

# Exact Maximum Likelihood Estimation of Structured or Unit Root Multivariate Time Series Models\*

Guy Mélard<sup>†</sup>    Roch Roy<sup>‡</sup>    Abdessamad Saidi<sup>§</sup>

**CRM-3129**

June 2004

---

\*The authors thank Saïd Nsiri for helpful comments. This work has been partly funded by the co-operation between the Commissariat général aux Relations internationales, Communauté Française Wallonie-Bruxelles and the Government of the Province of Québec, with a grant from the Action de Recherche Concertée 96/01/201, Communauté Française Wallonie-Bruxelles, and IAP-network of the Belgian federal Government contract P51/24. It was also supported by grants from the Natural Science and Engineering Research Council of Canada, the Network of Centres of Excellence on The Mathematics of Information Technology and Complex Systems (MITACS) and the Fonds québécois de la recherche sur la nature et les technologies (FQRNT).

<sup>†</sup>ISRO and Faculty of Social Sciences, Politics and Economics, Université Libre de Bruxelles, Campus Plaine, C.P. 210, Bd du Triomphe, B-1050 Bruxelles, Belgium (e-mail: gmelard@ulb.ac.be).

<sup>‡</sup>Département de mathématiques et de statistique and Centre de recherches mathématiques, C.P. 6128, succ. Centre-ville, Montréal, Québec, H3C 3J7, Canada (e-mail: royr@dms.umontreal.ca).

<sup>§</sup>Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, succ. Centre-ville, Montreal, Québec, H3C 3J7, Canada (e-mail: saidiabd@dms.umontreal.ca).



### **Abstract**

The exact likelihood function of a Gaussian vector autoregressive-moving average (VARMA) model is evaluated in two non standard cases: (a) a parsimonious structured form, such as obtained in the echelon form structure or the scalar component model (SCM) structure; (b) a partially non stationary (unit root) model in error-correction form. The starting point is Shea's algorithm (1987, 1989) for standard stationary and invertible VARMA models. Our algorithm also provides the parameter estimates and their standard errors. The small sample properties of our algorithm were studied by Monte Carlo methods. Examples with real data are provided.

**Keywords:** Gaussian likelihood estimation, ARMA echelon form, Scalar component model, Cointegrated model, Kalman filter, Chandrasekhar-type recursions.

### **Résumé**

La fonction de vraisemblance gaussienne exacte d'un modèle autorégressif-moyenne mobile vectoriel (VARMA) est évalué dans deux cas non standard : (a) une représentation parcimonieuse structurée comme la représentation ARMA échelon ou celle des modèles à composantes scalaires (SCM) ; (b) la représentation à correction d'erreur d'un modèle partiellement non stationnaire (à racine unitaire). Le point de départ est l'algorithme de Shea (1987, 1989) pour l'estimation d'un modèle VARMA standard, stationnaire et inversible. Notre algorithme fournit aussi les estimations des paramètres et leurs écarts-type. Les propriétés de notre algorithme pour de courtes séries ont été étudiées par des méthodes de Monte Carlo. Des exemples avec des données réelles sont présentées.



# 1 Introduction

The purpose of this paper is to extend the exact maximum likelihood estimation procedure for the Gaussian vector autoregressive-moving average (VARMA) model when the model is stated in a parsimonious structured form or when the process has unit roots.

Stationary VARMA models of order  $(p, q)$  and dimension  $k$  are characterized by a large number of parameters which are the  $k \times k$  elements of the  $p + q$  coefficient matrices plus the  $k(k + 1)/2$  elements of the innovation covariance matrix. Some elements of the coefficient matrices may be specified (*e.g.* to equal zero) without effect on the computational procedure. Detailed algorithms have been provided independently by Solo (1984), Dugré *et al.* (1986), Shea (1987, 1989), Mittnik (1991), Mauricio (1995, 1997, 2002) and others. The original procedure given by Tunnicliffe Wilson (1973) and Reinsel (1979) was not based on the exact likelihood but rather on its conditional form where the first values of the innovations are assumed to be known, instead of being determined conditionally on the observed series. Some other slightly improved procedures have been suggested by Hillmer and Tiao (1979) (see also Tiao and Box, 1981), and Nicholls and Hall (1979). More recently, that approach has been transformed by Mauricio (1995, 1997) into an exact and computationally efficient algorithm. Asymptotically, all these estimation methods are equivalent but their small sample properties known by Monte Carlo studies in the univariate case (Ansley and Newbold, 1980) are quite different. Although the autoregressive (AR) coefficients are estimated roughly with the same success, the moving average (MA) parameters suffer from a higher bias when the exact likelihood is not used. Some of the exact algorithms make use of a state space form and the fast Chandrasekhar-type equation recursions of Morf *et al.* (1974) (see also Lindquist, 1974, and Rissanen, 1973). The latter recursions were first advocated for that purpose in the case of univariate time series by Caines and Rissanen (1974) and Pearlman (1980), and implemented in that case by Mélard (1984) and Shea (1984). The general principle of using the state space approach and the Kalman filter for evaluating the likelihood of time series models goes back to Schweppe (1965) but the innovations form has first been used by Rissanen and Barbosa (1969) and Aasnaes and Kailath (1973), see also Ansley (1979), Mélard (1985) and Brockwell and Davis (1991). Less computationally efficient estimation procedures for VARMA models rely on the Kalman filter but they are also useful in the case of missing data, see Ansley and Kohn (1983).

Since the number of parameters of VARMA models is large, several problems arise: (a) the statistical precision of the estimators is low, hence the possibility of interpreting the fitted model is reduced; (b) the numerical optimization procedure is more prone to failure; (c) the computation time is large. The severity of the problems is higher when mixed models are used (*i.e.* models with both an autoregressive, or AR, and a moving average, or MA, polynomials). In the univariate case, it is known that the accuracy of the estimates suffers in the case of near canceling roots in the AR and MA polynomials. With multivariate models, the plausibility of near equal roots increases with  $k$ ,  $p$  and  $q$ .

The remedy which has been found is to obtain a more parsimonious specification. This is easier said than done. Many parsimonious specification procedures have been proposed since the initial stepwise model building approach of Tiao and Box (1981). It seems that two approaches emerge at the present time, (a) the echelon form structure implied by Kronecker indices, and (b) the scalar component model structure (SCM) of Tiao and Tsay (1989). In addition, several other specification procedures have been proposed recently, based upon a canonical ARMA echelon form (Tsay, 1991, Poskitt, 1992, Nsiri and Roy, 1992, 1996).

It will be shown in Section 2 that the echelon form and the SCM share the same general appearance, despite the fact that the two specification procedures are not equivalent and generally do not give rise to the same model specification. Tsay (1989b) has argued that the SCM may lead to more parsimonious parameterizations but there is little theoretical justification nor a large scale empirical study which confirms that conjecture yet.

Therefore, we shall handle both approaches simultaneously. Tsay (1989b) described an estimation procedure on an example which unfortunately cannot be generalized. Indeed, he made use of an algorithm for standard VARMA models, by multiplying both sides of the equation of the echelon form representation by a matrix which depends on the parameters, therefore leading to a two-stage method. Reinsel (1997)

has given a description of several procedures for estimating VARMA model. One of these procedures is for VARMA models under linear restrictions with a specialization to VARMA models in the echelon canonical form. However, the procedure is not based on the exact likelihood but rather on its conditional version. Therefore, the procedure may suffer from common known problems of conditional maximum likelihood especially in the case of short series and when the roots of the MA operator are near the unit circle. This is very frequent with univariate time series and there is no reason why things should improve in the multivariate case. For instance, in the example given by Reinsel (1997, p. 164), the MA polynomial is nearly noninvertible. Dufour and Jouini (2003) have proposed a two-stage linear procedure for estimating stationary and invertible VARMA echelon form models.

As we have explained earlier, the evaluation of the likelihood is based on recurrence equations. These equations need initial values which need to be determined. When a state space approach is invoked, these initial conditions are related to the stable solution of the state space model, for instance some Riccati equation for the covariance matrix of the state vector in the case where the Kalman filter is used. In the univariate case, the determination of these initial conditions has been the weakest point of the algorithm of Gardner *et al.* (1980). A much better solution has been found by Akaike (1978) and Jones (1980), and also by Mélard (1984) who has proposed to make use of the autocorrelations of the ARMA models computed with the algorithm of Tunnicliffe Wilson (1979). The latter is an improvement over the direct approach of McLeod (1975). Demeure and Mullis (1990) developed a simpler algorithm. In the multivariate case, no equivalent procedure has been found yet in order to obtain the initial conditions. Shea (1987, 1989) makes use essentially of the direct procedure first described formally by Ansley (1980), for computing the serial covariance matrices of a VARMA process, even without the improvement suggested by Kohn and Ansley (1982). Nevertheless, some progress has been obtained recently with the contributions of Mittnik (1990, 1993) and Tunnicliffe Wilson (1993). Harti, Mélard and Pham (2004) have provided an algorithm which seems to be superior, at least for high dimensional vectors. See also Söderström *et al.* (1998).

The case of a VARMA process with unit roots (also called a co-integrated VARMA process) is generally handled through an error correction form (Engle and Granger, 1987). It is supposed that the determinant of the autoregressive polynomial has one or several roots equal to 1 and the other roots lying outside of the unit circle. It is equivalent to saying that the autoregressive polynomial evaluated at point 1 is a singular matrix. The error correction factor is generally of reduced rank and is often parameterized as a product of two rectangular matrices. The parameters of the model are the autoregressive coefficients of the error-correction form (related but not identical with those of the corresponding VARMA representation), the moving average coefficients, the two rectangular matrices (although there may be some normalization in order to obtain a unique representation) and the innovation covariance matrix. The parameters are generally estimated using least squares (in the pure autoregressive case, Ahn and Reinsel, 1990), approximated least squares (*e.g.* Poskitt and Lütkepohl, 1995) or conditional maximum likelihood (Reinsel, 1997, Yap and Reinsel, 1995). However, the small sample properties of these methods are not good for the moving average coefficients, at least. Yap and Reinsel (1995) discuss full-rank and reduced-rank estimation of the error-correction form of the model, both using a Gaussian conditional maximum likelihood approach. They obtain the asymptotic properties in both cases and consider the likelihood ratio test for testing the number  $d$  of unit roots. They observe that, although the addition of the MA terms does not alter the asymptotic distribution of the likelihood ratio test statistic, finite sample performance of the test is affected by the MA terms.

In this paper, we describe computational procedures for the evaluation of the exact likelihood of parsimonious stationary VARMA models and partially nonstationary unit root VARMA models. In the case of a VARMA model written under a parsimonious echelon form or as a SCM, this implies mainly that the zero-lag coefficients are no longer the unit matrix but invertible matrices. Under the SCM form, the zero-lag coefficient is supposed to be obtained using the original canonical correlation approach. Two procedures are given, one based on the Kalman filter which remains useful in the case where at least one observation is missing (although we do not include a detailed treatment) and the other one based on the Chandrasekhar-type recursions, which is slightly more efficient. In the case of a unit root VARMA model, the parameters which are estimated are those of the error correction form. A transformation of the data is used and it does depend on the parameters.

To summarize, it happens that the two parsimonious structured forms and the unit root model have somewhat related features as far as the evaluation of the exact Gaussian likelihood is concerned. In our implementation, we have therefore treated the three cases together and added them to a procedure designed for stationary VARMA models in standard form. We have used the published program of Shea (1989) as a starting point. It is based on Chandrasekhar-type recursions. We have kept the approximate form of the likelihood proposed called ‘switching to constant recurrences’ by Gardner *et al.* (1980). Note however that the conditional method which is advertised by Shea (1989) does not work as such and has been modified.

