THE EXACT QUASI-LIKELIHOOD OF TIME-DEPENDENT ARMA MODELS WITH APPLICATIONS TO SOME NON-LINEAR MODELS (revised expanded version)[1]

Rajae AZRAK and Guy MELARD

Institut de Statistique
Université Libre de Bruxelles
Campus Plaine CP 210, Boulevard du Triomphe
B-1050 BRUXELLES

E-mail: gmelard@ulb.ac.be
Phone: +32-2-6505890
Fax: +32-2-6505899

Abstract. The purpose of the paper is to propose a simple and efficient algorithm to evaluate the exact quasi-likelihood of (possibly marginally heteroscedastic) ARMA models with time-dependent coefficients. The algorithm is based on the Kalman filter and is therefore simpler than a previous algorithm based on a Cholesky factorisation. Computational efficiency is obtained by taking the ARMA structure into account. Empirical evidence is given. It is also shown how the algorithm can be used as an approximation in the following non-linear models: conditionally heteroscedastic ARMA models (with GARCH errors) and threshold ARMA models, in order to improve the treatment of the initial observations when the parameters of these models are estimated.

1 Introduction

The computation of the exact likelihood function of a Gaussian ARMA (autoregressive-moving average) process of order (p, q) has been studied by several authors, for example Ljung and Box (1979), Ansley (1979), Gardner, Harvey and Phillips (1980), Pearlman (1980), Mélard (1984). These methods consist essentially in expressing the exact marginal distribution of the first m observations (m = max(p, q), max(p, q + 1) or p + q according the method). That procedure is possible because the distribution is multinormal and is thus characterised by the covariances between the first observations. Among those approaches, that of Gardner *et al.* (1980) (denoted by GHP in the sequel) is based on the Kalman filter but requires a large number of operations, mainly for computing the covariance matrix P_0 of the state vector at the initial time t = 0.

Monte Carlo experiments (e.g. Ansley and Newbold, 1980) have shown that, for ARMA models and relatively short series of length 50 or 100, exact maximum likelihood estimation is far superior to conditional maximum likelihood estimation (or least-squares estimation), which either assumes that the pre-sample observations and errors have known fixed values, or starts estimation at $t = m + 1$, with $m = \max(p, q)$ and $e_t = 0$ for $t < m$. It seems therefore plausible that conditional maximum likelihood suffers similarly when dealing with more general time-dependent or non-linear models.

[1] A preliminary version of this paper was presented at the "Rencontres Franco-Belges", Brussels, November 23-24, 1995. A reduced version of this paper, without the applications to non-linear models, is submitted. We thank the three anonymous referees and the editors of the special issue for their useful comments.

The exact likelihood function of a Gaussian *evolutive* ARMA process is the subject of this paper. An evolutive ARMA process is characterised by time-dependent coefficients and marginal heteroscedasticity, i.e. a time-dependent innovation standard deviation. By time-dependent, we mean a deterministic function of time depending on a finite number of parameters, as opposed to random coefficients (e.g. Nicholls and Quinn, 1982) and time-dependent estimators of constant parameters (e.g. Ljung and Söderström, 1983).

The interest towards that kind of models is renewed thanks to new asymptotic results about consistency and normality which have been obtained recently. Also, Grillenzoni (1990) and Dahlhaus (1996) have presented non-standard, approximate and somewhat complex methods and have applied them to the estimation of the parameters of models with time-dependent coefficients for real and artificial time series. Here, we consider the quasi-likelihood estimation method, compute the exact likelihood, not an approximation, and using a renowned algorithm. Implementation of that algorithm is however not trivial because of the determination of the initial conditions, the requirement of computational efficiency (since the algorithm is used many times in an optimisation procedure, and another algorithm already exists), and the generality of the model. Moreover, there are potential applications for the estimation of non-linear models. We do not raise here identification problems which can be serious. Coefficients can be polynomial (Grillenzoni, 1990), periodical (Dahlhaus, 1996) functions of time but other parametrisations can be considered.

The method for computing the likelihood which is described here is based on the Kalman filter. It is a generalisation of GHP with an improvement for the computation of P_0. The proofs are collected in Appendix 1. The main details of the new algorithm are given in Appendix 2. The method will be compared with another exact method given by Mélard (1982) which extends the Cholesky factorisation of Ansley (1979). That method is sketched in Appendix 3.

In several fields of applications, such as finance, meteorology, hydrology and biology, the class of ARMA models has shortcomings. A more satisfactory way to represent these series is to resort to non-linear models, using stochastic processes defined by non-linear equations. There are several ways to introduce non-linearity. We consider two of them: models with generalised conditionally autoregressive heteroscedastic (GARCH) errors (Bollerslev, 1986), and threshold ARMA models (Tong, 1990). The algorithm for computing the exact likelihood function for evolutive ARMA processes, which is described in this article, can be used as an approximation for estimating the parameters of these models.

Contrarily to the evolutive ARMA model which is *marginally* heteroscedastic, the ARMA model with GARCH errors, or ARMA-GARCH, is *conditionally* heteroscedastic. We propose a new estimation procedure for ARMA-GARCH models characterised by better initial values for the recurrences used in the quasi-maximum likelihood method. The procedure is an improvement with respect to Azrak *et al.* (1993).

We also consider threshold ARMA models, which provide a generalisation of threshold autoregressive or TAR models (Tong, 1978, 1983). These models are defined by several ARMA submodels. The submodel used at time t depends on the regime, defined by the location of a variable with respect to one or several thresholds. In the estimation method proposed by Mélard and Roy (1988), the first observations are lost since they are used as initial values in the recurrences. We propose a procedure which has not that inconvenience.

In the conclusions, we briefly indicate how the method can be extended to a wider class of models.

2 Evolutive ARMA models and non-linear models

The evolutive ARMA model considered here is a special case of the extended ARIMA model defined by Azrak and Mélard (1993). A non-stationary white noise process $(e_t, t \in Z)$ is a sequence of independent random variables, with mean zero and variance σ_t^2. We let $\sigma_t = g_t\sigma$, where σ is a scale factor, and g_t is a deterministic functions of t. Then, an evolutive ARMA process is defined as a (generally non-stationary) stochastic process which is the solution of the equation

$$w_t - \sum_{i=1}^{p} \phi_{ti} w_{t-i} = e_t - \sum_{j=1}^{q} \theta_{tj} e_{t-j} \quad , \tag{2.1}$$

where the coefficients $\phi_{t1}, \ldots, \phi_{tp}, \theta_{t1}, \ldots, \theta_{tq}$ and g_t are deterministic functions of t, depending on a finite number of unknown parameters, and p and q are integer constants. Let us denote by ν the vector of parameters, σ being not included.

An ARMA process, also called here an ARMA process with constant coefficients, is a special case defined by the equation

$$w_t - \sum_{i=1}^{p} \phi_i w_{t-i} = e_t - \sum_{j=1}^{q} \theta_j e_{t-j} \quad , \tag{2.2}$$

where the innovations have mean zero and a constant variance σ^2.

The statistical properties of the models with time-dependent coefficients are now better known. Kwoun and Yajima (1986), Hamdoune (1995), and Dahlhaus (1996) have shown estimators which are consistent and asymptotically normal in the homoscedatic case, that is to say when $\sigma_t = \sigma$. Heteroscedastic evolutive models have been scarcely treated in the literature.

A GARCH(q_1, p_1) process is defined (Bollerslev, 1986) as a sequence of martingale differences $(e_t, t \in Z)$, such that $E(e_t/I_{t-1}) = 0$, where I_t denotes the σ-field generated by the process up to time t, $(e_s, s \le t)$ and where the conditional variance satisfies

$$\text{var}(e_t/I_{t-1}) = h_t = \alpha_0 + \sum_{j=1}^{p_1} \alpha_j e_{t-j}^2 + \sum_{i=1}^{q_1} \beta_i h_{t-i} \quad , \tag{2.3}$$

with the constraints

$$\alpha_0 > 0, \quad \alpha_{p_1} > 0, \quad \alpha_j \ge 0, \quad (j = 1, \ldots, p_1 - 1), \beta_{q_1} > 0, \quad \beta_i \ge 0, \quad (i = 1, \ldots, q_1 - 1).$$

The ARMA (p, q) models with GARCH (q_1, p_1) errors have the same form as an ARMA model, except that the errors don't constitute a white noise but a GARCH process, satisfying (2.3).

