EXACT MAXIMUM LIKELIHOOD ESTIMATION FOR EXTENDED ARIMA MODELS[1]

Rajae AZRAK and Guy MÉLARD

UNIVERSITE LIBRE DE BRUXELLES,
Institut de Statistique
Campus Plaine U.L.B. CP 210,
Boulevard du Triomphe, B-1050 BELGIUM

This paper is offered to Maurice Priestley whose work on time dependent models has been fundamental.

ABSTRACT

Several extensions to autoregressive integrated moving average (ARIMA) models have been considered in the recent years. Many of them are special cases of the extended ARIMA model treated in this paper. The main features are time-dependent coefficients in the autoregressive and moving average polynomials, various types of interventions (including the usual Box and Tiao form and the innovation interventions but also interventions acting on the scale), trends on the level or on the scale, built-in deterministic seasonal components and variable transformations. In the past, estimation procedures were limited to least squares although the evaluation of the likelihood function is available for special cases, including the ARMA model with time dependent coefficients. This paper deals with maximum likelihood estimation when several features of the extended ARIMA model are taken together.

0. INTRODUCTION

Several extensions of ARIMA models have been considered in the recent years, including

(a) the use of time-dependent coefficients in the autoregressive and moving average polynomials (Quenouille, 1957; Whittle, 1965; Abdrabbo and Priestley, 1967; Miller, 1968 and 1969; Subba Rao, 1970; Mélard and Kiehm, 1981; Tyssedal and Tjøstheim, 1982; Grillenzoni, 1990),

(b) various types of interventions, including the usual Box and Tiao (1975) formulation and the innovational interventions (Fox, 1972) but also interventions acting on the scale (Mélard, 1981a; Tsay, 1988),

(c) additive (level) or multiplicative (scale) trend (Mélard, 1977),

(d) built-in deterministic seasonal components on the variable (Abraham and Box, 1978) or on the innovation (Mélard, 1981b),

(e) variable transformations (Box and Cox, 1964).

[1] We are grateful to the referee for his/her comments, and more especially for the summary that we have used at the end of Section 1.

The purpose of that model is to encompass several deterministic variations with respect to time in the framework of the usual stochastic ARIMA models. Other extensions not explicitly considered in this paper are ARMA models with GARCH errors (Bollerslev, 1986), threshold AR models (Tong, 1983), bilinear models (Subba Rao, 1981), fractional differencing ARIMA models (Granger and Joyeux, 1980). It should be noted, however, that some of these extensions can be handled using the same approach. For instance, threshold ARMA models (Mélard and Roy, 1988) can be seen as time-dependent ARMA models. Other approaches for time-dependent models include spectral density estimation (Priestley 1981, 1988), recursive estimation (Ljung and Söderström, 1983; Young, 1984) and models with random coefficients (Nicholls and Quinn, 1982; Bougerol, 1992).

Motivations for the extended ARIMA model which is used here have already been discussed elsewhere (Mélard, 1982a, 1985a). An illustration has already been provided (Mélard, 1985b). The estimation procedure was however limited to the conditional least squares approach, generalizing the approach of Box and Jenkins (1976). In this paper, an algorithm for the evaluation of the exact likelihood function is described, in the case where the innovation process is Gaussian.

1. THE MODEL

The following notations will be used:

- $\{z_t; t \in Z\}$ is the stochastic process which generates the time series $\{z_t; t = 1, \ldots, n\}$;
- $\{w_t; t \in Z\}$ is a second order stochastic process derived from $\{z_t; t \in Z\}$; it is supposed to be Gaussian, but its mean is not necessarily constant;
- $\{b_t; t \in Z\}$ is the generalized innovation process; the innovations b_t are assumed to be normally distributed independent stochastic variables, but do not necessarily constitute a stationary stochastic process with a zero mean;
- $\{a_t; t \in Z\}$ is a Gaussian white noise process in the strict sense with mean zero and variance σ^2;
- $y_t^I, y_t^D, y_t^W, y_t^M$, μ_t, and $\mu_t^{'}$ are arbitrary functions of time;
- f_t, g_t, y_t^S and y_t^V are strictly positive functions of time;
- m_t and $m_t^{'}$ are periodic functions of time;
- ϕ_{ti} and θ_{tj} are either constants or functions of time;
- $C_\lambda(.)$ is an instantaneous transformation which depends on an unknown parameter set λ;
- θ_0 is a constant;