Empirical evidence is shown on two aspects. First, examples given by Tiao and Tsay (1989), Reinsel (1997), Yap and Reinsel (1995) have been fitted using the new procedure: the mink and muskrat data, the flour price data, the quarterly AAA corporate bonds and commercial paper series (Haugh and Box, 1977), the U.S. monthly housing data (Hillmer and Tiao, 1979), the U.S. monthly interest rates (Stock and Watson, 1988) and a simulated series (Nsiri and Roy, 1996). Second, limited Monte Carlo comparisons between our procedure and the conditional procedure described by Reinsel (1997), based on the accuracy of the estimators are presented in two cases: the echelon form and the unit root ARMA model. The models in these examples have been previously fitted using a conditional maximum likelihood method or a two-stage method within which not all the parameters have been estimated by exact maximum likelihood.

The paper is subdivided as follows. In Section 2, we summarize the concept of a structured VARMA model in relation to a specification procedure, with the main emphasis on the echelon form based on Kronecker indices and the SCM. In Section 3, we describe partially nonstationary unit root VARMA models and explain how they can be transformed in standard stationary VARMA models. In Section 4, we derive the recurrence equations with respect to time and the initial conditions which contribute to the evaluation of the exact likelihood of VARMA models of that type. In Section 5, we describe briefly the new implementations. In Section 6, we handle some examples drawn from the literature. In Section 7, a Monte Carlo simulation study is conducted, establishing the superiority of the exact maximum likelihood method over the conditional one. Section 8 is devoted to concluding remarks.

## 2 Structured VARMA models

In the usual VARMA setup, we suppose that the  $k$ -dimensional process  $\{Y_t : t \in \mathbb{Z}\}$  is a solution of the following recurrence equation:

$$Y_t - \sum_{h=1}^p \Phi_h Y_{t-h} = \epsilon_t - \sum_{h=1}^q \Theta_h \epsilon_{t-h}, \quad (2.1)$$

where  $\{\epsilon_t\}$  is the innovation process, a sequence of independent and identically distributed (iid) random vectors with mean 0 and an invertible covariance matrix  $\Sigma$ . Denote by  $B$ , the lag operator, such that  $B^h Y_t = Y_{t-h}$ , the autoregressive and moving average polynomials by  $\Phi(z) = I_k - \Phi_1 z - \dots - \Phi_p z^p$ , and  $\Theta(z) = I_k - \Theta_1 z - \dots - \Theta_q z^q$ . We suppose that the VARMA model is identifiable (*e.g.* Hannan, 1969; Hannan and Deistler, 1988, Chap. 2) and invertible, that is the roots of  $\det\{\Theta(z)\}$  are greater than one in modulus. In the stationary case, the roots of  $\det\{\Phi(z)\}$  must also be outside the unit disk.

In the sequel, we assume normality for the distribution of the innovations but with non-Gaussian innovations whose density satisfies some regularity conditions, the quasi-likelihood estimators obtained by maximising the Gaussian likelihood have the same asymptotic properties as in the Gaussian case (*e.g.* Brockwell and Davis, 1991, Chap. 8).

### 2.1 The ARMA echelon form representation

More generally, consider a  $k$ -dimensional process  $\{Y_t : t \in \mathbf{Z}\}$ , which is wide-sense stationary, purely non-deterministic and centered. If  $\{\epsilon_t : t \in \mathbf{Z}\}$  denotes the innovation process of  $\{Y_t\}$ , then we have

$$Y_t = \epsilon_t + \sum_{i \geq 1} \Psi_i \epsilon_{t-i}, \quad (2.2)$$

where the  $\Psi_i$  are  $k \times k$  matrices such that  $\sum_{i \geq 1} \|\Psi_i\|^2 < \infty$  and  $\|\cdot\|$  denotes the Euclidean matrix norm (or a compatible one). The covariance matrix  $\Sigma$  of  $\epsilon_t$  is assumed to be invertible. This implies, in particular, that the matrices  $\Psi_i$  are uniquely defined by Equation (2.2). Let us define the vector of future values  $\mathcal{F}_{t+1}^\infty$  and the vector of past and present values  $\mathcal{P}_{-\infty}^t$  of the process  $\{Y_t\}$ , at time  $t$ , by

$$\mathcal{F}_{t+1}^\infty = (Y'_{t+1}, Y'_{t+2}, \dots)' \quad , \quad \mathcal{P}_{-\infty}^t = (Y'_t, Y'_{t-1}, \dots)'.$$

Let us denote  $Y_{t+i|t}$  the orthogonal projection, component by component, of  $Y_{t+i}$  onto the space generated by the components of  $\mathcal{P}_{-\infty}^t$ . It follows from (2.2) that

$$F_{t+1|t} = \Psi_\infty^\infty E_t \tag{2.3}$$

where

$$F_{t+1|t} = \begin{bmatrix} Y_{t+1|t} \\ Y_{t+2|t} \\ \vdots \end{bmatrix} \quad , \quad \Psi_\infty^\infty = \begin{bmatrix} \Psi_1 & \Psi_2 & \dots \\ \Psi_2 & \Psi_3 & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \quad , \quad E_t = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \\ \vdots \end{bmatrix}.$$

The matrix  $\Psi_\infty^\infty$  is called the Hankel matrix associated with the sequence  $\{\Psi_i\}$ , whereas the space generated by the components of  $F_{t+1|t}$  is called the space of predictors and is denoted  $\mathcal{P}_t$ . The dimension of  $\mathcal{P}_t$  is called the dynamic dimension or the McMillan degree of the process  $\{Y_t\}$ . We can show that a necessary and sufficient condition for the process  $\{Y_t\}$  to admit an ARMA representation is that its dynamic dimension be finite. Also when  $\dim \mathcal{P}_t < \infty$ , for any basis of  $\mathcal{P}_t$  we can find a corresponding ARMA representation of  $\{Y_t\}$  (see for example Deistler 1985 or Gouriéroux and Monfort 1990, chap. 8). A natural way to choose such a basis, when  $\dim \mathcal{P}_t = n$ , is to consider the one formed by the first  $n$  linearly independent components of the vector of predictors  $F_{t+1|t}$ . If these components are found at positions  $i_1, \dots, i_n$ , then it is easy to show that the set  $I_Y = \{i_1, \dots, i_n\}$  has the following property

$$(\forall i \in \mathbf{N}) : i \notin I_Y \Rightarrow (i+k) \notin I_Y . \tag{2.4}$$

We deduce from the above property that there exist non negative integers  $n_1, \dots, n_k$ , such that  $I_Y = \{1, 1+k, \dots, 1+(n_1-1)k; 2, 2+k, \dots, 2+(n_2-1)k; \dots; k, k+k, \dots, k+(n_k-1)k\}$ , with the convention that  $n_i = 0$  if and only if  $i \notin I_Y$ , for  $i = 1, \dots, k$ . The integers  $n_1, \dots, n_k$  are called the Kronecker indices, or the dynamic indices of the process  $\{Y_t\}$ . The sum of these indices is equal to the dynamic dimension of the process. Moreover, since it is assumed that the matrix  $\Sigma$  is regular, it follows from (2.3) that linear independence among the components of  $F_{t+1|t}$  is equivalent to linear independence of the corresponding rows of  $\Psi_\infty^\infty$ . Thus, to determine the Kronecker indices, it is sufficient to determine the basis formed by the first  $n$  linearly independent rows of  $\Psi_\infty^\infty$ . The ARMA representation associated with this basis is called the echelon form canonical representation, (see Deistler, 1985, or Hannan and Deistler, 1988).

These indices allow us to define, in a unique manner, an ARMA( $r, r$ ) model

$$\tilde{\Phi}_0 Y_t - \sum_{h=1}^r \tilde{\Phi}_h Y_{t-h} = \tilde{\Theta}_0 \epsilon_t - \sum_{h=1}^r \tilde{\Theta}_h \epsilon_{t-h},$$

referred to as an ARMA echelon form model, which is abbreviated by ARMA $_E$ . The order  $r$  is such that  $r = \max n_i, i = 1, \dots, k$ . The elements  $\tilde{\Phi}_{ij}(z)$  and  $\tilde{\Theta}_{ij}(z)$  of the matrix polynomials  $\tilde{\Phi}(z) = \tilde{\Phi}_0 - \sum_{h=1}^r \tilde{\Phi}_h z^h$  and  $\tilde{\Theta}(z) = \tilde{\Theta}_0 - \sum_{h=1}^r \tilde{\Theta}_h z^h$  of the ARMA $_E$  model are of the form, see Tsay (1991),

$$\tilde{\Phi}_{ij}(z) = \delta_{ij} - \sum_{h=n_i+1-n_{ij}}^{n_i} \tilde{\Phi}_{h,ij} z^h \tag{2.5}$$

and

$$\tilde{\Theta}_{ij}(z) = \delta_{ij} - \sum_{h=n_i+1-m_{ij}}^{n_i} \tilde{\Theta}_{h,ij} z^h, \quad (2.6)$$

where the  $\tilde{\Phi}_{h,ij}$  and  $\tilde{\Theta}_{h,ij}$  are real parameters,

$$n_{ij} = \begin{cases} \min(n_i + 1, n_j) & \text{if } j < i, \\ \min(n_i, n_j) & \text{otherwise,} \end{cases}$$

and

$$m_{ij} = \begin{cases} n_i + 1 & \text{if } i > j \text{ and } n_i < n_j, \\ n_i & \text{otherwise.} \end{cases}$$

Given that  $\tilde{\Phi}_0 = \tilde{\Theta}_0$ , relation (2.6) can be equivalently written as (see Nsiri and Roy, 1992),

$$\tilde{\Theta}_{ij}(z) = \tilde{\Phi}_{0,ij} - \sum_{h=1}^{n_i} \tilde{\Theta}_{h,ij} z^h.$$

Once we know the Kronecker indices  $n_1, \dots, n_k$  of the process  $\{Y_t\}$ , we can define the ARMA<sub>E</sub> model (ARMA echelon form model) by employing relations (2.5) and (2.6). The  $i$ -th equation of this model is given by

$$\tilde{\Phi}_{0,i} Y_t - \sum_{h=1}^{n_i} \tilde{\Phi}_{h,i} Y_{t-h} = \tilde{\Theta}_{0,i} \epsilon_t - \sum_{h=1}^{n_i} \tilde{\Theta}_{h,i} \epsilon_{t-h} \quad (2.7)$$

where  $\tilde{\Phi}_{h,i}$  and  $\tilde{\Theta}_{h,i}$  respectively denote the  $i$ -th row of the matrices  $\tilde{\Phi}_h$  and  $\tilde{\Theta}_h$ . The relation (2.7) can be written in the ‘‘refined’’ form

$$\tilde{\Phi}_{0,i} Y_t - \sum_{h=1}^{p_i} \tilde{\Phi}_{h,i} Y_{t-h} = \tilde{\Theta}_{0,i} \epsilon_t - \sum_{h=1}^{q_i} \tilde{\Theta}_{h,i} \epsilon_{t-h},$$

where  $p_i$  and  $q_i$  are such that  $\tilde{\Phi}_{p_i,i} \neq 0$  and  $\tilde{\Theta}_{q_i,i} \neq 0$ , and thus we have  $\max(p_i, q_i) = n_i$ .