A threshold autoregressive model for a time series $(w_t, t \in Z)$ is composed of l regimes $R_k (k = 1, \ldots, l)$, autoregressive of order p_k, determined by another series $(y_t, t \in Z)$. Let $I(t)$ be such that $I(t) = k$ if and only if $r_{k-1} \le y_t < r_k$, where the thresholds r_k are real numbers such that $r_0 = -\infty < r_1 < \ldots < r_{l-1} < r_l = \infty$. The model is defined (Tong, 1983) by

$$w_t = \theta_0(I(t)) + \sum_{i=1}^{p_{I(t)}} \phi_i(I(t)) w_{t-i} + e_t, \tag{2.4}$$

where $(e_t, t \in Z)$ is a non-stationary white noise process with variance $\text{var}(e_t) = \sigma_{I(t)}^2$. The coefficients $\phi_i(k), i = 1, \ldots, p_k, \theta_0(k)$ and the standard deviations σ_k, are the parameters of the k-th regime, k = 1, ..., l. Tong's approach has some inconveniences, particularly the purely autoregressive specification and the fact that the parameters of the different regimes need to be functionally independent from each other, leading to a large number of parameters and complicating estimation.

These considerations and others have lead Mélard and Roy (1988) to propose a quasi-maximum likelihood estimation method for the threshold ARMA (TARMA) models (e. g. Tong, 1990, p. 101). If we forget the constants, TARMA model is defined similarly as (2.4) as follows

$$w_t = \sum_{i=1}^{p_{I(t)}} \phi_i(I(t)) w_{t-i} + e_t - \sum_{j=1}^{q_{I(t)}} \theta_j(I(t)) e_{t-j}. \tag{2.5}$$

3 Quasi-likelihood function of evolutive ARMA models

We suppose that the process is stationary and invertible in the past. This means that, for all $t \leq 0$, $\phi_{ti} = \phi_i$, $\theta_{tj} = \theta_j$, and $g_t = g_0$, and the zeros of the polynomials in the complex variable z

$$1 - \sum_{i=1}^{p} \phi_i z^i \quad , \quad 1 - \sum_{j=1}^{q} \theta_j z^j$$

are all of modulus greater than 1. The assumption of invertibility in the past implies that the non-stationary process is invertible (Cramér, 1961, Hallin, 1978), i. e. for all finite t, e_t can be expressed as a mean-square convergent linear combination of w_t, w_{t-1}, w_{t-1}, …. Without that property, the process remains unindentifiable and inefficient forecasts are produced by the model (Hallin, 1986).

The quasi-likelihood function is the density function of the observations $w = (w_1 \ldots w_n)'$, considered as a function of the parameters of the model, assuming that the process is Gaussian. It can be written as follows

$$L(\nu, \sigma^2; w) = (2\pi)^{-n/2} (\det \Gamma_w)^{-1/2} \exp\left(-\frac{1}{2} w' \, \Gamma_w^{-1} \, w\right) \quad , \tag{3.1}$$

where Γ_w is the $n \times n$ covariance matrix of w, whose element (t, s) is $\text{cov}(w_t, w_s)$ $(t, s = 1, \ldots, n)$. The computation of that function requires inverting Γ_w, which requires a computation time proportional to n^3.

We shall develop a very fast algorithm with a number of operations of order n instead of n^3, allowing therefore to evaluate the likelihood of an evolutive ARMA process exactly and without difficulty. Another algorithm based on the Cholesky factorisation of a band matrix (Mélard, 1982) is summarised in Appendix 2.

4 Fast algorithm based on the Kalman filter

In an article devoted to the likelihood of multivariate ARMA models in the context of missing and/or aggregated data, Ansley and Kohn (1983) have used the Kalman filter for a state space model where the matrices depend on time. They haven't handled explicitly ARMA models with time-dependent coefficients. Mélard (1985) has given a sketch of an algorithm for evaluating the likelihood of a multivariate ARMA model with time-dependent coefficients. It is an improved version of that algorithm which is offered here in the univariate case.

Let us consider a discrete time stochastic process $(W_t, t \in N)$ satisfying the equation:

$$W_t = F_t W_{t-1} + G_t e_t \quad , \tag{4.1}$$

where the $r \times 1$ state vector W_t represents the state of the system at time t, F_t is called the $r \times r$ transition matrix, G_t is a $r \times 1$ matrix, and $(e_t, t \in N)$ is a white noise process, with zero mean and variance σ_t^2. Suppose that the state of the system can be observed through the observation w_t, of dimension 1:

$$w_t = H_t W_t, \tag{4.2}$$

where H_t is a $1 \times r$ matrix (usually a noise is added to the right hand side of (4.2) but it is not necessary here). The specification of the state space form (4.1-2) is completed by the two conditions:

- the initial state vector, W_0, with mean α_0 and covariance matrix P_0;
- that the e_t are uncorrelated with the initial state vector W_0, i. e. $E(e_t W_0') = 0$ for $t \in N$.

The Kalman filter is a recurrence procedure for obtaining an estimator of the state vector at time t, based on the information at time $t-1$. We follow the presentation of Harvey (1989, pp. 104-107), slightly improved. Let us consider the model composed of (4.1) and (4.2). Denote I_t the σ-field generated by the random variables $(w_u, u \le t)$, $\alpha_{t-1} = E(W_{t-1}/I_{t-1})$ the optimal estimator of W_{t-1}, based on the observations up to time $t-1$, and $\alpha_{t/t-1}$ the estimator of W_t based on the same observations:

$$\alpha_{t/t-1} = E(W_t/I_{t-1}) = F_t \alpha_{t-1} \quad , \tag{4.3}$$

taking (4.1) into account. Let P_t be the covariance matrix of the estimation error defined by:

$$P_t = E\{(W_t - \alpha_t)(W_t - \alpha_t)'\} \quad ,$$

and $P_{t/t-1}$ the prediction error covariance matrix

$$P_{t/t-1} = E\{(W_t - \alpha_{t/t-1})(W_t - \alpha_{t/t-1})'\}$$

$$= F_t P_{t-1} F_t' + G_t \sigma_t^2 G_t' \quad . \tag{4.4}$$

Equations (4.3-4) are known as the prediction equations from which the updating equations can be deduced (Jazwinski, 1970):

$$\alpha_t = \alpha_{t/t-1} + P_{t/t-1} H_t' \hat{\sigma}_t^{-2} \hat{e}_t \quad , \tag{4.5}$$

$$P_t = P_{t/t-1} - P_{t/t-1} H_t' \hat{\sigma}_t^{-2} H_t P_{t/t-1} \quad , \tag{4.6}$$

where

$$\hat{\sigma}_t^2 = H_t P_{t/t-1} H_t' \tag{4.7}$$

and the sample innovations, or residuals, are defined by

$$\hat{e}_t = w_t - H \alpha_{t/t-1} \quad . \tag{4.8}$$

Let us denote by K_t the gain matrix given by

$$K_t = F_{t+1} P_{t/t-1} H_t' \hat{\sigma}_t^{-2} \quad . \tag{4.9}$$

The Kalman recurrence estimation technique, or Kalman filter, is based on equations (4.3-8). They allow us to write the following recurrence relations:

$$\alpha_{t+1/t} = (F_{t+1} - K_t H_t) \alpha_{t/t-1} + K_t w_t \quad , \tag{4.10}$$

$$P_{t+1/t} = F_{t+1} P_{t/t-1} F'_{t+1} + G_{t+1} \sigma_{t+1}^2 G_{t+1}' - K_t \hat{\sigma}_t^2 K'_t \quad . \tag{4.11}$$

In principle, the initial values of the Kalman filter are given by the mean and the covariance matrix of the distribution of the state vector. They can be specified by $\alpha_0 = 0$ and P_0, or $\alpha_{1/0} = 0$ and $P_{1/0} = P_0$. To use that recurrence algorithm in order to evaluate the likelihood function of the evolutive ARMA model defined by (2.1), it is necessary to transform (2.1) in a state space representation. That transformation can be done using the following theorem (Mélard, 1985).

