] Whittle, P., 1953. The analysis of multiple stationary time series, J.Royal Statist. Soc. Ser. B. 15, 125-139.

[14] Zadrozny, P. A. , 1989. Analytic Derivatives for Estimation of Linear Dynamic Models, Computers and Mathematics with Applications 18, 539-553.

[15] Zadrozny, P. A., 1992. Errata to "Analytical derivatives for estimation of linear dynamic models", Computers and Mathematics with Applications 24, 289-290.

[16] Zadrozny, P. A., Mittnik, S. 1994. Kalman filtering methods for computing information matrices for time-invariant periodic and generally time-varying VARMA models and samples, Computers and Mathematics with Applications 28, 107-119.