The echelon form structure leads to

$$(\tilde{\Phi}_0 - \sum_{j=1}^p \tilde{\Phi}_j B^j) Y_t = (\tilde{\Theta}_0 - \sum_{j=1}^q \tilde{\Theta}_j B^j) \epsilon_t, \quad (2.8)$$

where  $\tilde{\Phi}_0 = \tilde{\Theta}_0$  is a lower triangular matrix with 1’s along the main diagonal. In addition, when a refined ARMA echelon form is obtained, there are some linear constraints, generally that some coefficients are equal to zero, thereby reducing the number of parameters that need to be estimated. Note that (2.8) can be transformed in (2.1), by letting  $\Phi_j = \tilde{\Phi}_0^{-1} \tilde{\Phi}_j$ ,  $j = 1, \dots, p$ ,  $\Theta_h = \tilde{\Phi}_0^{-1} \tilde{\Theta}_h$ ,  $h = 1, \dots, q$ .

## 2.2 The scalar component model

Another way to write down a canonical form is given by the so-called scalar component model (SCM) approach. Let us consider a VARMA( $p, q$ ) process satisfying (2.1). A linear combination  $X_t^{(1)} = v_1' Y_t$  follows a SCM( $p_1, q_1$ ) model where  $p_1 \leq p$  and  $q_1 \leq q$ , if  $v_1' \Phi_{p_1} \neq 0$ ,  $v_1' \Phi_j = 0$  for  $j = p_1 + 1, \dots, p$ , and  $v_1' \Theta_{q_1} \neq 0$ ,  $v_1' \Theta_j = 0$  for  $j = q_1 + 1, \dots, q$ . Hence

$$v_1' Y_t - \sum_{j=1}^{p_1} v_1' \Phi_j Y_{t-j} = v_1' \epsilon_t - \sum_{j=1}^{q_1} v_1' \Theta_j \epsilon_{t-j}.$$

If we can write  $k$  such independent linear combinations and stack the vectors  $v'_1, \dots, v'_k$  in an invertible matrix  $\tilde{\Phi}_0$ , it corresponds to the following equation

$$(\tilde{\Phi}_0 - \sum_{j=1}^p \tilde{\Phi}_j B^j) Y_t = (\tilde{\Phi}_0 - \sum_{j=1}^q \tilde{\Theta}_j B^j) \varepsilon_t, \quad (2.9)$$

where some of the rows of  $\tilde{\Phi}_j$  and  $\tilde{\Theta}_j$  are equal to zero. The SCM approach was introduced by Tiao and Tsay (1989) and Tsay (1989a); see also Reinsel (1997) and Tsay (1991). We are thus led to an equation which bears some resemblance which (2.8) except that now the matrix  $\tilde{\Phi}_0$  is not considered as being composed of parameters but rather is deduced from the data in order to simplify the model. Left multiplying (2.9) by  $\tilde{\Phi}_0^{-1}$ , the usual VARMA form (2.1) is obtained but the corresponding coefficients  $\tilde{\Phi}_0^{-1} \tilde{\Phi}_j$  and  $\tilde{\Phi}_0^{-1} \tilde{\Theta}_j$  have a particular reduced-rank structure.

For convenience, we use the same notations for the coefficients of the ARMA<sub>E</sub> and SCM representations (2.8) and (2.9) respectively. However, it is important to remember that these two specification procedures generally give rise to two different models for a given series. Indeed in (2.8),  $\tilde{\Phi}_0$  is a lower triangular matrix whilst in (2.9),  $\tilde{\Phi}_0$  is unrestricted.

### 3 Unit root VARMA processes

Let us again consider a VARMA model (2.1) but assume that  $\det \{\Phi(z)\} = 0$  has  $d$  unit roots (roots equal to 1),  $0 < d < k$ , and that all other roots lie outside the unit circle. It is also assumed that  $\text{rank}\{\Phi(1)\} = k - d$ , and that the roots of  $\det \{\Theta(z)\} = 0$  are all outside the unit circle.

Following Engle and Granger (1987) and Yap and Reinsel (1995), we use the parameterization provided by the error-correction representation

$$\Phi^*(B)(I_k - B)Y_t = CY_{t-1} + \Theta(B)\varepsilon_t, \quad (3.1)$$

where

$$\Phi^*(z) = I_k - \Phi_1^* z - \dots - \Phi_{p-1}^* z^{p-1}, \quad \Phi_j^* = - \sum_{l=j+1}^p \Phi_l,$$

and  $C = -\Phi(1) = -(I_k - \sum_{j=1}^p \Phi_j)$ . We shall use a reduced-rank parameterization structure for  $C$  advocated by Ahn and Reinsel (1990) and Yap and Reinsel (1995), among others, by taking  $C = C_1 C_2$ , where  $C_1$  and  $C_2$  are full-rank matrices of dimensions  $k \times (k - d)$  and  $(k - d) \times k$ , respectively. In order to obtain a unique parameterization,  $C_2$  can be normalized, possibly by interchanging the components of  $Y_t$ , so that  $C_2 = [I_{k-d}, C_0]$ , where  $C_0$  is a  $(k - d) \times d$  matrix. The parameters of the unit root VARMA model are the coefficients in  $\Phi^*(B)$ ,  $\Theta(B)$ ,  $C_1$ ,  $C_0$ , and the innovation covariance matrix  $\Sigma$ . It is possible to write the model in the VARMA form (2.1), using  $\Phi(B) = \Phi^*(B)(I_k - B) - CB$ . Although there is a one-to-one mapping from  $\Phi_1, \dots, \Phi_p$  to  $\Phi_1^*, \dots, \Phi_{p-1}^*, C$ , the parameterization provided by the error correction form seems to be easier.

Using the Jordan canonical form of  $I_k + C$ , Ahn and Reinsel (1990), and Yap and Reinsel (1995) have suggested a conditional method for estimating the error correction representation. Contrarily to what they say, the assumption on  $\Phi(1)$  does not imply that  $\sum_{j=1}^p \Phi_j = I_k + C$  has  $d$  unit eigenvalues.

There are examples where that procedure does not work, as Pham, Roy and Cédra (2003) have pointed out. We shall therefore follow the approach described in this latter paper. Let  $P_1$  be a  $k \times d$  matrix such that  $\Phi(1)P_1 = 0$ , and  $P_2$  be any  $k \times (k - d)$  matrix such that the  $k \times k$  matrix  $P = [P_1, P_2]$  is invertible. Let  $Q = [Q'_1, Q'_2]'$  be the inverse of  $P$ , with  $Q_1$  and  $Q_2$  having respectively  $d$  and  $k - d$  rows. Then, denoting  $\Phi^{**}(B) = \Phi^*(B)(I_k - P_2 Q_2 B) + \Phi(1)B$ , we have  $\Phi(B) = \Phi^{**}(B)(I_k - P_1 Q_1 B)$  and  $\det \{\Phi(z)\} = (1 - z)^d \det \{\Phi^{**}(z)\}$ . By the assumption on the roots of  $\det \{\Phi(z)\}$ , all the roots of  $\det \{\Phi^{**}(z)\}$  are outside the unit circle. From Pham, Roy and Cédra (2003, Section 2), we also have

$$(I_k - P_1 Q_1 B)Y_t = \Phi^{**}(B)^{-1} \Theta(B) \varepsilon_t, \quad (3.2)$$

which can be written, since  $Q_1 P_1 = I_{k-d}$ ,  $Q_2 P_1 = 0$ ,

$$Q_1(Y_t - Y_{t-1}) = Q_1 \Phi^{**}(B)^{-1} \Theta(B) \epsilon_t, \quad (3.3)$$

$$Q_2 Y_t = Q_2 \Phi^{**}(B)^{-1} \Theta(B) \epsilon_t. \quad (3.4)$$

Let us define the transformed process  $X_t = \begin{bmatrix} X_t^{(1)} \\ X_t^{(2)} \end{bmatrix} = Q Y_t$ , with  $X_t^{(1)} = Q_1 Y_t$  and  $X_t^{(2)} = Q_2 Y_t$ . Hence,  $X_t^{(1)}$  has stationary first differences whereas  $X_t^{(2)}$  is stationary. Note also that  $\Phi(1)Y_t = \Phi(1)P_2 X_t^{(2)}$  shows that there are  $k-d$  linear combinations of  $Y_t$  that are jointly stationary. Here, we shall use (3.2) using the transformation  $\tilde{X}_t = (I_k - P_1 Q_1 B) Y_t$ . The process  $\tilde{X}_t$  has a stationary and invertible VARMA representation.

Yap and Reinsel's (1995) contribution already mentioned in the introduction plus the known fact that MA estimation is better with the exact quasi-likelihood than with conditional likelihood or the least squares method, has led us to investigate the computation of the exact Gaussian likelihood for model (3.1) for given values of the parameters  $\Phi^*(B)$ ,  $\Theta(B)$ ,  $C_1$ ,  $C_0$ , and  $\Sigma$ . The procedure is as follows:

1. given  $\Phi(1)$ , obtain  $P = [P_1, P_2]$  and  $Q = [Q_1', Q_2']' = P^{-1}$ ;
2. compute  $\tilde{X}_t = (I_k - P_1 Q_1 B) Y_t$ ;
3. evaluate the coefficients of  $\Phi^{**}(B) = \Phi^*(B)(I_k - P_2 Q_2 B) + \Phi(1)B$ ;
4. compute the Gaussian likelihood of the model  $\Phi^{**}(B) \tilde{X}_t = \Theta(B) \epsilon_t$ ;
5. if necessary, evaluate the autoregressive polynomial  $\Phi(B) = \Phi^{**}(B)(I_k - P_1 Q_1 B)$  of the original unit root VARMA model.

Of course, the matrices  $P$  and  $Q$  are not unique. The following Theorem, whose proof is in the appendix, ensures the unicity of  $\tilde{X}_t$  and  $\Phi^{**}(B)$ , and consequently the unicity of representation (3.2).

**Theorem 3.1** *Let  $\Phi(1)$  be a  $k \times k$  matrix such that  $\text{rank}\{\Phi(1)\} = k - d = r$ , where  $0 < d < k$ . Let  $P_1$  be any  $k \times d$  matrix such that  $\Phi(1)P_1 = 0$  and  $P_2$  be any  $k \times r$  matrix such that  $P = [P_1, P_2]$  is invertible and the columns of  $P_2$  are orthogonal to those of  $P_1$ . If  $P^{-1} = Q = [Q_1', Q_2']'$  stands for the inverse of  $P$ , with  $Q_1$  and  $Q_2$  having  $d$  and  $r$  rows respectively, then the matrices  $P_1 Q_1$  and  $P_2 Q_2$  are uniquely defined.*

## 4 Evaluation of the likelihood

### 4.1 Some recurrences with respect to time

In this section, we obtain recurrences with respect to time in order to evaluate the Gaussian likelihood function of the model, given a time series  $\{Y_t; t = 1, \dots, T\}$ . As mentioned in Section 1, one of the most computationally efficient way to handle the problem is by using the Chandrasekhar-type recursions, provided that the model can be written under an invariant state space form. However, we start with the Kalman filter recursions which are more general, and can be used for instance when there are missing data (although we shall not cover that case here, see Ansley and Kohn, 1983).