Theorem 1

Let $(w_t, t \in \mathrm{N})$ be an evolutive ARMA(p, q) process defined by (2.1) and such that:

$$\phi_{ti} = 0 \quad , \quad p < i < r \quad , \qquad \theta_{tj} = 0 \quad , \quad q < j < r-1 \quad ,$$

with $r = \max(p, q+1)$. It can be represented (not uniquely) in state space form (4.1-2), where

$$F_t = \begin{pmatrix} \phi_{t,1} & | & & & & \\ \phi_{t+1,2} & | & & & & \\ . & | & & I_{r-1} & & \\ . & | & & & & \\ . & | & & & & \\ -- & -- & -- & -- & -- & -- \\ \phi_{t+r-1,r} & | & 0 & \ldots & \ldots & 0 \end{pmatrix} \quad , \quad G_t = \begin{pmatrix} 1 \\ -\theta_{t+1,1} \\ -\theta_{t+2,2} \\ . \\ . \\ . \\ -\theta_{t+r-1,r-1} \end{pmatrix} \quad , \qquad (4.12)$$

$H_t = H$ is a constant matrix equal to $H = (1, 0, \ldots, 0)$, and $\sigma_t = g_t\sigma$.

By applying the projection theorem (Jazwinski, 1970), the residuals $\hat{e}_t$ are mutually uncorrelated random variables, with mean zero and variance (4.7) which defines $b_t^2 = \hat{\sigma}_t^2/\sigma^2$. If we suppose that the process is Gaussian, the residuals have a normal distribution and are mutually independent. Changing the variable w_t by $\hat{e}_t$ in the density is a transformation which yields a lower triangular Jacobian matrix containing ones on the main diagonal. Consequently, the Jacobian is equal to 1. The logarithm of the likelihood function deduced from (3.1) can thus be written as follows:

$$\log L(\nu, \sigma^2; w) = c - \frac{1}{2}\sum_{t=1}^{n} \log(\sigma^2 b_t^2) - \frac{1}{2}\sum_{t=1}^{n} \frac{\hat{e}_t^2}{\sigma^2 b_t^2} \quad , \qquad (4.13)$$

where c is a constant.

It is well known (for example Tunnicliffe Wilson, 1973) that estimating the variance simultaneously with the other parameters can be difficult. We consider an alternative procedure which makes use of a non-linear least squares procedure such as the one proposed by Marquardt (1963). Solving the likelihood equation for σ^2, and substituting σ^2 by the solution

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n} \frac{\hat{e}_t^2}{b_t^2} \quad ,$$

we obtain the concentrated log-likelihood

$$c' - \frac{1}{2}\sum_{t=1}^{n} \log(b_t^2) - \frac{n}{2}\log\left\{\sum_{t=1}^{n} \frac{\hat{e}_t^2}{b_t^2}\right\} \quad , \qquad (4.14)$$

where c’ is another constant. From that expression, parameter estimation is done as indicated by Azrak and Mélard (1993).

In order to compute $\hat{e}_t^2$ and b_t^2 in (4.14), the recurrences (4.7-11) can be used, in the given order, avoiding the estimation error covariance matrix P_t, see e. g. Pearlman (1980), Mélard (1985). In the sequel, we propose an algorithm employing simultaneously $P_{t/t-1}$ and P_t, which is justified by the following theorem (inspired by the Fortran code in GHP in the stationary case).

Theorem 2

Let $(w_t, t \in \mathrm{N})$ be an evolutive ARMA(p, q) process expressed in the form of the linear dynamic model of Theorem 1. The elements of the matrix $V_t = G_t\sigma_t^2 G_t'$ have the following form

$$V_t(i,j) = V_t(j,i) = \begin{cases} \theta_{t+i-1,i-1}\theta_{t+j-1,j-1}\sigma_t^2, & 2 \le i \le j \le r, \\ -\theta_{t+j-1,j-1}\sigma_t^2, & i = 1, \quad 1 < j \le r, \\ \sigma_t^2, & i = j = 1. \end{cases} \qquad (4.15)$$

The elements of the matrix $P_{t/t-1}$ are given by the recurrence:

$$P_{t/t-1}(i,j) = V_t(i,j) + P_{t-1}(i+1,j+1), \qquad i,j = 1,\dots,r-1 \quad , \tag{4.16a}$$

$$P_{t/t-1}(r,r) = V_t(r,r) \quad , \tag{4.16b}$$

and the elements of matrix P_t are given by the following recurrence:

$$P_t(i,j) = P_{t/t-1}(i,j) - \frac{P_{t/t-1}(i,1)P_{t/t-1}(1,j)}{P_{t/t-1}(1,1)} \qquad i,j \geq 2 \quad , \tag{4.17a}$$

$$P_t(1,j) = P_t(i,1) = 0. \tag{4.17b}$$

In order to start the recurrences, we need initial values which are respectively $\alpha_0 = 0$ and $P_0 = E[W_0 W'_0]$. If the process were stationary, hence $F = F_t$, $G = G_t$, $\sigma_t^2 = \sigma^2$, $g_t = 1$ $(t = 1,\dots,n)$, we could determine P_0 as the solution of the equation

$$P_0 = FP_0F' + G\sigma^2G' = C + V \quad , \tag{4.18}$$

where $C = FP_0F'$ and $V = G\sigma^2G'$. In the subroutine STARMA, GHP have rewritten (4.18) in the form $\mathrm{Vec}(V) = S\,\mathrm{Vec}(P_0)$, with an appropriate definition of Vec(.) and some matrix S, in such a way to solve a system of $r(r+1)/2$ linear equations. It is one of the reasons why their algorithm is not efficient for large models, the other one being the use of the Kalman filter instead of the Chandrasekhar recurrences (Morf *et al*., 1974, Pearlman, 1980, Mélard, 1984).

The algorithm proposed in this paper for computing the initial covariance matrix P_0 of the state vector (Akaike, 1978, Jones, 1980, Mélard, 1984) allows to improve the AS154 program for computing the likelihood function of an ARMA model with constant coefficients using the GHP method. The autocovariances $\gamma_k = \mathrm{cov}(w_t, w_{t-k})$ for $k = 0,\dots,r$, are determined by the algorithm of Tunnicliffe Wilson (1979). A subproduct of that algorithm (see Mélard, 1984) consists in the covariances $\lambda_k = E(w_t e_{t-k})$ computed by the recurrence

$$\lambda_k = -\theta_k\sigma^2 + \sum_{j=1}^{\min(p,k)} \phi_j \lambda_{k-j} \quad , k = 1,\dots,q, \tag{4.19}$$

with $\lambda_0 = \sigma^2$. The first element of matrix P_0 is given by

$$P_0(1,1) = \mathrm{cov}((W_0)_1, (W_0)_1) = \mathrm{var}(w_0) = \gamma_0. \tag{4.20}$$

In the case of an ARMA(p, q) process with constants coefficients, (A1.1) (see Appendix 1) is written in the form

$$(W_t)_i = \sum_{j=i}^{r} (\phi_j w_{t+i-1-j} - \theta_{j-1} e_{t+i-j}) \quad .$$

Hence, the components $P_0(i,1)$ of P_0 can be obtained by the following relation:

$$P_0(i,1) = \mathrm{cov}((W_0)_i, (W_0)_1) = E((W_0)_i w_0)$$

$$= \sum_{j=i}^{r} \{\phi_j E(w_{-j+i-1} w_0) - \theta_{j-1} E(e_{-j+i} e_0)\}$$

$$= \sum_{j=i}^{r} (\phi_j \gamma_{j-i+1} - \theta_{j-1} \lambda_{j-i}) \quad . \tag{4.21}$$

Since P_0 is a symmetric matrix, there remains to determine the diagonal and subdiagonal elements of the columns after the first one, which can be done using the following theorem.