- p, q, d and D are positive integers, and s is a strictly positive integer;
- ∇ is the regular difference operator;
- ∇_s is the seasonal difference operator with periodicity s;
- B is the backshift operator such that $B \bullet_t = \bullet_{t-1}$; it is assumed that the operator acts only on the right, e.g. $f_t B g_t = f_t g_{t-1}$.

All the functions of time included in the model have a specified analytical expression depending on a finite number of unknown parameters.

Using these notations, the extended ARIMA model for a time series is defined by the equation

$$\left(1 - \sum_{i=1}^{p} \phi_{ti} B^i\right) \frac{\nabla^d \nabla_s^D \{C_\lambda(z_t - y_t^I)/f_t\} - (y_t^D + \mu_t + m_t)}{y_t^V}$$

$$= \theta_0 + y_t^W + \left(1 - \sum_{j=1}^{q} \theta_{tj} B^j\right) [g_t \{y_t^S a_t + (y_t^M + \mu_t' + m_t')\}] . \qquad (1.1)$$

This is slightly different from the model considered by Mélard (1982a,1985b). It can be subdivided into three submodels (SM): the innovation SM, the extended ARMA SM, and the variable SM, defined as follows:

- the variable SM :

$$\frac{\nabla^d \nabla_s^D \{C_\lambda(z_t - y_t^I)/f_t\} - (y_t^D + \mu_t + m_t)}{y_t^V} = w_t; \qquad (1.2)$$

- the extended ARMA SM :

$$w_t - \sum_{i=1}^{p} \phi_{ti} w_{t-i} = \theta_0 + y_t^W + b_t - \sum_{j=1}^{q} \theta_{tj} b_{t-j}; \qquad (1.3)$$

- the innovation SM :

$$b_t = g_t \{y_t^S a_t + (y_t^M + \mu_t' + m_t')\} . \qquad (1.4)$$

Note that solving (1.4) for a_t gives

$$a_t = \frac{(b_t/g_t) - (y_t^M + \mu_t' + m_t')}{y_t^S} . \qquad (1.5)$$

The motivations behind the extended ARIMA model are the following. Using the data z_t as a starting point, the variable SM consists in the following scheme.

• The function y_t^I is subtracted; it is the intervention, in the sense of Box and Tiao (1975), or the intervention on the variable, in the sense of Tsay (1988); the y_t^I can be equal to zero at times where there is no intervention effect.

• An instantaneous non linear transformation $C_\lambda(.)$ is applied (see Box and Cox, 1964); in this paper it will be illustrated by the power family but this is not restrictive.

• A division by the deterministic function f_t is performed; that function represents a multiplicative trend (Mélard, 1977).

• The regular differences and the seasonal differences are applied in order to obtain a second order (possibly nonstationary) stochastic process.

• The function y_t^D is subtracted; it is the intervention on the data after differences. For a series which requires differencing, a level change can be represented by a pulse on y_t^D (Mélard, 1981a).

• The deterministic trend μ_t is subtracted; it is either a constant, or a function of time which represents an additive trend.

• The deterministic seasonal component m_t is subtracted; it is a periodic function of time, with a sum equal to zero on an interval of length equal to the period (Abraham and Box, 1978).

• A division by y_t^V is performed; it is an intervention function acting on the standard deviation of the variable (Hipel and McLeod, 1980).

The extended ARMA SM consists in the following stages.

• A deterministic function y_t^W is taken into account; it represents a special type of intervention which is added to the constant θ_0 used by Box and Jenkins (1976) in their ARMA model.