We shall follow the presentation given by Shea (1987), using a state vector  $\alpha_t$  of dimension  $rk \times 1$ , where  $r = \max(p, q)$ . We suppose that  $\tilde{\Phi}_0$  is invertible. When we have the echelon form,  $\tilde{\Phi}_0$  is a lower triangular matrix with 1's along the main diagonal, then its inverse  $\tilde{\Phi}_0^{-1}$  is computed by elimination and from (2.8), we write

$$(I_k - \sum_{j=1}^p \tilde{\Phi}_0^{-1} \tilde{\Phi}_j B^j) Y_t = (I_k - \sum_{j=1}^q \tilde{\Phi}_0^{-1} \tilde{\Theta}_j B^j) \epsilon_t. \quad (4.1)$$

For the scalar component model of Section 2.2,  $\tilde{\Phi}_0$  is obtained using a canonical analysis (see Tsay 1989) and is not considered as a parameter of the model. Besides that, the treatment is similar.

Let us denote  $\Phi_j = \tilde{\Phi}_0^{-1}\tilde{\Phi}_j$ ,  $j = 1, \dots, p$ ,  $\Phi_j = 0$ ,  $j = p + 1, \dots, r$ ,  $\Theta_j = \tilde{\Phi}_0^{-1}\tilde{\Theta}_j$ ,  $j = 1, \dots, q$ ,  $\Theta_j = 0$ ,  $j = q + 1, \dots, r$ . The state space form is written

$$Y_t = h'\alpha_t + \epsilon_t \quad (4.2)$$

$$\alpha_t = T\alpha_{t-1} + R\epsilon_{t-1}, \quad (4.3)$$

where

$$T = \begin{pmatrix} \Phi_1 & I_k & & & \\ \Phi_2 & & I_k & & \\ \vdots & & & \ddots & \\ \Phi_r & 0_k & & \cdots & 0_k \end{pmatrix}, \quad R = \begin{pmatrix} \Phi_1 - \Theta_1 \\ \Phi_2 - \Theta_2 \\ \vdots \\ \Phi_r - \Theta_r \end{pmatrix}, \quad \text{and} \quad h = \begin{pmatrix} I_k \\ 0_k \\ \vdots \\ 0_k \end{pmatrix}. \quad (4.4)$$

Let us denote by  $\mathcal{F}_1^t$  the Hilbert space spanned by the components of  $\{Y_1, Y_2, \dots, Y_t\}$  with the covariance as the scalar product. The vector composed of the orthogonal projection of the components of  $\alpha_t$  in the sub-space  $\mathcal{F}_1^{t-1}$  is denoted by  $\hat{\alpha}_{t|t-1}$ . Let  $V_t$  be the difference between  $\alpha_t$  and  $\hat{\alpha}_{t|t-1}$  which is orthogonal to  $\mathcal{F}_1^{t-1}$ . It is called the sample innovation at time  $t$ . Let  $F_t$  be its covariance matrix  $F_t = E(V_t V_t')$ .

The  $\hat{\alpha}_{t|t-1}$  and the  $F_t$  are computed using the Kalman filter recurrences as follows

$$\hat{\alpha}_{t|t-1} = T\hat{\alpha}_{t-1|t-2} + K_{t-1}F_{t-1}^{-1}V_{t-1}, \quad (4.5)$$

$$K_t = TP_{t-1}hF_{t-1}^{-1}, \quad (4.6)$$

$$P_t = TP_{t-1}T' + K_{t-1}F_{t-1}^{-1}K_{t-1}', \quad (4.7)$$

$$F_t = h'P_t h, \quad (4.8)$$

$$V_t = Y_t - h'\hat{\alpha}_{t|t-1}, \quad (4.9)$$

where  $K_t$  and  $P_t$  are auxiliary matrices of respective dimensions  $rk \times k$ ,  $rk \times rk$ . Note that it is the need to update the matrix  $P_t$  that prevents the Kalman filter from being competitive when the Chandrasekhar recurrences are valid.

The  $\hat{\alpha}_{t|t-1}$  and the  $F_t$  can also be computed using the Chandrasekhar recurrences as follows

$$\hat{\alpha}_{t|t-1} = T\hat{\alpha}_{t-1|t-2} + K_{t-1}F_{t-1}^{-1}V_{t-1}, \quad (4.10)$$

$$K_t = K_{t-1} + TL_{t-1}M_{t-1}L_{t-1}'h, \quad (4.11)$$

$$F_t = F_{t-1} + h'L_{t-1}M_{t-1}L_{t-1}'h, \quad (4.12)$$

$$M_t = M_{t-1} - M_{t-1}L_{t-1}'hF_{t-1}^{-1}h'L_{t-1}M_{t-1}, \quad (4.13)$$

$$L_t = TL_{t-1} - K_{t-1}F_{t-1}^{-1}h'L_{t-1}, \quad (4.14)$$

$$V_t = Y_t - h'\hat{\alpha}_{t|t-1}, \quad (4.15)$$

where  $K_t$ ,  $M_t$ , and  $L_t$  are auxiliary matrices of respective dimensions  $rk \times k$ ,  $k \times k$ , and  $rk \times k$ . The initial conditions are described in the Section 4.2.

With the notations introduced, the logarithm of the Gaussian likelihood function from time 1 to time  $T$  is expressed by

$$\log L = -\frac{Tk}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^T \log(\det F_t) - \frac{1}{2}\sum_{t=1}^T V_t'F_t^{-1}V_t. \quad (4.16)$$

## 4.2 Derivation of the initial conditions

The initial conditions of the Chandrasekhar recurrences are given by:  $\hat{\alpha}_{1|0} = 0$ ,  $K_1 = L_1 = TP_{1|0}h + R\Sigma$ ,  $F_1 = h'P_{1|0}h + \Sigma$  and  $M_1 = -F_1^{-1}$  where  $P_{1|0}h = E(\alpha_1\alpha_1')h$ . Because  $h$  contains  $I_k$  on the first block and 0 otherwise, we only evaluate the first  $k$  columns of  $P_{1|0}$ . From Jones (1980), we have

$$\alpha_t^i = \sum_{j=i}^p \Phi_j Y_{t+i-j-1} - \sum_{j=i}^q \Theta_j \epsilon_{t+i-j-1},$$

$$(P_{1|0}h)^i = \sum_{j=i}^p \Phi_j \Gamma(j-i+1) - \sum_{j=i}^q \Theta_j \delta'(j-i+1), \quad i = 1, \dots, r,$$

where  $\delta(j) = E(Y_t \varepsilon'_{t-j}) = -\Theta_j \Sigma + \sum_{i=1}^{\min(j,p)} \Phi_i \delta(j-i)$  and  $\Gamma(j) = E(Y_{t-j} Y'_t)$ , the autocovariance function,  $j = 1, 2, \dots, q$ , ( $\delta(0) = \Sigma$ ). The matrices  $\Gamma(j)$ ,  $j = 1, 2, \dots, p$ , are calculated using the the procedure mentioned below.

For univariate processes, the computation of the initial conditions of the Chadrsekhar recursions is reasonably fast, due to the existence of fast algorithms (see Mélard, 1984, Kohn and Ansley, 1985), which need approximately  $O(p^2)$  operations instead of  $O(p^3)$  for the direct method described by McLeod (1975). This is not true for multivariate processes. The method of Ansley (1980), which is used by Shea (1989) consists of solving a linear system of  $pk^2$  equations and is of order  $O(p^3 k^6)$ . Kohn and Ansley (1982) have given an improved algorithm by slightly reducing the number of equations. Mittnik (1993) has proposed a block-Levinson procedure using a pure moving average representation of the process which requires a number of operations of order  $O(p^2 d^6)$ . Tunnicliffe Wilson (1993) has tried to improve that procedure but the method is iterative and is therefore not exact.

## 5 Implementations

Two Fortran 77 implementations have been compared in order to check the algorithms and the corresponding programs. Both are based on the original Shea's (1989) program. Implementation A was initiated by Harti (1996) and is home made except the optimization routine which is based on the library GQOPT (Goldfeld and Quandt, 1972). With respect to the Shea (1989) program, the following changes have been implemented in A:

1. the procedure described in Harti, Mélard and Pham (2004) has been used for computing the autocovariance matrices, since it is generally faster and more secure than Shea's (1989) COVARs;
2. instead of a sufficient condition for stationarity, we have used the necessary and sufficient condition, which is provided free by the algorithm of Harti, Mélard and Pham (2004);
3. a necessary and sufficient condition of invertibility based on the condition for stationarity but applied on the moving average coefficients instead of the autoregressive coefficients is also checked in the program.

Implementation B relies on the NAG library (The Numerical Algorithms Group, 1999). Indeed, the subroutine E04UCF of the NAG library was used as optimization routine. For the unit root case, F02WEF was used to compute the matrix  $P = [P_1, P_2]$ . Moreover, to compute matrix inverses, the subroutines F07ADF and F07AJF of NAG library were used. A check for stationarity and invertibility is performed using G13DXF. The source files are available at the first author's web page (<http://homepages.ulb.ac.be/~gmelard>).

Both implementations make use of the following features:

1. instead of the elements of the innovation covariance matrix, a  $k \times k$  lower triangular matrix  $U$  has been used as a parameter, so that  $\Sigma = U'U$ ;
2. some modifications have been operated in routines DECOMP and SOLVE (Moler, 1972) in order to avoid comparisons of real numbers with 0;
3. the original conditional maximum likelihood procedure included in Shea's (1989) program does not work and has been corrected.

The empirical results presented here were obtained using implementation B but those of implementation A are very close. The simulations described in Section 7 were produced using implementation B. For our vector of parameters, the optimization procedure gives the observed Fisher information (or Hessian) matrix of the estimators. Their estimated standard errors are computed by inverting the observed Hessian matrix at the final estimate of the parameters.

## 6 Empirical comparisons

This Section illustrates the proposed algorithm by fitting models that have been used in the literature. These models are fitted by the conditional method (Reinsel, 1997), and the exact maximum likelihood method. For the SCM model, a two-stage method is used as explained above. In the following, the symbol ‘X’, indicates a coefficient to be estimated, ‘0’ or ‘1’, that the corresponding coefficient is identically equal to 0 or 1, respectively, and ‘\*’ means that the coefficient is either a fixed constant different from 0 and 1, or is dependent on another coefficient of the model.