Theorem 3

Let $(w_t, t \in N)$ be an ARMA(p, q) process with constant coefficients satisfying (2.1), represented in the form of a linear dynamic model (4.1) with matrices $F_t = F, G_t = G$ composed of the coefficients ϕ_i, θ_i, respectively. The elements $P_0(i, j)$ of the covariance matrix of the state vector at time 0 satisfy the relation:

$$P_0(i+1, j+1) = P_0(i, j) - \phi_j\{\phi_i P_0(1, 1) + P_0(i+1, 1)\} - \phi_i P_0(1, j+1) - V(i, j) \quad , \quad (4.22)$$

for $1 \le i, j \le r - 1$, where the elements of $V = \sigma^2 G G'$ are given by:

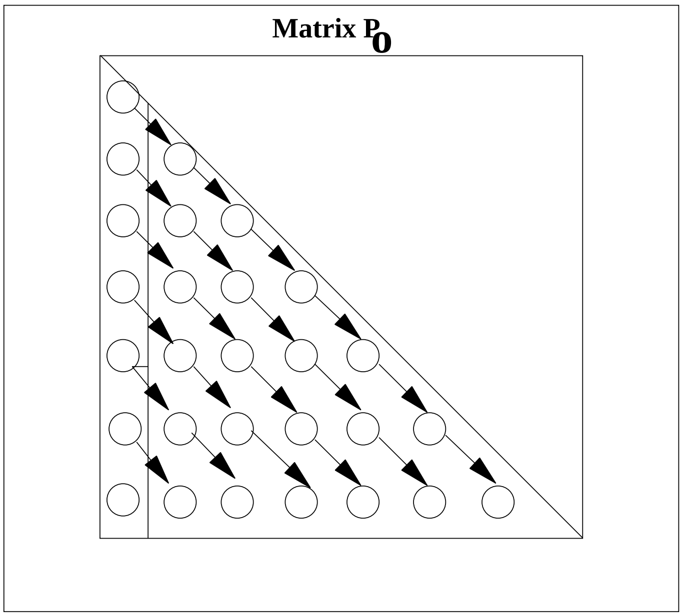
$$V(i, j) = V(j, i) = \theta_{i-1}\theta_{j-1}\sigma^2 \quad 2 \le i \le r, \quad i \le j \le r \quad ,$$

$$V(1, j) = -\theta_{j-1}\sigma^2 \quad 1 < j \le r \quad ,$$

$$V(1, 1) = \sigma^2 \quad .$$

Remark. The recurrence relation expressed in Theorem 3 allows to determine the elements within each diagonal, in the manner represented by Figure 1.

Figure 1.



Scheme of the recurrence

Let us now consider the evolutive ARMA process defined by (2.1). Suppose that it is stationary in the past. More precisely, suppose that $F_t = F_1$, $G_t = G_1$, and $\sigma_t = g_1\sigma$ for $t \le 1$, which implies that:

$$\phi_{tj} = \phi_{jj} \quad , \theta_{tj} = \theta_{jj} \quad , \quad \text{for} \quad t < j \quad . \quad (4.23)$$

Equation (4.18) is replaced by

$$P_0 = F_1 P_0 F_1' + G_1 g_1^2 \sigma^2 G_1' = C + V \quad . \quad (4.24)$$

The method described in Theorem 3 can then be used, with some modifications resulting from the replacement of σ by $g_1\sigma$.

In all the relations which have been given, σ is involved but it is not known until the other parameters have been estimated. Consequently, σ should not appear which implies e.g. that $P_{t/t-1}/\sigma^2$ is expressed instead of $P_{t/t-1}$in (4.4), and similarly for the other equations.

The final algorithm, without the initialisation stage which has just been described, is given in Appendix 2. At time t, we use the notation α for both $\alpha_{t/t-1}$ and α_t, P for both $P_{t/t-1}/\sigma^2$ and P_t/σ^2, and V for V_t/σ^2. The order of computation is $\alpha_{t/t-1}$ (4.3), $P_{t/t-1}/\sigma^2$ (4.16), b_t^2 (4.7), $\hat{e}_t^2$ (4.8), α_{t+1} (4.5), P_{t+1}/σ^2 (4.17), V_{t+1}/σ^2 (4.15).

The AS154 program of GHP for computing the likelihood function of ARMA models with constant coefficients is composed of two subroutines, STARMA which computes P_0/σ^2, and KARMA, which evaluates the likelihood function and the covariance matrix P_t. STARMA requires a large number of operations, and consequently a large execution time. We have replaced it by an improved subroutine called RECPO which implements the preceding method. Besides, we have adapted KARMA renamed TKALMAR so as to handle evolutive models. Note finally that the diffuse Kalman filter technique of De Jong and Chu-Chun-Lin (1994) is not necessary in our context.

5 Empirical evidence

In this Section, a comparison of the new algorithm with the one based on the Cholesky factorisation is made and a limited Monte Carlo study is presented to compare the unconditional quasi-maximum likelihood method with the conditional quasi-maximum likelihood method.

In order to compare the methods for evaluating the likelihood function, it is necessary to compute the number of operations, counting multiplications and divisions. The best method will be the one with the smallest number of operations, which can be decomposed as follows:

$$N_n(p,q) = N_0(p,q) + n N_t(p,q),$$

where $N_0(p,q)$ indicates the number of operations at the start (at time $t \le 0$), and $N_t(p,q)$ means the number of operations at time t.

In the case of evolutive models, there is another available algorithm LIKAMT, due to Mélard (1982), which is based on a Cholesky factorisation of a band matrix (LIKAMT) (see Appendix 3). It is compared here in Table 1 with the algorithm proposed in Section 4. Only the dominating terms with respect to n, p and q are shown. Moreover, p and q are neglected before n, and 1 is neglected before p or q. Let s be defined by $s = \min(p,q)$.

Table 1. Number of operations used by algorithms LIKAMT and TKALMAR

Algorithms	references	number of multiplications and divisions
LIKAMT	Mélard (1982)	$\frac{7}{6}p^3 + p^2(q-p) + \frac{p}{2}(r^2 - p^2) + n(q^2 + p)$
Proposed method TKALMAR	Azrak (1996)	$\left(\frac{p^2}{2} + \frac{q^2}{2} + 2pr + qs + \frac{3}{2}r^2\right) + n\left(\frac{r^2}{2} + \frac{q^2}{2} + \frac{3}{2}r + \frac{3}{2}q + p\right)$

Note that without Theorem 2, the number of operations of TKALMAR needed at each time t would be higher: of order $3r^3/2$, if no care is taken of the structural zeros in F_t, or $2pr + r^2/2 + q^2/2$, otherwise, instead of $r^2/2 + q^2/2$.

We have included LIKAMT in the ANSECH-PC module of the software called Time Series Expert (Mélard and Pasteels, 1994). We have compared it with TKALMAR based on the modification of AS154 (GHP) (see Section 4). As a matter of illustration, execution time in milliseconds are reproduced in Tables 2 to 5. They have been obtained as averages over 1000 trials on a personal computer with an Intel Pentium processor clocked at 60 MHz, while evaluating the exact likelihood function of ARMA(p, q) models with time-dependent coefficients for series of length 50, 100 and 200.

Table 2. Large ARMA(p, q) models for a series of length 50: execution time (ms).

Models	(13,0)	(12,1)	(1,12)	(0,13)	(13,13)	(24,13)	(13,24)	(24,24)
LIKAMT	13.5	12.2	15.0	15.8	33.1	90.3	66.9	129.7
TKALMAR	18.0	16.1	18.0	20.0	21.8	52.3	55.9	57.9

Table 3. Large ARMA(p, q) models for a series of length 100: execution time (ms).

Models	(13,0)	(1,12)	(12,1)	(0.13)	(13,13)	(24,13)	(13,24)	(24,24)
LIKAMT	14.8	29.1	14.7	31.5	50.1	108.3	110.5	174.4
TKALMAR	34.7	35.0	30.8	38.9	41.6	100.1	107.4	110.1

Table 4. Large ARMA(p, q) models for a series of length 200: execution time (ms).

Models	(13,0)	(12,1)	(1,12)	(0.13)	(13,13)	(24,13)	(13,24)	(24,24)
LIKAMT	17.6	19.6	57.4	63.0	84.0	144.2	197.8	263.8
TKALMAR	67.8	60.4	68.9	76.7	81.3	195.6	210.4	214.8

Table 5. Ratio of execution times for LIKAMT over TKALMAR (according to the series length, 50, 100, 200)

Length	(13,0)	(12,1)	(1,12)	(0,13)	(13,13)	(24,13)	(13,24)	(24,24)
50	0.9	0.7	0.8	0.8	1.5	1.7	1.2	2.2
100	0.4	0.5	0.8	0.8	1.2	1.1	1.0	1.6
200	0.3	0.3	0.8	0.8	1.0	0.7	0.9	1.2

We notice that algorithm LIKAMT is the fastest for some models. On the contrary, for models with orders p and q equal or larger than 13, we observe a superiority of TKALMAR, at least for short series. Even if is not as fast for some models, it is based on a simple principle and is easy to manipulate.

One of the reasons why algorithm LIKAMT is fast, despite the fact it is complex, is that it is optimised as a function of p and of q, separately, whereas TKALMAR depends on the size r of the state vector. TKALMAR can still be improved, mainly for long series, by making use of the fast recurrences (*quick recursions*) implemented in GHP. We have not shown results in that direction because they depend on the specific models being used.