• An time-invariant ARMA or an evolving ARMA filter (Priestley, 1981) is applied. Note that if at least one coefficient depends on time, or if y_t^S and g_t are not identically equal to 1, then the ARMA filter is evolving.

Finally, the innovation SM consists in the following steps.

• A division of the residuals by the deterministic function $g_t$is performed; it is a trend on the standard deviation of the innovations; the purpose of this operation is to stabilize the innovation variance (Herbst, 1963; Mélard 1977).

• Y_t^M is subtracted, which is the intervention function acting on the innovation mean (Fox, 1972; Mélard, 1981; Tsay, 1988).

• $\mu_t^{'}$ is subtracted, which represents the trend of the innovation mean (Mélard 1981b).

- $m_t^{'}$ is subtracted, which is a seasonal component of the innovations.

- A division by y_t^S is performed; y_t^S is an intervention function acting on the standard deviation of the innovations in order to represent, for example, a troubled time interval during which the scatter is higher (Mélard, 1981a; Wichern *et al*., 1976).

Parameters to be estimated are usually included in the functions y_t^I, y_t^D, y_t^W, y_t^M, μ_t, $\mu_t^{'}$, f_t, g_t, y_t^S, y_t^V, m_t, $m_t^{'}$, ϕ_{ti}, θ_{tj}, and $C_\lambda(.)$; ϕ_i, θ_j, and σ^2 themselves are parameters.

The usual ARIMA model in the sense of Box and Jenkins is a special case when:

$$y_t^I = y_t^D = y_t^W = y_t^M = \mu_t = \mu_t^{'} = m_t = m_t^{'} = 0$$

$$f_t = g_t = y_t^S = y_t^V = 1$$

$$\phi_{ti} = \phi_i, \qquad i = 1, \ldots, p$$

$$\theta_{tj} = \theta_j, \qquad j = 1, \ldots, q$$

for all t, and $C_\lambda(.)$ is the identity transformation.

The order in which the various elements are introduced is justified by several considerations:

(a) not all extensions are expected to be used together; as a matter of fact, most of the time, only one or the other component will be used;

(b) the data are first adjusted for the intervention y_t^I, transformed, and divided by the multiplicative trend f_t;

(c) when the series is homogeneous, differencing is used to stabilize the level changes; at that point, the additive components (trend μ_t, seasonal m_t and intervention y_t^D) are removed;

(d) a multiplicative intervention component, y_t^V, is then taken into account, since local variations on the scale can be better perceived when the series is stable;

(e) comparison of (1.2) and (1.5) reveals a strong analogy between the variable and the innovation submodels, where $\nabla^d \nabla_s^D \{C_\lambda(z_t - y_t^I)/f_t\}$ is replaced by b_t/g_t.

In this paper, pseudo maximum likelihood parameter estimation is considered instead of the conditional least squares method in Mélard (1982a). The exact likelihood function is computed as if the innovation process were Gaussian. The process $\{z_t; t \in Z\}$ which generates the time series is not Gaussian in general. It may not even be a second order process. On the contrary, the derived process $\{w_t; t \in Z\}$ is a Gaussian process, although it is nonstationary, since the mean is not constant and the covariance between w_t and w_s does not depend only on $t - s$.

In order to give a better understanding of the problem, it is useful to consider a simpler form of the model

$$\left(1-\sum_{i=1}^{p}\phi_{ti}B^{i}\right)F_t(z_t)=\xi_t+\left(1-\sum_{j=1}^{q}\theta_{tj}B^{j}\right)\{\gamma_t a_t+\alpha_t\}, \tag{1.6}$$

where

$$F_t(z_t)=\frac{\nabla^d\nabla_s^D\{C_\lambda(z_t-y_t^I)/f_t\}-(y_t^D+\mu_t+m_t)}{y_t^V} \tag{1.7}$$