**Example 1** This is the artificial bivariate series of length  $T = 120$  used by Nsiri and Roy (1996) to illustrate their method for identifying the refined echelon form representation of a VARMA model. The series was generated from the model

$$Y_t - \begin{bmatrix} 0.6 & 0.0 \\ 0.5 & -0.5 \end{bmatrix} Y_{t-1} = \epsilon_t - \begin{bmatrix} 0.8 & -0.2 \\ 0.0 & 0.0 \end{bmatrix} \epsilon_{t-1} - \begin{bmatrix} -0.85 & -0.8 \\ 0.0 & 0.0 \end{bmatrix} \epsilon_{t-2}, \quad (6.1)$$

where the normal innovations  $\{\epsilon_t\}$  have mean zero and  $\Sigma = I_2$ . The dynamic dimension is,  $n = 2$ , the Kronecker indices are  $n_1 = 2$ ,  $n_2 = 1$ , and  $\tilde{\Phi}(0) = I_2$ . The method of Nsiri and Roy (1996) has given the right values for the Kronecker indices. We have fitted model (6.1) with the appropriate ARMA<sub>E</sub> specification except that we have added the element (2,1) of  $\tilde{\Phi}(0)$  as an additional parameter. More precisely we fit the following ARMA<sub>E</sub> model

$$\begin{bmatrix} 1 & 0 \\ X & 1 \end{bmatrix} Y_t - \begin{bmatrix} X & 0 \\ X & X \end{bmatrix} Y_{t-1} = \begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix} \epsilon_t - \begin{bmatrix} X & X \\ 0 & 0 \end{bmatrix} \epsilon_{t-1} - \begin{bmatrix} X & X \\ 0 & 0 \end{bmatrix} \epsilon_{t-2}.$$

The conditional method has given the following results, with the estimated standard errors (see Section 5) given after the  $\pm$  symbol,

$$\begin{aligned} & \begin{bmatrix} 1 & 0 \\ 0.090 & 1 \\ (\pm 0.055) & \end{bmatrix} Y_t - \begin{bmatrix} 0.555 & 0 \\ (\pm 0.065) & -0.533 \\ 0.570 & (\pm 0.041) \\ (\pm 0.051) & \end{bmatrix} Y_{t-1} \\ & = \epsilon_t - \begin{bmatrix} 0.731 & -0.165 \\ (\pm 0.066) & (\pm 0.083) \\ 0 & 0 \end{bmatrix} \epsilon_{t-1} - \begin{bmatrix} -0.902 & -0.798 \\ (\pm 0.067) & (\pm 0.082) \\ 0 & 0 \end{bmatrix} \epsilon_{t-2}, \end{aligned}$$

with the estimate of  $\text{vec}(\Sigma) : ( 1.102 \ 0.081 \ 0.081 \ 1.006 )'$ . The exact maximum likelihood has given the following results

$$\begin{aligned} & \begin{bmatrix} 1 & 0 \\ 0.086 & 1 \\ (\pm 0.057) & \end{bmatrix} Y_t - \begin{bmatrix} 0.557 & 0 \\ (\pm 0.068) & -0.532 \\ 0.565 & (\pm 0.041) \\ (\pm 0.051) & \end{bmatrix} Y_{t-1} \\ & = \epsilon_t - \begin{bmatrix} 0.745 & -0.166 \\ (\pm 0.072) & (\pm 0.081) \\ 0 & 0 \end{bmatrix} \epsilon_{t-1} - \begin{bmatrix} -0.890 & -0.816 \\ (\pm 0.069) & (\pm 0.083) \\ 0 & 0 \end{bmatrix} \epsilon_{t-2}, \end{aligned}$$

with the estimate of  $\text{vec}(\Sigma) : ( 1.068 \ 0.073 \ 0.073 \ 1.000 )'$ .

The two methods give very close results. This conclusion is not surprising given the sample size considered. The exact estimators are closer to the true values than the conditional ones, except in one case.

**Example 2** This is the example of the logarithms of the mink and muskrat data (Reinsel, 1997, p. 164), with  $T = 62$ . Using the Kronecker indices, Reinsel obtained the following echelon form representation

$$\begin{aligned} \begin{bmatrix} 1 & 0 \\ X & 1 \end{bmatrix} Y_t - \begin{bmatrix} X & 0 \\ X & X \end{bmatrix} Y_{t-1} - \begin{bmatrix} X & X \\ 0 & 0 \end{bmatrix} Y_{t-2} = \begin{bmatrix} X \\ X \end{bmatrix} \\ + \begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix} \epsilon_t - \begin{bmatrix} X & X \\ X & X \end{bmatrix} \epsilon_{t-1} - \begin{bmatrix} X & X \\ 0 & 0 \end{bmatrix} \epsilon_{t-2}, \end{aligned}$$

and he used that specification to illustrate his conditional method for fitting structured models. After eliminating a few parameters whose estimates were not significant (they are indicated by 0.000), he obtained the following results

$$\begin{aligned} \hat{\Phi}_0 = \hat{\Theta}_0 &= \begin{bmatrix} 1 & 0 \\ 0.955 & 1 \\ (\pm 0.260) & \end{bmatrix}, & \hat{\Phi}_1 &= \begin{bmatrix} 1.307 & 0 \\ (\pm 0.148) & \\ 0.000 & 0.984 \\ & (\pm 0.140) \end{bmatrix}, \\ \hat{\Phi}_2 &= \begin{bmatrix} -0.704 & 0.046 \\ (\pm 0.086) & (\pm 0.038) \\ 0 & 0 \end{bmatrix}, & \hat{\Theta}_1 &= \begin{bmatrix} 0.789 & -0.666 \\ (\pm 0.127) & (\pm 0.113) \\ 0.000 & -0.913 \\ & (\pm 0.196) \end{bmatrix}, \\ \hat{\Theta}_2 &= \begin{bmatrix} -0.289 & 0.323 \\ (\pm 0.091) & (\pm 0.102) \\ 0 & 0 \end{bmatrix} & \text{and } \hat{\Sigma} &= \begin{bmatrix} 0.0423 & 0.0199 \\ 0.0199 & 0.0580 \end{bmatrix}. \end{aligned}$$

With the exact method, we have obtained the following results

$$\begin{aligned} \hat{\Phi}_0 = \hat{\Theta}_0 &= \begin{bmatrix} 1 & 0 \\ 1.063 & 1 \\ (\pm 0.253) & \end{bmatrix}, & \hat{\Phi}_1 &= \begin{bmatrix} 1.336 & 0 \\ (\pm 0.142) & \\ 0.000 & 0.840 \\ & (\pm 0.203) \end{bmatrix}, \\ \hat{\Phi}_2 &= \begin{bmatrix} -0.677 & 0.031 \\ (\pm 0.065) & (\pm 0.027) \\ 0 & 0 \end{bmatrix}, & \hat{\Theta}_1 &= \begin{bmatrix} 0.901 & -0.748 \\ (\pm 0.166) & (\pm 0.117) \\ 0.000 & -1.245 \\ & (\pm 0.253) \end{bmatrix}, \\ \hat{\Theta}_2 &= \begin{bmatrix} -0.380 & 0.498 \\ (\pm 0.153) & (\pm 0.190) \\ 0 & 0 \end{bmatrix} & \text{and } \hat{\Sigma} &= \begin{bmatrix} 0.0407 & 0.0198 \\ 0.0198 & 0.0566 \end{bmatrix}. \end{aligned}$$

In general, the two estimators are closer for the autoregressive parameters than for the moving average parameters. This is not surprising since the sample size is small.

**Example 3** The data consist of the logarithms of indices of monthly flour prices in three cities, Buffalo, Minneapolis and Kansas City, over the period from August 1972 to November 1980. In this case,  $T = 100$ . Tiao and Tsay (1989) have specified a VARMA(1,1) model using the SCM approach. The SCM representation (SCM(1,0), SCM(1,0), SCM(1,1)) is given by the following equation

$$\tilde{\Phi}_0 Y_t - \begin{bmatrix} X & X & X \\ X & X & X \\ X & X & X \end{bmatrix} Y_{t-1} = \begin{bmatrix} X \\ X \\ X \end{bmatrix} + \tilde{\Theta}_0 \epsilon_t - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ X & X & X \end{bmatrix} \epsilon_{t-1}. \quad (6.2)$$

The transformation matrix  $\tilde{\Phi}_0$  obtained by Tiao and Tsay (1989) is

$$\tilde{\Phi}_0 = \tilde{\Theta}_0 = \begin{bmatrix} -0.40 & 0.83 & -0.40 \\ 0.61 & -0.51 & -0.60 \\ 0.55 & 0.83 & -0.06 \end{bmatrix}.$$

Probably because the package which they used for model fitting did not cope with the presence of  $\tilde{\Phi}_0$ , they used a two-stage approach. More precisely, after the computation of the matrix  $\tilde{\Phi}_0$  and given that the transformation  $X_t = \tilde{\Phi}_0 Y_t$  does not alter the parameter specification, they fitted the model for the transformed series  $X_t$ . Replacing in their results  $X_t$  by  $\tilde{\Phi}_0 Y_t$ , we obtain

$$\begin{aligned} & \begin{bmatrix} -0.40 & 0.83 & -0.40 \\ 0.61 & -0.51 & -0.60 \\ 0.55 & 0.83 & -0.06 \end{bmatrix} Y_t - \begin{bmatrix} -0.364 & 0.741 & -0.340 \\ 0.531 & -0.271 & -0.722 \\ 0.774 & 0.136 & 0.352 \end{bmatrix} Y_{t-1} = \begin{bmatrix} -0.02 \\ -0.16 \\ 0.26 \end{bmatrix} \\ & + \begin{bmatrix} -0.40 & 0.83 & -0.40 \\ 0.61 & -0.51 & -0.60 \\ 0.55 & 0.83 & -0.06 \end{bmatrix} \epsilon_t - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1.518 & -1.332 & -0.069 \end{bmatrix} \epsilon_{t-1}. \end{aligned}$$

Using the exact method of this paper, we obtain the following fitted model

$$\begin{aligned} & \begin{bmatrix} -0.40 & 0.83 & -0.40 \\ 0.61 & -0.51 & -0.60 \\ 0.55 & 0.83 & -0.06 \end{bmatrix} Y_t - \begin{bmatrix} -0.362 & 0.738 & -0.345 \\ 0.494 & -0.235 & -0.733 \\ 0.825 & 0.075 & 0.374 \end{bmatrix} Y_{t-1} = \begin{bmatrix} -0.007 \\ -0.124 \\ 0.215 \end{bmatrix} \\ & + \begin{bmatrix} -0.40 & 0.83 & -0.40 \\ 0.61 & -0.51 & -0.60 \\ 0.55 & 0.83 & -0.06 \end{bmatrix} \epsilon_t - \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1.421 & -1.256 & -0.062 \end{bmatrix} \epsilon_{t-1}. \end{aligned}$$

The results are close but not identical because the so-called exact estimation method of SCA (Lui and Hudak, 1986) computes, in fact, the exact likelihood for an autoregressive model, not for an ARMA model.

**Example 4** We consider the bivariate time series of seasonally adjusted monthly U.S. housing data consisting of housing starts and housing sold during the period from 1965 to 1974 (Hilmer and Tiao, 1979), with  $T = 120$ . The error correction form model specified by Reinsel (1997, p. 206) is

$$(1 - B)Y_t = \begin{bmatrix} X \\ X \end{bmatrix} \begin{bmatrix} 1 & X \end{bmatrix} Y_{t-1} + \epsilon_t \quad (6.3)$$

and he obtained the following estimates

$$\hat{C}_1 = \begin{bmatrix} -0.523 \\ 0.141 \end{bmatrix}, \hat{C}_2 = \begin{bmatrix} 1 & -1.872 \end{bmatrix}, \hat{\Sigma} = \begin{bmatrix} 26.59 & 5.97 \\ 5.97 & 9.87 \end{bmatrix}.$$

The model fitted by the exact (conditionally on the first observation) Gaussian maximum likelihood method gives the following results

$$\hat{C}_1 = \begin{bmatrix} -0.517 \\ (\pm 0.082) \\ 0.142 \\ (\pm 0.048) \end{bmatrix}, \hat{C}_2 = \begin{bmatrix} 1 & -1.872 \\ & (\pm 0.086) \end{bmatrix}, \hat{\Sigma} = \begin{bmatrix} 26.69 & 6.03 \\ 6.03 & 9.87 \end{bmatrix}.$$

This is not surprising because the two methods are generally close for estimating pure VAR models even if the sample size is small.