Now, we consider a limited Monte Carlo study to show that the unconditional quasi-maximum likelihood method is superior to the conditional quasi-maximum likelihood method. We are interested in speed of convergence of the estimators to the true value of the parameters, either when the innovation distribution is normal (corresponding to exact maximum likelihood), or it is not normal, considering as an example the case where the law is double exponential.

A marginally heteroscedastic MA(1) model, defined by

$$w_t = e_t - \theta e_{t-1},$$

$$\mathrm{var}(e_t) = \sigma_t^2 = \exp(2\gamma t),$$

has been simulated, with $\theta = 0.9$, and using a normal or double exponential distribution for the e_t. The length of the series varies from 25 to 400. The value of γ has been chosen in correspondence to n, such that the product γn is roughly constant. A number of 10000 series have been generated. The results are given in Table 6 and 7 in the case of a normal distribution, and in Table 8 for the comparison between double exponential and normal distributions. The bias is smaller with the unconditional method than with the conditional method, and decreases faster to zero when n increases. Also, there is no difference when the innovation distribution is compatible with the law used in the quasi-maximum likelihood, i.e. normal, or when it is not compatible, i.e. double exponential.

Table 6. Estimated parameters for order one moving average model with normal innovation distribution obtained by the conditional or unconditional quasi-maximum likelihood methods (n = 25 or 50; 10000 replications)

Observations	$n = 25$		$n = 50$	
Method	$\theta = 0.900$	$\gamma = 0.054$	$\theta = 0.900$	$\gamma = 0.0027$
Conditional	0.851	0.048	0.871	0.0025
Unconditional	0.880	0.053	0.905	0.0027

Table 7. Estimated parameters for order one moving average model with normal innovation distribution obtained by the conditional or unconditional quasi-maximum likelihood methods (n = 100, 200 or 400; 10000 replications)

Observations	$n = 100$		$n = 200$		$n = 400$	
Methods	$\theta = 0.900$	$\gamma = 0.0130$	$\theta = 0.900$	$\gamma = 0.0060$	$\theta = 0.900$	$\gamma = 0.0030$
Conditional	0.883	0.0120	0.889	0.0060	0.893	0.0029
Unconditional	0.910	0.0130	0.905	0.0060	0.902	0.0030

Table 8. Estimated parameters for order one moving average model with normal or double exponential innovation distribution obtained by the conditional or unconditional quasi-maximum likelihood methods (n = 50; 10000 replications)

Innovation distribution	double exponential		normal	
Method	$\theta = 0.900$	$\gamma = 0.027$	$\theta = 0.900$	$\gamma = 0.027$
Conditional	0.872	0.025	0.871	0.025
Unconditional	0.905	0.027	0.905	0.027

6 Application to the ARMA model with GARCH errors

The joint density function of the observations $w = (w_1, \ldots, w_n)$ is expressed in the form of the product of the marginal density of the first observation by the conditional densities of the following observations. The main contribution of this Section is to show a natural approximation of the likelihood function which allows using the algorithm for evaluating the likelihood function of evolutive processes such as described in Section 4.

Parameter estimation is achieved by the quasi-maximum likelihood method, where the likelihood function is computed as if the conditional distributions of the $(e_t, t \in Z)$ were normal. The ARMA-GARCH model can then be rewritten as

$$w_t - \sum_{i=1}^{p} \phi_i w_{t-i} = e_t - \sum_{j=1}^{q} \theta_j e_{t-j} \quad , \tag{6.1}$$

$$e_t / I_{t-1} \sim N\left(0, h_t = \alpha_0 + \sum_{j=1}^{p_1} \alpha_j e_{t-j}^2 + \sum_{i=1}^{q_1} \beta_i h_{t-i}\right) \quad . \tag{6.2}$$

We temporarily suppose that all the variables for $t < 1$ have known fixed values.

The parameters $\phi_i (i = 1, \ldots, p)$, $\theta_j (j = 1, \ldots, q)$, $\alpha_j (j = 1, \ldots, p_1)$, and $\beta_i (i = 1, \ldots, q_1)$ are stored in vector ν. Noted that α_0 is not included in ν, and $\alpha_0^{1/2}$ can be interpreted as a scale factor, like σ in Section 2.

We should therefore exhibit a parametrisation which doesn't involve $\alpha_0{}^2$. Let us denote $g_t^2 = h_t / \alpha_0$ and the new parameters $\alpha'_j = \alpha_j / \alpha_0$. Equation (2.3) implies the following expression for the conditional variance

$$g_t^2 = 1 + \sum_{j=1}^{p_1} \alpha_j' e_{t-j}^2 + \sum_{i=1}^{q_1} \beta_i g_{t-i}^2 \quad . \tag{6.3}$$

Let $\nu_1 = (\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q, \alpha_1', \ldots, \alpha_{p_1}', \beta_1, \ldots, \beta_{q_1})$ be the new vector of parameters. The logarithm of the quasi-likelihood function can thus be written as

$$\log l(\nu_1, \alpha_0; w) = c - \frac{1}{2} \sum_{t=1}^{n} \log \alpha_0 g_t^2 - \frac{1}{2} \sum_{t=1}^{n} \frac{e_t^2}{\alpha_0 g_t^2} \quad . \tag{6.4}$$

As in Section 4, we obtain the concentrated log-likelihood with respect to α_0:

$$c' - \frac{1}{2} \sum_{t=1}^{n} \log g_t^2 - \frac{n}{2} \log \left\{ \sum_{t=1}^{n} \frac{e_t^2}{g_t^2} \right\} \quad , \tag{6.5}$$

where the e_t and g_t are computed by using respectively (6.1) and (6.3). After having obtained the estimators $\hat{\nu}_1$ and $\hat{\alpha}_0$, we derive the estimators of the original parameters $\alpha_j = \alpha_0 \alpha'_j$.

In the preceding paragraphs, we have supposed that the variable and the errors have known fixed values for $t < 1$. Our purpose here is to describe briefly an estimation method which doesn't make that restrictive assumption.

Let us consider time t. Conditionally to I_t, the model has the form (2.1), where the coefficients $\phi_{ti} = \phi_i, i = 1, \ldots, p$, $\theta_{tj} = \theta_j, j = 1, \ldots, q$ and g_t is given by (6.3). This is a special case of an evolutive ARMA model, called ARMAG model by Mélard (1977, 1985), intrinsically a marginally heteroscedastic ARMA model. Therefore the quasi-likelihood function (4.13) can be computed by using the method described in Section 4, which consists in removing the link between relations (6.1) and (6.3). We consider the errors e_t as if they were the innovations of an ARMAG model where g_t is deterministic instead of being random, and we consider the g_t as if they were obtained from expression (6.3), i. e. using the past errors instead of being deterministic. The method computes the residuals $\hat{e}_t$ and their variance $\sigma^2 b_t^2$ using the improved Kalman filter deduced from Theorem 2, and described in Appendix 2. Theorem 3 can be used to start the recurrences.

That procedure avoids the loss of observations at the beginning of the series. The only assumption which is needed is that $g_t = 1$ for $t \le 1$.

---

[2] That approach has been developed in collaboration with O. Scaillet, see Azrak *et al.* (1993).

7 Application to threshold ARMA models

Mélard and Roy (1988) have considered an estimation method for threshold ARMA models (Tong, 1990, p. 101) by conditional quasi-maximum likelihood: they have supposed normality of the process $(e_t, t \in Z)$ but have assumed that the values of the variable w_t and e_t, $t \le m = \max(p, q)$, are zero, where $p = \max(p_1, \ldots, p_l)$ and $q = \max(q_1, \ldots, q_l)$. In this Section, unconditional quasi-maximum likelihood is considered, under the assumptions that the thresholds are known, that the ARMA models for the l regimes are stationary and invertible, and that the regime for time 1 is also the regime for $t < 1$. This provides an interpretation of the first e_t as innovations of the process.

We can proceed as in Section 6, computing the quasi-likelihood function by the product of normal conditional densities. At time t, conditionally to $y = (y_1, \ldots, y_t)$, the model defined by (2.5) is an evolutive ARMA model, since the coefficients depend on time, through the regime $I(t)$ at time t. Indeed, (2.5) corresponds to (2.1), where the coefficients $\phi_{ti} = \phi_i(I(t)), i = 1, \ldots, p$, $\theta_{tj} = \theta_j(I(t)), j = 1, \ldots, q$ and $g_t\sigma = \sigma_{I(t)}$.

Consequently, we consider the errors e_t as if they were the innovations of an evolutive ARMA model. The quasi-likelihood function is of the form (4.13) and is computed by using the method described in Section 4. More precisely, the residuals $\hat{e}_t$ and their variance $\sigma^2 b_t^2$ are obtained using the recurrences in Appendix 2, and Theorem 3 is used to start the recurrences. Contrarily to Tong (1983) and Mélard and Roy (1988), the impact of the initial conditions is reduced because the likelihood is no longer conditional on the first observations and errors.

In principle, the algorithm LIKAMT based on the Cholesky factorisation (Mélard, 1982) (see Appendix 3) could be used instead but it is very complex and the computation of the coefficients of the model appears at several places. On the contrary, the Kalman filter method improved in Section 4 allows more easily to insert the computation of the coefficients for the different regimes, while reducing the total computation time for high-order models. However, the assumption (4.23) should be slightly modified.

8 Conclusion

In this paper, a new algorithm called TKALMAR has been developed for quasi-maximum likelihood estimation of non-stationary models. That algorithm is based on the Kalman filter with a number of improvements. It is compared with the algorithm LIKAMT (Mélard, 1982) which relies on the Cholesky factorisation of a band matrix. The results for the computation time are not in favour of a single algorithm. TKALMAR is better than LIKAMT for high-order models, when the series length is not too large. TKALMAR, which is conceptually simple, can be used instead of non-standard, approximate and somewhat complex methods proposed and applied recently by Grillenzoni (1990) and Dahlhaus (1996). Besides linear models with time-dependent coefficients, TKALMAR described in this paper can also be used for estimating the parameters in ARMA-GARCH models and in threshold ARMA models.

At the time of writing, the algorithm is not implemented yet in the programme ANSECH of Time Series Expert (TSE version 2.2), see Mélard and Pasteels (1994), but well LIKAMT, based on the Cholesky factorisation. ARMA models with coefficients which depend linearly or exponentially on time, with an innovation standard deviation which varies linearly or exponentially on time, can be fitted by quasi-maximum likelihood. For ARMA models with GARCH errors, three estimation methods are proposed for estimating simultaneously the parameters of an ARMA-GARCH model:

- a conditional maximum likelihood method using initial errors equal to zero in order to start the recurrences ("algorithm with taking off");

- the method described in Section 6, but using the Cholesky factorisation instead of the Kalman filter ("algorithm for time-dependent ARMA models");

• a compromise method, based for the first observations on the exact evaluation of a process with time-dependent coefficients, in order to start the recurrence relations of the conditional method ("algorithm with time-dependent startup").

For threshold ARMA models, the THRESH programme that Mélard and Roy (1988) have implemented and which has been converted to personal computers by Azrak (1996) is still based on conditional quasi-likelihood estimation and relies on LIKAMT.

The approach which has been developed in this paper can be extended without difficulty to the following models:

• the evolutive ARMA(p, q) model with GARCH(q_1, p_1) errors, defined by the following equation:

$$w_t = \sum_{i=1}^{p} \phi_{ti} w_{t-i} + e_t - \sum_{j=1}^{q} \theta_{tj} e_{t-j} \quad ,$$

with

$$E(e_t / I_{t-1}) = 0$$

$$V(e_t / I_{t-1}) = h_t = \alpha_0 + \sum_{j=1}^{p_1} \alpha_j e_{t-j}^2 + \sum_{i=1}^{q_1} \beta_i h_{t-i} \quad .$$

• the threshold evolutive ARMA model, defined by the equation

$$w_t = \sum_{i=1}^{p_{(I(t))}} \phi_{ti}(I(t)) w_{t-i} + e_t - \sum_{j=1}^{q_{(I(t))}} \theta_{tj}(I(t)) e_{t-j} \quad .$$

We have not yet investigated the asymptotic properties of the estimators. Except for the thresholds, they can be derived by combining the approach of Azrak (1996) with that of Klimko and Nelson (1978), and Tjøstheim (1984b, 1986).

References

AKAIKE, H. (1978) Covariance matrix computation of the state variable of a stationary Gaussian process, *Ann. Inst. Statist. Math.* **30 (B)**, 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., and KOHN, R. (1983) Exact likelihood of vector autoregressive-moving average process with missing or aggregated data, *Biometrika* **70**, 275-278.

ANSLEY, C. F., and NEWBOLD, P. (1980) Finite sample properties of estimators for autoregressive moving average models, *J. Econometrics* **13**, 159-183.

AZRAK, R. (1996) Contributions à l'estimation de modèles non stationnaires, thèse de doctorat, Université Libre de Bruxelles, Bruxelles.

AZRAK, R., and MELARD, G. (1993) Exact likelihood estimation for extended ARIMA models, in honour of M. B. Priestley, T. Subba Rao (ed.) *Developments in Time Series Analysis*, Chapman & Hall, London, pp. 110-123.

AZRAK, R., MELARD, G., and SCAILLET, O. (1993) Estimation of models with ARCH errors and applications, Communication at ORBEL 7 (Seventh National Congress on Quantitative Methods for Decision Making), Bruxelles, January 21-22.

BOLLERSLEV, T. (1986) Generalized autoregressive conditional heteroskedasticity, *J. Econometrics* **31**, 307-327.

CRAMER, H. (1961) On some classes of non-stationary stochastic processes, in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley and Los Angeles, Vol. 2, pp. 57-78.

DAHLHAUS, R. (1996) Fitting time series models to nonstationary processes, *Ann. Statist.*, forthcoming.

DE JONG, P., and CHU-CHUN-LIN, S. (1994) Fast likelihood evaluation and prediction for nonstationary state space models, *Biometrika* **81**, 133-142.

GARDNER, G. , HARVEY, A. C., and PHILLIPS, G. D. A. (1980) Algorithm AS 154, An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering, *J. Roy. Statist. Soc. Ser. C, Appl. Statist.* **29,** 311-322.

GRILLENZONI, C. (1990) Modeling time-varying dynamical systems, *J. Amer. Statist. Assoc.* **85**, 499-507.

HALLIN, M. (1978) Mixed autoregressive-moving average multivariate processes with time-dependent coefficients, *J. Multivariate Anal.* **8**, 567-572.

HALLIN, M. (1986) Non-stationary q-dependent processes and time-varying moving average models: invertibility properties and the forecasting problem, *Adv. Appl. Prob.* **18**, 170-210.

HAMDOUNE, S. (1995) Etude des problèmes d'estimation de certains modèles ARMA évolutifs, Ph.D. Thesis, Université Henri Poincaré, Nancy 1.

HARVEY, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.

JAZWINSKI, A.H. (1970) *Stochastic Processes and Filtering Theory*, Academic Press, New York.

JONES, R.H. (1980) Maximum likelihood fitting of ARMA models to time series with missing observations, *Technometrics C* **22**, 389-395.

KLIMKO, L.A., and NELSON, P.I. (1978) On conditional least squares estimation for stochastic processes, *The Annals of Statistics* **6**, 629-642.

KWOUN, G.H., and YAJIMA, Y. (1986) On an autoregressive model with time-dependent coefficients, *Ann. Inst. Statist. Math.* **38**, Part A, 297-309.

LJUNG, G., and BOX, G. E. P. (1979). The likelihood function of stationary autoregressive-moving average models, *Biometrika* **66**, 265-270.

LJUNG, L. and SÖDERSTRÖM, T. (1983) *Theory and Practice of Recursive Identification, MIT Press*, Cambridge MA.

MARQUARDT, D.W. (1963) An algorithm for least-squares estimation of non-linear parameters, *Journal of the Society of Industrial Applied Mathematics* **11**, 431-441.

MELARD, G. (1977) Sur une classe de modèles ARIMA dépendant du temps, *Cahiers du Centre d'Etudes de Recherche Opérationnelle* **19**, 285-295.

MELARD, G. (1982) The likelihood function of a time-dependent ARMA model, in O.D. ANDERSON and M.R. PERRYMAN (eds), *Applied Time Series Analysis*, North-Holland, Amsterdam, pp. 229-239.

MELARD, G. (1984) Algorithm AS197: A fast algorithm for the exact likelihood of autoregressive-moving average models, *J. Roy. Statist. Soc. Ser. C Appl. Statist.* **33**, 104-114.

MELARD, G. (1985) *Analyse de données chronologiques*, Coll. Séminaire de mathématiques supérieures - Séminaire scientifique OTAN (NATO Advanced Study Institute) n89, Presses de l'Université de Montréal, Montréal.