**Example 5** The data consists of the quarterly AAA corporate bonds and commercial paper series from 1953 to 1970 (Haugh and Box, 1977; Reinsel, 1997, p. 306), where  $T = 72$ . The model specified in error correction form by Reinsel (1997, p. 213) is

$$(1 - B)Y_t = \begin{bmatrix} X \\ X \end{bmatrix} \begin{bmatrix} 1 & X \end{bmatrix} Y_{t-1} + \begin{bmatrix} X & X \\ X & X \end{bmatrix} (1 - B)Y_{t-1} + \begin{bmatrix} X & X \\ X & X \end{bmatrix} (1 - B)Y_{t-2} + \epsilon_t.$$

The results obtained using the conditional method are as follows

$$\hat{C} = \hat{C}_1 \hat{C}_2 = \begin{bmatrix} -0.039 \\ 0.040 \end{bmatrix} \begin{bmatrix} 1 & -1.376 \end{bmatrix},$$

$$\hat{\Phi}_1^* = \begin{bmatrix} 0.628 & -0.060 \\ 1.186 & 0.507 \end{bmatrix}, \quad \hat{\Phi}_2^* = \begin{bmatrix} -0.496 & 0.072 \\ -0.739 & -0.083 \end{bmatrix},$$

with the residual covariance matrix estimate  $\hat{\Sigma} = \begin{bmatrix} 0.0189 & 0.0270 \\ 0.0270 & 0.0838 \end{bmatrix}$ .

The exact Gaussian maximum likelihood method gives the following results

$$\hat{C} = \hat{C}_1 \hat{C}_2 = \begin{bmatrix} -0.041 \\ (\pm 0.021) \\ 0.085 \\ (\pm 0.044) \end{bmatrix} \begin{bmatrix} 1 & -1.265 \\ & (\pm 0.063) \end{bmatrix},$$

$$\hat{\Phi}_1^* = \begin{bmatrix} 0.687 & -0.072 \\ (\pm 0.131) & (\pm 0.067) \\ 1.284 & 0.476 \\ (\pm 0.277) & (\pm 0.129) \end{bmatrix}, \quad \hat{\Phi}_2^* = \begin{bmatrix} -0.465 & 0.073 \\ (\pm 0.152) & (\pm 0.065) \\ -0.691 & -0.038 \\ (\pm 0.310) & (\pm 0.128) \end{bmatrix},$$

with the residual covariance matrix estimate  $\hat{\Sigma} = \begin{bmatrix} 0.0188 & 0.0261 \\ 0.0261 & 0.0798 \end{bmatrix}$ .

The conclusions are similar to those of Example 4.

**Example 6** Our final example consists in three U.S. monthly interest rates from 1960 to 1979 (Stock and Watson, 1988; Reinsel, 1997, p. 307), where  $T = 240$ . Yap and Reinsel (1995) fitted the following reduced rank model in error correction form

$$(1 - B)Y_t = \begin{bmatrix} X & X \\ X & X \\ X & X \end{bmatrix} \begin{bmatrix} 1 & 0 & X \\ 0 & 1 & X \end{bmatrix} Y_{t-1} + \epsilon_t - \begin{bmatrix} X & X & X \\ X & X & X \\ X & X & X \end{bmatrix} \epsilon_{t-1}.$$

They used that specification to illustrate their conditional estimation method for fitting partially nonstationary ARMA models. They obtained the following results.

$$\hat{C} = \hat{C}_1 \hat{C}_2 = \begin{bmatrix} -0.199 & 0.250 \\ 0.023 & -0.082 \\ 0.041 & 0.027 \end{bmatrix} \begin{bmatrix} 1 & 0 & -1.398 \\ 0 & 1 & -1.148 \end{bmatrix}, \quad \hat{\Theta}_1 = \begin{bmatrix} -0.143 & 0.237 & -0.463 \\ -0.224 & 0.118 & -0.317 \\ -0.125 & 0.037 & -0.330 \end{bmatrix}.$$

The model fitted by the exact Gaussian maximum likelihood method gives the following results

$$\hat{C} = \hat{C}_1 \hat{C}_2 = \begin{bmatrix} -0.156 & 0.287 \\ (\pm 0.054) & (\pm 0.155) \\ 0.011 & -0.123 \\ (\pm 0.040) & (\pm 0.119) \\ -0.013 & -0.006 \\ (\pm 0.040) & (\pm 0.104) \end{bmatrix} \begin{bmatrix} 1 & 0 & -1.199 \\ & & (\pm 0.108) \\ 0 & 1 & -1.008 \\ & & (\pm 0.040) \end{bmatrix},$$

$$\hat{\Theta}_1 = \begin{bmatrix} -0.014 & -0.225 & -0.193 \\ (\pm 0.076) & (\pm 0.262) & (\pm 0.246) \\ -0.208 & 0.086 & -0.190 \\ (\pm 0.067) & (\pm 0.158) & (\pm 0.141) \\ -0.167 & -0.025 & -0.195 \\ (\pm 0.059) & (\pm 0.144) & (\pm 0.135) \end{bmatrix}, \quad \hat{\Sigma} = \begin{bmatrix} 0.2125 & 0.0650 & 0.0678 \\ 0.0650 & 0.1022 & 0.0830 \\ 0.0678 & 0.0830 & 0.0918 \end{bmatrix}.$$

This example confirms the general fact that the exact method and the conditional one lead to different estimates of the moving average parameters even for a moderate sample size.

## 7 Some Monte Carlo results

In this Monte Carlo study, we mainly investigate the finite sample properties of both the conditional method and the Gaussian exact maximum likelihood method. We compare their performances and we check the adequacy of their standard errors. In order to do that, we realized two different experiments to illustrate the results in each of the two cases: ARMA echelon and unit root models. The standard ARMA case was omitted because it was widely studied by Shea (1989). In the first experiment we consider a bivariate  $\text{ARMA}_E(1,4)$  model that satisfies both the stationarity and invertibility conditions. In the second experiment we consider a trivariate partially non-stationary  $\text{ARMA}(1,4)$  model with one unit root ( $d=1$ ). The two series lengths considered were 50 and 100. For each of these two experiments and for each of the two series lengths, the G05HDF subroutine of the NAG library (NAG, 1995) was used to generate Gaussian series from the corresponding standard VARMA data generating process (Barone, 1987, and Shea, 1988). Then, suitable transformations were applied to obtain the series from the non standard VARMA processes, For each replication, the two estimation methods (exact and conditional) were performed using the implementations of Section 5.

In order to perform a valid comparison, for both experiments, we kept only the first 100 replications for which

- (i) the optimization procedure E04UCF of the NAG library reaches an optimal solution for both methods (conditional and exact), *i.e.* E04UCF ended without error indicators or warnings for both methods;
- (ii) for both methods, the optimal solution obtained satisfies the stationarity and invertibility conditions.

Moreover, to insure that initial values of the parameters do not affect our conclusions, those values for both methods were taken equal to the true parameter values.

For each parameter, the following results are reported:

- (1) MEAN: the average of the estimates over the replications;
- (2) OSD: the Observed Standard Deviation that is defined as the square root of the average over the replications of the variances obtained by the optimization procedure;
- (3) ESD: the Empirical Standard Deviation across the replications.

### 7.1 Experiment 1

In the first experiment, we generate data from the following  $\text{ARMA}_E$  model

$$Y_t - \begin{bmatrix} 0.9 & 0.0 \\ 0.2 & 0.9 \end{bmatrix} Y_{t-1} = \epsilon_t - \begin{bmatrix} 0.0 & 0.0 \\ 0.0 & 0.1 \end{bmatrix} \epsilon_{t-3} - \begin{bmatrix} 0.9 & -0.8 \\ 0.0 & 0.0 \end{bmatrix} \epsilon_{t-4},$$

where the innovations  $\{\epsilon_t\}$  have mean zero and  $\Sigma = I_2$ . The main purpose is to fit the following  $\text{ARMA}_E$  model

$$\begin{bmatrix} 1 & 0 \\ X & 1 \end{bmatrix} Y_t - \begin{bmatrix} X & 0 \\ X & X \end{bmatrix} Y_{t-1} = \begin{bmatrix} 1 & 0 \\ * & 1 \end{bmatrix} \epsilon_t - \begin{bmatrix} 0 & 0 \\ 0 & X \end{bmatrix} \epsilon_{t-3} - \begin{bmatrix} X & X \\ 0 & 0 \end{bmatrix} \epsilon_{t-4}, \quad (7.1)$$

for each replication. Of course, the (2,1)-element of  $\tilde{\Phi}_0$  is a parameter in the model. The results are reported in Table 1. The latter is presented as a  $3 \times 1$  block matrix, the first block gives the true parameter values and the second and third ones present the results of the two estimation methods (conditional and exact) for the sample sizes 50 and 100 respectively. For each estimated coefficient, we provide the MEAN, OSD and ESD based on 100 independent realizations. For the coefficients that are not estimated such as the elements of  $\tilde{\Phi}_0$  except (2,1), or the element (1,2) of  $\tilde{\Phi}_1$  and some elements in  $\tilde{\Theta}_3$ , and  $\tilde{\Theta}_4$ , the MEAN corresponds to the true value and we put a dash for the OSD and ESD. The lower triangular matrix  $U$  being estimated instead of  $\Sigma$ , we provide the results for the elements of  $U$ . Since  $\Sigma = I_2$  in this model, the

true  $U$  is also equal to  $I_2$ . The OSD and ESD are presented only for the lower part of  $U$ . To underline the differences between the two methods, the conditional and exact estimates averages that are far apart or far from the true value were put in bold.

Inspection of Table 1 reveals that, for both sample sizes, the averages obtained from the exact method are closer to the true values than those obtained by the conditional method. The greatest differences between the two methods appear for the (1, 1)-element of  $\tilde{\Theta}_4$  and the (2, 1)-element of  $\tilde{\Theta}_3$ . The conditional method is particularly inefficient for estimating  $\tilde{\Theta}_4$ . Note that for both methods, as the sample size increases, the averages become closer to the true values. Moreover, for the series length  $T = 100$ , the bias of the exact estimates are smaller than those of the conditional ones. Note also that for the larger sample size ( $T = 100$ ), the OSD and ESD are close for all parameters. However, with  $T = 50$ , for a few parameters, the OSD and ESD differ when the exact method is performed. Experiments for other sample sizes were considered but the results are not reported. They indicate that the OSD and ESD are much closer as the sample size increases.

			The true model									
			$\tilde{\Phi}_0$		$\tilde{\Phi}_1$		$\tilde{\Theta}_3$		$\tilde{\Theta}_4$		$U$	
			1	0	.900	0	0	0	.900	-.800	1.000	0
			.000	1	.200	.900	0	.100	0	0	.000	1.000
			The estimated model									
Size	Method		$\tilde{\Phi}_0$		$\tilde{\Phi}_1$		$\tilde{\Theta}_3$		$\tilde{\Theta}_4$		$U$	
50	COND	MEAN	1	0	.858	0	0	0	<b>.640</b>	-.746	1.166	0
			-.057	1	.173	.880	0	<b>.078</b>	—	—	.094	1.054
		OSD	—	—	.065	—	—	—	.134	.149	.116	0
			.159	—	.140	.043	—	.125	—	—	.155	.106
		ESD	—	—	.075	0	—	—	.166	.194	.168	—
			.127	—	.111	.044	—	.136	—	—	.246	.192
	EXACT	MEAN	1	0	.874	0	0	0	<b>.813</b>	-.854	.998	0
			-.057	1	.171	.883	0	<b>.108</b>	—	—	-.009	.927
		OSD	—	—	.056	—	—	—	.233	.150	.130	0
			.117	—	.105	.035	—	.127	—	—	.154	.098
		ESD	—	—	.065	0	—	—	.141	.145	.120	—
			.119	—	.106	.040	—	.141	—	—	.167	.010
100	COND	MEAN	1	0	.880	0	0	0	<b>.741</b>	-.775	1.112	0
			-.020	1	.190	.891	0	<b>.085</b>	—	—	.059	1.050
		OSD	—	—	.039	—	—	—	.082	.086	.076	0
			.089	—	.080	.023	—	.075	—	—	.105	.074
		ESD	—	—	.040	0	—	—	.119	.117	.092	—
			.079	—	.069	.025	—	.080	—	—	.143	.133
	EXACT	MEAN	1	0	.887	0	0	0	<b>.900</b>	-.825	.990	0
			-.019	1	.189	.893	0	<b>.102</b>	—	—	-.004	.983
		OSD	—	—	.035	—	—	—	.116	.078	.077	0
			.073	—	.063	.019	—	.073	—	—	.107	.071
		ESD	—	—	.034	0	—	—	.062	.077	.063	—
			.069	—	.064	.022	—	.067	—	—	.106	.072

**Table 1.** Comparison of the conditional (COND) and EXACT likelihood estimation methods for the bivariate ARMA echelon model (7.1) based on 100 replications for each of the two sample sizes. For each estimated parameter, we report the mean of the 100 estimates (MEAN), the observed standard deviation (OSD) and the empirical standard deviation (ESD).