MELARD, G., and PASTEELS, J-M. (1994) User's manual of Time Series Expert (TSE version 2.2), Institut de Statistique, Université Libre de Bruxelles.

MELARD, G., and ROY, R. (1988) Modèles de séries chronologiques avec seuils, *Revue de Statistique Appliquée* **36**, 5-24.

MORF, M., SIDHU, G. S., and KAILATH, T. (1974) Some new algorithms for recursive estimation on constant, linear, discrete-time systems, *I.E.E.E. Trans. Automatic Control* **AC-19**, 315-323.

NICHOLLS, D. F. and QUINN, B. G. (1982) *Random Coefficient Autoregressive Models: An Introduction*, Spinger-Verlag, New York.

PEARLMAN, J. G. (1980) An algorithm for the exact likelihood of a high-order autoregressive-moving average process, *Biometrika* **67**, 232-233.

TJØSTHEIM, D. (1984a) Estimation in linear time series models I: stationary series, Departement of Mathematics, University of Bergen 5000 Bergen, Norway and Departement of Statistics, University of North Carolina Chapel Hill, North Carolina 27514.

TJØSTHEIM, D. (1984b) Estimation in linear time series models II: some nonstationary series, Departement of Mathematics, University of Bergen 5000 Bergen , Norway and Departement of Statistics, University of North Carolina Chapel Hill, North Carolina 27514

TJØSTHEIM, D. (1986) Estimation in nonlinear time series models, *Stochastic Processes and their Applications* **21**, 251-273.

TONG, H. (1978) On a threshold model, In C.H. Chen (Ed.), *Pattern recognition and signal processing*, Sythoff and Noordhoff, Alphen aan den Rijn, pp. 575-586.

TONG, H. (1983) *Threshold Models in Non-linear Time Series Analysis*, Springer-Verlag, New York.

TONG, H. (1990) *Non-linear Time Series: A Dynamical System Approach*, Oxford University Press, Oxford.

TUNNICLIFFE WILSON, G. (1973) The estimation of parameters in multivariate time series models, *J. Roy. Statist. Soc. Series B* **35**, 76-85.

TUNNICLIFFE WILSON, G. (1979) Some efficient computational procedures for high order ARMA models, *J. Statist. Comput. Simul.* **8**, 303-309.

Appendix 1 Proofs

Proof of Theorem 1

The matrices F_t and G_t depend on parameters while $H_t = H$ is a constant matrix equal to $H = (1, 0, \ldots, 0)$. We have thus $w_t = (W_t)_1$ as the observation equation. Consequently,

$$(W_t)_1 = \sum_{j=1}^{r} \phi_{tj} w_{t-j} + e_t - \sum_{j=1}^{r-1} \theta_{tj} e_{t-j} \quad ,$$

$$= \phi_{t1}(W_{t-1})_1 + (W_{t-1})_2 + e_t \quad ,$$

provided we let

$$(W_t)_2 = \sum_{j=2}^{r} \phi_{t+1,j} w_{t+1-j} - \sum_{j=1}^{r-1} \theta_{t+1,j} e_{t+1-j}$$

$$= \phi_{t+1,2}(W_{t-1})_1 - \theta_{t+1,1} e_t + (W_{t-1})_3 \quad ,$$

with the convention $\sum_{i=a}^{b} \bullet_i = 0 \quad \text{if} \quad a > b$. We deduce the i-th element:

$$(W_t)_i = \sum_{j=i}^{r} (\phi_{t+i-1,j} w_{t+i-1-j} - \theta_{t+i-1,j-1} e_{t+i-j}) \quad . \tag{A1.1}$$

We go on up to $(W_t)_r$ which has the following expression

$$(W_t)_r = \phi_{t+i-1,r} w_{t-1} - \theta_{t+r-1,r-1} e_t \quad .$$

Proof of Theorem 2

First notice that (4.15) is a direct consequence of (4.12). Consider equation (4.6). Since the variance $\hat{\sigma}_t^2$ defined by (4.7) is simply equal to $P_{t/t-1}(1,1)$, the explicit computation of the covariance matrix P_t yields the following form:

$$\begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & P_{t/t-1}(2,2)-\dfrac{P_{t/t-1}(2,1)P_{t/t-1}(1,2)}{P_{t/t-1}(1,1)} & \dots & P_{t/t-1}(2,r)-\dfrac{P_{t/t-1}(2,1)P_{t/t-1}(1,r)}{P_{t/t-1}(1,1)} \\ : & : & : & : \\ 0 & P_{t/t-1}(r,2)-\dfrac{P_{t/t-1}(r,1)P_{t/t-1}(1,2)}{P_{t/t-1}(1,1)} & \dots & P_{t/t-1}(r,r)-\dfrac{P_{t/t-1}(r,1)P_{t/t-1}(1,r)}{P_{t/t-1}(1,1)} \end{pmatrix},$$

Hence (4.17) is proved. Introducing the matrix P_{t-1} in expression (4.4), we obtain after straightforward algebra, given the structure of F_t, the first term in the form

$$\begin{pmatrix} P_{t/t-1}(2,2)-\dfrac{P_{t/t-1}(2,1)P_{t/t-1}(1,2)}{P_{t/t-1}(1,1)} & \dots & P_{t/t-1}(2,r)-\dfrac{P_{t/t-1}(2,1)P_{t/t-1}(1,r)}{P_{t/t-1}(1,1)} & 0 \\ : & : & : & : \\ P_{t/t-1}(r,2)-\dfrac{P_{t/t-1}(r,1)P_{t/t-1}(1,2)}{P_{t/t-1}(1,1)} & \dots & P_{t/t-1}(r,r)-\dfrac{P_{t/t-1}(r,1)P_{t/t-1}(1,r)}{P_{t/t-1}(1,1)} & 0 \\ 0 & : & 0 & 0 \end{pmatrix}$$

which proves (4.16).

Proof of Theorem 3

The elements of matrix $V = G\sigma^2 G'$ have a very simple form. They are given by the following relations:

$$V(i,j) = V(j,i) = \theta_{i-1}\theta_{j-1}\sigma^2 \quad , 2 \le i \le r \quad , \quad i \le j \le r$$

$$V(1,j) = -\theta_{j-1}\sigma^2 \quad , \quad 1 < j \le r$$

$$V(1,1) = \sigma^2.$$

We notice that the elements of the matrix C, $C = FP_0F'$, have the following form:

- the elements of the first column satisfy:

$$C(i,r) = \phi_r\{\phi_i P_0(1,1) + P_0(i+1,1)\} \quad , \quad 1 \le i \le r \quad ;$$

- by symmetry of the matrix, we have

$$C(r,j) = \phi_r\{\phi_j P_0(1,1) + P_0(1,j+1)\} \quad , \quad 1 \le j \le r \quad ;$$

- the other elements are obtained by:

$$C(i,j) = \phi_j\{\phi_i P_0(1,1) + P_0(i+1,1)\} + \phi_i P_0(1,j+1) + P_0(i+1,j+1) \quad , \quad 1 \le i,j \le r-1 \quad .$$

We can observe that all the terms which appear in the definition of the elements of matrix C, except for $P_0(i+1,j+1)$, don't depend on the elements of the first row (or of the first column) of matrix P_0. Since

$$P_0(i,j) = C(i,j) + V(i,j) \quad , \quad 1 \le i,j \le r$$

and given the relations for matrix $C(i,j)$, we can write the following relation:

$$P_0(i,j) = \phi_j\{\phi_i P_0(1,1) + P_0(i+1,1)\} + \phi_i P_0(1,j+1) + P_0(i+1,j+1) + V(i,j) \quad , \quad 1 \le i,j \le r \quad ,$$

which yields the recurrence relation between the elements of matrix P_0

$$P_0(i+1,j+1) = P_0(i,j) - (\phi_j(\phi_i P_0(1,1) + P_0(i+1,1) + \phi_i P_0(1,j+1)) - V(i,j)$$

for $1 \le i,j \le r-1$.