## 7.2 Experiment 2

In this experiment, we generate data from a partially non-stationary ARMA(1,4) model with one unit root

$$Y_t - \begin{bmatrix} 0.602 & 0.433 & 0.110 \\ 0.121 & 0.660 & 0.066 \\ 0.103 & 0.166 & 0.838 \end{bmatrix} Y_{t-1} = \epsilon_t - \begin{bmatrix} 0.6 & 0 & 0.2 \\ 0 & 0.9 & 0 \\ -0.4 & -0.4 & 0.8 \end{bmatrix} \epsilon_{t-4},$$

where the innovations  $\{\epsilon_t\}$  have zero mean and

$$\Sigma = \begin{bmatrix} 1.0 & 0.5 & 0.4 \\ 0.5 & 1.0 & 0.7 \\ 0.4 & 0.7 & 1.0 \end{bmatrix}, \quad \text{so} \quad U = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 0.866 & 0 \\ 0.4 & 0.577 & 0.712 \end{bmatrix}.$$

			The true model								
			$C_1$		$C_0$	$\Theta_4$			$U$		
			-.398	.433	-.800	.600	0	.200	1.000	0	0
			.121	-.340	-.480	0	.900	0	.500	.897	0
			.103	.166		-.400	-.400	.800	.400	.577	.712
			The estimated model								
Size	Method		$C_1$		$C_0$	$\Theta_4$			$U$		
50	COND	MEAN	-.430	.411	-.811	.573	0	.172	.992	0	0
			.119	-.430	-.488	0	<b>.742</b>	0	.548	.897	0
			.157	.131		-.411	<b>-.091</b>	<b>.432</b>	.367	.454	.816
		OSD	.109	.133	.217	.122	—	.116	.103	—	—
			.112	.141	.134	—	.088	—	.145	.090	—
			.114	.138		.149	.162	.140	.147	.132	.080
	ESD	.138	.169	.231	.136	—	.134	.096	—	—	
		.162	.177	.158	—	.099	—	.138	.110	—	
		.119	.144		.169	.198	.200	.171	.167	.108	
	EXACT	MEAN	-.414	.419	-.798	.587	0	.253	.957	0	0
			.123	-.401	-.484	0	<b>.862</b>	0	.510	.834	0
			.143	.135		-.452	<b>-.551</b>	<b>.980</b>	.394	.566	.582
OSD		.101	.126	.167	.137	—	.146	.105	—	—	
		.100	.132	.105	—	.086	—	.136	.092	—	
		.105	.128		.150	.220	.254	.147	.122	.095	
ESD	.112	.150	.210	.147	—	.150	.097	—	—		
	.120	.141	.144	—	.060	—	.121	.095	—		
	.117	.143		.173	.228	.204	.155	.141	.092		
100	COND	MEAN	-.409	.437	-.793	.598	0	.179	1.001	0	0
			.129	-.371	-.472	0	<b>.799</b>	0	.533	.901	0
			.132	.155		-.409	<b>-.151</b>	<b>.520</b>	.393	.492	.806
		OSD	.068	.084	.055	.068	—	.073	.071	—	—
			.065	.086	.034	—	.048	—	.099	.062	—
			.072	.088		.085	.096	.102	.099	.089	.053
	ESD	.071	.092	.053	.074	—	.079	.074	—	—	
		.076	.094	.039	—	.060	—	.109	.061	—	
		.072	.100		.101	.136	.140	.108	.100	.071	
	EXACT	MEAN	-.410	.446	-.800	.611	0	.229	.980	0	0
			.119	-.352	-.478	0	<b>.894</b>	0	.510	.854	0
			.122	.152		-.404	<b>-.495</b>	<b>.931</b>	.392	.585	.642
OSD		.063	.079	.045	.066	—	.075	.072	—	—	
		.051	.071	.022	—	.035	—	.096	.063	—	
		.065	.082		.076	.114	.121	.101	.085	.057	
ESD	.065	.088	.038	.073	—	.078	.075	—	—		
	.059	.080	.018	—	.032	—	.103	.054	—		
	.072	.091		.083	.130	.145	.109	.081	.064		

**Table 2.** Comparison of the conditional (COND) and EXACT likelihood estimation methods for the trivariate nonstationary ARMA(1,4) model (7.2) based on 100 replications for each of the two sample sizes. For each estimated parameter, we report the mean of the 100 estimates (MEAN), the observed standard deviation (OSD) and the empirical standard deviation (ESD).

We can easily check that the matrices  $C_1$  and  $C_0$  that characterize the *error-correction* form are given by

$$C_1 = \begin{bmatrix} -0.398 & 0.433 \\ 0.121 & -0.340 \\ 0.103 & 0.166 \end{bmatrix} \quad \text{and} \quad C_0 = \begin{bmatrix} -0.80 \\ -0.48 \end{bmatrix}.$$

We fit the following *error-correction* form model

$$W_t = Y_t - Y_{t-1} = \begin{bmatrix} X & X \\ X & X \\ X & X \end{bmatrix} \begin{bmatrix} 1.0 & 0.0 & X \\ 0.0 & 1.0 & X \end{bmatrix} Y_{t-1} + \epsilon_t - \begin{bmatrix} X & 0 & X \\ 0 & X & 0 \\ X & X & X \end{bmatrix} \epsilon_{t-4}. \quad (7.2)$$

The first observation  $Y_0$  was set equal to zero. The likelihood function is conditional on this first observation. The results are reported in Table 2 whose presentation is similar to Table 1. For each sample size and each method, we report the average estimates of the matrices  $C_1$ ,  $C_0$ ,  $\Theta_4$ , and  $U$  and their respective OSD and ESD. The main conclusions are similar to those of Experiment 1. The exact method performs better than the conditional one for both sample sizes. The difference between the two methods is specially striking for  $\Theta_4$  at both series lengths. In particular, the true value of the (3,2)-element is  $-0.400$  whilst the conditional estimates averages are  $-0.091$  at  $T = 50$  and  $-0.151$  at  $T = 100$ . The corresponding exact estimates averages are respectively  $-0.551$  and  $-0.495$ .

## 8 Conclusions

This paper describes algorithms for exact maximum likelihood estimation of parsimonious VARMA models and partially nonstationary unit root VARMA. The main difference between the VARMA echelon and the standard VARMA representation is that a triangular matrix  $\tilde{\Phi}(0)$  is substituted to the identity matrix in the autoregressive and moving average polynomials and that constraints of nullity are imposed to some coefficients. The constraints do not have an impact on the likelihood function itself. However, the presence of  $\tilde{\Phi}(0)$  modifies the evaluation of the likelihood but the change is smaller than it could be anticipated.

Similarly, the coverage of unit root or cointegrated VARMA models for exact maximum likelihood estimation (conditional, of course on the the first observation of the vector series) require some minor changes in the code. In the two cases, these features are not available in the existing software packages.

We have been able to fit several models shown in the literature and to investigate in a small Monte Carlo study the effect of the use of the exact likelihood instead of the conditional likelihood. In particular, with short series of 50 and 100 observations, the difference between the conditional and the exact estimations can be quite large especially for high-order moving average parameters.

There should be no problem to handle unit root VARMA models in structured form although it is not currently implemented. A VARMA cointegrated model in echelon form can be specified using the method of Poskitt and Lütkepohl (1995) with Kronecker indices estimated using Bartel and Lütkepohl (1997). Another procedure for obtaining a parsimonious cointegration representation is due to Hunter and Dislis (1994).

## Appendix

*Proof of Theorem 3.1.* The singular value decomposition of  $\Phi(1)$  yields that

$$\Phi(1) = UDV' = [u_1, u_2, \dots, u_k] \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 \\ 0 & d_2 & 0 & & 0 \\ 0 & 0 & d_3 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & d_k \end{bmatrix} \begin{bmatrix} v'_1 \\ \vdots \\ v'_r \\ \vdots \\ v'_k \end{bmatrix},$$

where  $U$  and  $V$  are  $k \times k$  orthogonal matrices.

The null space of  $\Phi(1)$ , i.e., the set of vectors  $x$  such that  $\Phi(1)x = 0$ , is the space generated by the columns of the matrix  $G = [v_{r+1}, \dots, v_k]$ . Now let  $P_1$  and  $P_1^*$  be any  $k \times d$  matrices such that  $\Phi(1)P_1 = 0$

and  $\Phi(1)P_1^* = 0$ . Since  $P_1$  and  $P_1^*$  are full rank matrices, there exist full rank  $d \times d$  matrices  $\alpha$  and  $\alpha^*$  such that

$$P_1 = G\alpha \quad \text{and} \quad P_1^* = G\alpha^*.$$

Hence, we have

$$P_1^* = P_1\alpha^{-1}\alpha^*.$$

In a similar way, consider  $k \times r$  matrices  $P_2$  and  $P_2^*$  such that  $P = [P_1, P_2]$  and  $P^* = [P_1^*, P_2^*]$  are invertible, and the column vectors of  $P_2$  and  $P_2^*$  are orthogonal to those of  $P_1$  and  $P_1^*$  respectively. Of course, the space spanned by the column vectors of  $M = [v_1, \dots, v_r]$  is orthogonal to the one spanned by those of  $G$ . Hence, there exist full rank  $r \times r$  matrices  $\beta$  and  $\beta^*$  such that

$$P_2 = M\beta \quad \text{and} \quad P_2^* = M\beta^*.$$

Thus,

$$P_2^* = P_2\beta^{-1}\beta^*.$$

Let  $Q = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = P^{-1}$  and  $Q^* = \begin{bmatrix} Q_1^* \\ Q_2^* \end{bmatrix} = (P^*)^{-1}$ , we can easily check that

$$Q^* = \begin{bmatrix} (\alpha^*)^{-1}\alpha Q_1 \\ (\beta^*)^{-1}\beta Q_2 \end{bmatrix}.$$

Now turning back to the product  $P_1^*Q_1^*$ , we have that

$$P_1^*Q_1^* = P_1\alpha^{-1}\alpha^*(\alpha^*)^{-1}\alpha Q_1 = P_1Q_1.$$

In a similar way we can check that  $P_2^*Q_2^* = P_2Q_2$ , which completes the proof.