Appendix 2 The algorithm TKALMAR

We give here the contents of the loop over time of the algorithm TKALMAR.

```
α₁ ← α(1)
For j = 1, ..., r – 1:
        α(j) ← α(j + 1)
For j = 1, ..., p:
        α(j) ← α(j) + φ_{t+j–1,j} α₁
For k = 1, ..., r:
        For j = k, ..., r:
                P(k, j) ← V(k, j)
                If j < r then P(k, j) ← P(k, j) + P(k + 1, j + 1)
```

$b_t^2 \leftarrow P(1, 1)$
$\hat{e}_t \leftarrow w_t - \alpha_1$
$\alpha(1) \leftarrow w_t$
For $j = 2, ..., r$:
    $K \leftarrow P(1, j)/b_t^2$
    $\alpha(j) \leftarrow \alpha(j) + K\,\hat{e}_t$
    For $k = j, ..., r$:
        $P(k, j) \leftarrow P(k, j) - K\,P(1, k)$
For $k = 1, ..., r$:
    $P(1, k) \leftarrow 0$
For $j = 2, ..., q + 1$:
    $V(1, j) \leftarrow -\theta_{t+j,j}g_{t+1}$
$V(1, 1) \leftarrow g_{t+1}^2$
For $k = 2, ..., q + 1$:
    For $j = k, ..., q + 1$:
        $V(j, k) \leftarrow V(1, j)\,V(1, k)$
    $V(1, k) \leftarrow V(1, k)\,g_{t+1}$

Appendix 3 The algorithm LIKAMT

Ansley (1979) has given an algorithm for the computation of the exact likelihood function of an ARMA process with constant coefficients. In this appendix, we generalise that method to ARMA(p, q) models with time-dependent coefficients.

The algorithm LIKAMT (Mélard, 1982) is based on the Cholesky factorisation of a positive definite symmetric matrix but, to reduce the number of operations by an order of magnitude, the covariance matrix is put in the form of a band matrix. This is done by changing the variable in the density function of w such that $x_t = w_t$, for $t = 1, \ldots, p$, and $x_t = y_t$, for $t = p+1, \ldots, n$, where

$$y_t = w_t - \phi_{t1}w_{t-1} - \ldots - \phi_{tp}w_{t-p}\ , \qquad (A3.1)$$

which implies that

$$y_t = e_t - \theta_{t1}e_{t-1} - \ldots - \theta_{tq}e_{t-q}\ . \qquad (A3.2)$$

Let us consider the Wold-Cramér decomposition (Cramér, 1961), of the process $\{x_t, t \ge 1\}$. Denote $\hat{e}_t = b_t\tilde{e}_t$ the sample innovation at time t, where $\mathrm{var}(\tilde{e}_t) = \sigma^2$. The decomposition of x_t is given by

$$x_t = b_t\tilde{e}_t - \sum_{j=1}^{q(t)} \psi_{tj} b_{t-j}\tilde{e}_{t-j}\ , \qquad (A3.3)$$

where $q(t) = t-1$ for $1 \le t \le p$ and $q(t) = q$ for $t > p$, because of (A3.2). Note that the Jacobian of the transformation is equal to 1. Consequently, the likelihood function (3.1) has the form (4.13). The algorithm LIKAMT computes $\tilde{e}_t$ and b_t. The covariances

$$\gamma_{tk} = \mathrm{cov}(w_t, w_{t-k}),\quad \beta_{tk} = \mathrm{cov}(y_t, y_{t-k}),\quad \mu_{tk} = \mathrm{cov}(y_t, w_{t-k}),\quad \lambda_{tk} = \mathrm{cov}(x_t, x_{t-k})$$

are computed using the following relations:

1. the autocovariances of the MA part are computed by means of

$$\beta_{tk} = \sum_{j=k}^{q} \theta_{t-k,j-k}\theta_{tj}\sigma^2 g_{t-j}^2, \qquad (A3.4)$$

for $0 \le k \le q$, and $\beta_{tk} = 0$, for $k > q$, denoting $\theta_{t0} = -1$ for all t;

2. the covariance μ_{tk} are obtained, in the order $k = q, q-1, \ldots, 0$, by the recurrence

$$\mu_{tk} = \mathrm{cov}\left(y_t, y_{t-k} + \sum_{i=1}^{p} \phi_{t-k,i}w_{t-k-i}\right) = \beta_{tk} + \sum_{i=1}^{\min(p,q-k)} \phi_{t-k,i}\mu_{t,k+i}, \qquad (A3.5)$$

for $0 \le k \le q$, given that $\mu_{tk} = 0$, for $k > q$;

3. the autocovariances γ_{tk} of the process are obtained by the recurrences

$$\gamma_{tk} = \mathrm{cov}\left(y_t + \sum_{i=1}^{p} \phi_{ti}w_{t-i},\quad w_{t-k}\right) = \mu_{tk} + \sum_{i=1}^{p} \phi_{ti}\gamma_{t-i,k-i}, \qquad (A3.6)$$

including the variance γ_{t0};

4. the covariances λ_{tk} coincide with one of the previously mentioned covariances.

Using $(A3.3)$,we have a system of $q(t)+1$ equations

$$\lambda_{tk} = -\psi_{tk} b_{t-k}^2 \sigma^2 + \sum_{j=k+1}^{q(t)} \psi_{t-k,j-k} \psi_{tj} b_{t-j}^2 \sigma^2 \quad , \qquad (A3.7)$$

$$\lambda_{t0} = b_t^2 \sigma^2 + \sum_{j=1}^{q(t)} \psi_{tj}^2 b_{t-j}^2 \sigma^2 \quad . \qquad (A3.8)$$

We shall need the following basic algorithm for determining, at time t, the coefficients ψ_{tj} of the Wold-Cramér decomposition of $(x_t, t \geq 1)$, the standard deviation σb_t, and the sample innovation $b_t \tilde{e}_t$, as a function of the past coefficients $\psi_{t-k,j}$, standard deviations σb_{t-k}, and innovations $b_{t-k} \tilde{e}_{t-k}$, $0 \leq k \leq q(t)$, as follows:

1. $\psi_{tk} b_{t-k}^2$ is determined for $k = q(t), q(t)-1, \ldots, 1$, by using $(A3.7)$ with $\lambda_{t,q(t)}, \lambda_{t,q(t)-1}, \ldots, \lambda_{t,1}$ respectively;

2. b_t^2 is computed using $(A3.8)$;

3. $b_t \tilde{e}_t$ is obtained from $(A3.3)$.

The complete algorithm for evaluating the $\tilde{e}_t$ and the b_t, hence the likelihood, is composed of five stages.

• Stage 1

1. The coefficients $\phi_{ti} = \phi_{0i}$ are determined for $t = 1-p, \ldots, 0$ and $\theta_{tj} = \theta_{0j}$ for $t = 1-q, \ldots, 0$.

2. The values of g_t are stored for $t = 1-q, 2-q, \ldots, 0$.

3. The autocovariances γ_{tk} $(k = 0, \ldots, p)$ are determined for $t = 1-p, 2-p, \ldots, 0$, as if the process were stationary, with the constant coefficients equal to ϕ_{0i}, θ_{0j}, and g_0^2.

• Stage 2

For every $t = 1, \ldots, p$:

1. the covariances $(A3.4\text{-}6)$ are computed;

2. the basic algorithm is applied at time t, using $\lambda_{tk} = \gamma_{tk}$ $(k = 0, \ldots t-1)$.

• Stage 3

By means of expression $(A3.1)$, $x_t = y_t$ is computed for $t \geq p+1$.

• Stage 4

For every $t = p+1, \ldots, p+q$,

1. the covariances β_{tk} $(k = 0, \ldots, q)$ are computed, and the μ_{tk} $(k = q, q-1, \ldots, t-1)$ are deduced in that order, using $(A3.4\text{-}5)$;

2. the basic algorithm is applied at time t with the autocovariances $\lambda_{tk} = \beta_{tk}$ $(k = 0, \ldots, t-p-1)$, $\lambda_{tk} = \mu_{tk}$ $(k = t-p, \ldots, \min(t-1, q))$.

• Stage 5

For every $t = p+q+1$ to n, the basic algorithm is applied using the $\lambda_{t0} = \beta_{t0}, \ldots, \lambda_{t,t-q} = \beta_{t,t-q}$. Given $(A3.4)$, the equations $(A3.7\text{-}8)$ are replaced by the following ones:

$$\psi_{tk} b_{t-k}^2 = \theta_{tk} g_{t-k}^2 + \sum_{j=k+1}^{p} (\psi_{t-k,j-k} \psi_{tj} b_{t-j}^2 - \theta_{t-k,j-k} \theta_{tj} g_{t-j}^2) \ ,$$

$$b_t^2 = g_t^2 - \sum_{j=1}^{q} (\psi_{tj}^2 b_{t-j}^2 - \theta_{tj}^2 g_{t-j}^2) \ .$$