## References

- Aasnaes, H. B., and Kailath, T. (1973). An innovation approach to least-squares estimation - Part VII: Some applications of vector autoregressive moving average models. *IEEE Transactions on Automatic Control* **18**, 601-607.
- Ahn, S. K., and Reinsel, G. C. (1990). Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association* **85**, 813-823.
- Akaike, H. (1978). Covariance matrix computation of the state variable of stationary Gaussian process. *Annals of the Institute of Statistical Mathematics* **30**, 499-504.
- Ansley, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika* **66**, 59-65.
- Ansley, C. F. (1980). Computation of the theoretical autocovariance function for a vector ARMA process. *Journal of Statistical Computation and Simulation* **12**, 15-24.
- Ansley, C. F., and Newbold, P. (1980). Finite sample properties of estimators for autoregressive-moving average models. *Journal of Econometrics* **13**, 159-183.
- Ansley, C. F., and Kohn, R. (1983). Exact likelihood of vector autoregressive-moving average process with missing or aggregated data. *Biometrika* **70**, 275-278.
- Barone, P. (1987). A method for generating independent realization of a multivariate normal stationary and invertible ARMA( $p, q$ ) process. *Journal of Time Series Analysis* **8**, 125-130.
- Bartel, H. and Lütkepohl, H. (1997). Estimating the Kronecker indices of cointegrated echelon form VARMA models. *Econometrics Journal* **1**, C76-C99.
- Brockwell, P. J., and Davis, R. A. (1987). *Time Series: Theory and Methods*. New York: Springer-Verlag.
- Caines, P. E., and Rissanen, J. (1974). Maximum likelihood estimation of parameters in multivariate Gaussian stochastic processes. *IEEE Transactions on Information Theory* **20**, 102-104.

- Deistler M., (1985). General structure and parametrization of ARMA and state-space systems and its relation to statistical problems, in E.J. Hannan, P.R. Krishnaiah, M.M. Rao, Eds. *Handbook of Statistics*, Vol. 5. Amsterdam: North-Holland, pp. 257-277.
- Demeure, C. J., and Mullis, C. T. (1989). The Euclid algorithm and the fast computation of cross-covariance and autocovariance sequences. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **37**, 545-552.
- Dufour, J. M., and Jouini, T. (2003). Asymptotic distribution of a simple linear estimator for VARMA models in echelon form. In *Statistical Modeling and Analysis for Complex Data Problems* (Duchesne, P., and Rémillard, B., Eds.). Kluwer: The Netherlands. To appear.
- Dugré, J. P., Scharf, L. L., and Gueguen, C. (1986). Exact likelihood for stationary vector autoregressive moving average processes. *IEEE Transactions on Signal Processing* **11**, 105-118.
- Engle, R. F., and Granger, C. W. J. (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica* **55**, 251-276.
- Gardner, G., Harvey, A. C., and Phillips, G. D. A. (1980). An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. *Journal of the Royal Statistical Society C*, **29**, 311-322.
- Goldfeld, S. M., and Quandt, R. E. (1972). *Nonlinear Methods in Econometrics*. Amsterdam: North-Holland
- Gouriéroux, C., and Monfort, A. (1990). *Séries temporelles et modèles dynamiques*. Paris: Economica.
- Hannan, E. J. (1969). The identification of vector mixed autoregressive moving average systems. *Biometrika* **57**, 223-225.
- Hannan, E. J., and Deistler, M. (1988), *The Statistical Theory of Linear Systems*. New York: Wiley.
- Harti M. (1996). Algorithmes pour l'estimation par pseudo-maximum de vraisemblance exacte pour des modèles VARMA sous forme classique et sous forme structurée. Unpublished Ph.D. Thesis. Institut de Statistique et de Recherche Opérationnelle, Université Libre de Bruxelles.
- Harti, M., Melard, G., and Pham D. T. (2004). Computing covariances for scalar and vector ARMA processes. In preparation.
- Haugh, L. D., and Box, G. E. P. (1977). Identification of dynamic regression (distributed lag) models connecting two series. *Journal of the American Statistical Association* **72**, 121-130.
- Hillmer, S. C., and Tiao, G. C. (1979). Likelihood function of stationary multiple autoregressive moving average models. *Journal of the American Statistical Association* **74**, 652-660.
- Hunter, J., and Dislis, C. (1994). A parsimonious representation and VARMA estimator for cointegration. Working paper.
- Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22**, 389-395.
- Kohn, R., and Ansley, C. F. (1982). A note on obtaining the theoretical autocovariances of an ARMA process. *Journal of Statistical Computation and Simulation* **15**, 273-283.
- Kohn, R., and Ansley, C. F. (1985). Computing the likelihood and its derivatives for a Gaussian ARMA model. *Journal of Statistical Computation and Simulation* **22**, 229-263.
- Lindquist, A. (1974). A new algorithm for optimal filtering of discrete time stationary processes. *SIAM Journal on Control and Optimization* **12**, 736-746.
- Lui, L.-M. and Hudak, G.B. (1986). *The SCA Statistical System Analysis: Reference Manual for Forecasting and Time series Analysis*. Chicago: Scientific Computing Associates.
- Mauricio, J. A. (1995). Exact maximum likelihood estimation of stationary vector ARMA models. *Journal of the American Statistical Association* **90**, 282-291.
- Mauricio, J. A. (1997). The exact likelihood function of stationary vector autoregressive moving average model. *Journal of the Royal Statistical Society C*, **46**, 157-171.
- Mauricio, J. A. (2002). An algorithm for the exact likelihood of a stationary vector autoregressive-moving average model. *Journal of Time Series Analysis* **23**, 473-486.
- McLeod, A. I. (1975). Derivation of the theoretical autocovariance function of autoregressive-moving average time series. *Journal of the Royal Statistical Society C*, **24**, 255-256.

- Mélard, G. (1984). Algorithm AS197: A fast algorithm for the exact likelihood of autoregressive-moving average models. *Journal of the Royal Statistical Society C*, **33**, 104-114.
- Mélard, G. (1985). Analyse de données chronologiques, Séminaire de mathématiques supérieures, Séminaire scientifique OTAN (NATO Advanced Study Institute), Montréal: Presses de l'Université de Montréal.
- Mittnik, S. (1990). Computation of theoretical autocovariance matrices of multivariate autoregressive moving average time series. *Journal of the Royal Statistical Society B*, **52**, 151-155.
- Mittnik, S. (1991). Derivation of the unconditional state-covariance matrix for exact maximum-likelihood estimation of ARMA models. *Journal of Economic Dynamics and Control* **15**, 731-740.
- Mittnik, S. (1993). Computing theoretical autocovariances of multivariate autoregressive moving average models by using a block Levinson method. *Journal of the Royal Statistical Society B*, **55**, 435-440.
- Moler, C. B. (1972). Algorithm 423: Linear equation solver. *Commun Association for Computing Machinery* **15**, 274.
- Morf, M., Sidhu, G. S., and Kailath, T. (1974). Some new algorithms for recursive estimation on constant, linear, discrete-time systems. *IEEE Transactions on Automatic Control* **19**, 315-323.
- Nicholls, D. F., and Hall, A. D. (1979). The exact likelihood function of multivariate autoregressive moving average models. *Biometrika* **66**, 259-264.
- Nsiri, S., and Roy, R. (1992). On the identification of ARMA echelon form models. *Canadian Journal of Statistics* **20**, 369-386.
- Nsiri, S., and Roy, R. (1996). Identification of refined ARMA echelon form models for multivariate time series. *Journal of Multivariate Analysis* **56**, 207-231.
- Pearlman, J. G. (1980). An algorithm for the exact likelihood of a high-order autoregressive-moving average process. *Biometrika* **67**, 232-233.
- Pham, D. T., Roy R. and Cédras, L. (2003). Tests for non-correlation of two cointegrated ARMA time series. *Journal of Time Series Analysis* **24**, 553-577.
- Poskitt, D. S. (1992). Identification of echelon canonical form for vector linear processes using least squares. *Annals of Statistics* **20**, 195-215.
- Poskitt, D. S. and Lütkepohl, H. (1995). Consistent specification of cointegrated autoregressive moving-average systems. Preprint, Humbolt University, Berlin.
- Reinsel, G. C. (1979). FIML estimation of the dynamic simultaneous equations model with ARMA disturbances. *Journal of Econometrics* **9**, 263-281.
- Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*, Second Edition. New York: Springer-Verlag.
- Rissanen, J. (1973). Algorithm for triangular decomposition of block Hankel and Toeplitz matrices with application to factoring positive matrix polynomials. *Mathematics of Computation* **27**, 147-154.
- Rissanen, J., and Barbosa, L. (1969). Properties of infinite covariance matrices and stability of optimum predictors. *Information Sciences* **1**, 221-236.
- Schweppe, F. C. (1965). Evaluation of likelihood functions for Gaussian signals. *IEEE Transactions on Information Theory* **11**, 61-70.
- Shea, B. L. (1984). Maximum likelihood estimation of multivariate ARMA processes via the Kalman filter, in Anderson O. D., Ed, *Time Series Analysis: Theory and Practice*, Vol. 5. Amsterdam: North-Holland, pp. 91-101.
- Shea, B. L. (1987). Estimation of multivariate time series. *Journal of Time Series Analysis* **8**, 95-109.
- Shea, B. L. (1988). A note on the generation of independent realisations of a vector autoregressive moving average process. *Journal of Time Series Analysis* **9**, 403-410.
- Shea, B. L. (1989). The exact likelihood of a vector autoregressive moving average model. *Journal of the Royal Statistical Society C*, **38**, 161-184.
- Solo, V. (1984). The exact likelihood for a multivariate ARMA model. *Journal of Multivariate Analysis* **15**, 164-173.
- Söderström, T., Ježek, J. and , Kučera V. (1998). An efficient and versatile algorithm for computing the covariance function of an ARMA process. *IEEE Transactions on Signal Processing* **46**, 1591-1600.

- Stock, J. C. and Watson, M. W. (1988). Testing for common trends. *Journal of the American Statistical Association* **83**, 1097-1107.
- The Numerical Algorithms Group (1999). *Fortran 77 Library Mark 19 Manual*. Oxford: NAG Ltd.
- Tiao, G. C., and Box, G. E. P. (1981). Modeling multiple time series with applications. *Journal of the American Statistical Association* **76**, 802-816.
- Tiao, G. C., and Tsay, R. S. (1989). Model specification in multivariate time series. *Journal of the Royal Statistical Society B*, **51**, 157-213.
- Tong, D. Q. (1991). Méthodes d'estimation de paramètres de modèles autorégressifs multivariés". Unpublished Ph.D. Thesis. Université Joseph Fourier, Grenoble.
- Tsay, R. S. (1989a). Identifying multivariate time series models. *Journal of Time Series Analysis* **10**, 357-372.
- Tsay, R. S. (1989b). Parsimonious parametrization of vector autoregressive moving average models. *Journal of Business and Economic Statistics* **7**, 327-341.
- Tsay, R. S. (1991). Two canonical forms for vector ARMA processes. *Statistica Sinica* **1**, 247-269.
- Tunncliffe Wilson, G. (1973). The estimation of parameters in multivariate time series models. *Journal of the Royal Statistical Society B*, **35**, 76-85.
- Tunncliffe Wilson, G. (1979). Some efficient computational procedures for high order ARMA models. *Journal of Statistical Computation and Simulation* **8**, 303-309.
- Tunncliffe Wilson, G. (1993). Developments in multivariate covariance generation and factorization. in Subba Rao, T., Ed., *Developments in Time Series Analysis: in honor of Maurice B. Priestley*. London: Chapman and Hall, pp. 26-36.
- Yap, S. F., and Reinsel, G. C. (1995), "Estimation and testing for unit roots in partially nonstationary vector autoregressive moving average models. *Journal of the American Statistical Association* **90**, 253-267.