$$\xi_t=\theta_0+y_t^W \tag{1.8}$$

$$\gamma_t=y_t^S g_t \tag{1.9}$$

$$\alpha_t=g_t(y_t^M+\mu_t'+m_t'). \tag{1.10}$$

The three submodels can then be written as

- the variable SM :

$$F_t(z_t)=w_t; \tag{1.11}$$

- the extended ARMA SM :

$$\left(1-\sum_{i=1}^{p}\phi_{ti}B^{i}\right)w_t=\xi_t+\left(1-\sum_{j=1}^{q}\theta_{tj}B^{j}\right)b_t; \tag{1.12}$$

- the innovation SM :

$$b_t=\gamma_t a_t+\alpha_t, \tag{1.13}$$

where $\gamma_t>0$ for all t, and $\{a_t; t\in Z\}$ is a Gaussian white noise process with mean zero and variance σ^2.

The non linear transformation F_t and the deterministic sequences ϕ_{ti}, θ_{tj}, ξ_t, γ_t, and α_t depend on a finite number of parameters so that the model can be specified by a parameter vector of finite dimension denoted by $\underline{\beta}$ and the variance σ^2 of the white noise. Let $\delta=d+Ds$ be the number of observations lost by differencing. These observations are stored in a vector $\underline{z}_0=(z_1,z_2,\ldots,z_\delta)^{\mathrm{T}}$, where $^{\mathrm{T}}$ denotes transposition. The unconditional likelihood function which will be computed is nevertheless conditional on the $\delta=d+Ds$ first observations $\underline{z}_0$.

In the sequel of the paper, we express the likelihood function of $\underline{z}=(z_{\delta+1},\ldots,z_n)^{\mathrm{T}}$ conditional on $\underline{z}_0$ by using the density of $\underline{w}=(w_{\delta+1},\ldots\ldots,w_n)^{\mathrm{T}}$. This implies some additional assumptions on F_t (see Section 2) and an adequate treatment of a Jacobian (see Section 5). The process $\{w_t; t\in Z\}$ satisfies a time dependent ARMA model (1.12) where the b_t are independent normal random variables with mean α_t and variance $\gamma_t^2\sigma^2$. Hence the problem reduces to finding the joint density of $n - \delta$ consecutive values of a time dependent ARMA process (see Section 4). For this, it is necessary to center the process and thus to determine the mean of w_t, for all t (see Section 3).

2. THE VARIABLE SUBMODEL

Parameter estimation by the maximum likelihood method requires the computation of the exact likelihood function $L(\underline{\beta}, \sigma^2; \underline{z}/\underline{z}_0)$ which is the density of $\underline{z} = (z_{\delta+1}, \ldots, z_n)^{\mathrm{T}}$ conditional on $\underline{z}_0$. Using (1.2), it is equal to the density of $\underline{w} = (w_{\delta+1}, \ldots\ldots, w_n)^{\mathrm{T}}$, $f(\underline{w}; \underline{\beta}, \sigma^2)$, multiplied by the Jacobian of the transformation. Since w_t depends only on z_s for $s \le t$, the Jacobian matrix is triangular, and the diagonal elements are equal to $(\partial w_t)/(\partial z_t)$. Hence the Jacobian is

$$J(\underline{\beta}) = \prod_{t=\delta+1}^{n} \frac{\partial w_t}{\partial z_t} = \prod_{t=\delta+1}^{n} \frac{\partial F_t(z_t)}{\partial z_t} = \prod_{t=\delta+1}^{n} \frac{1}{f_t y_t^V} \left\{ \frac{\partial C_\lambda(z_t - y_t^I)}{\partial z_t} \right\}. \tag{2.1}$$

The function f_t will be restricted by the following condition :

$$\left(\prod_{t=\delta+1}^{n} f_t \right)^{\frac{1}{n-\delta}} = 1 \tag{2.2}$$

which means that its geometric mean over the interval from $\delta + 1$ to n is equal to 1. Similarly, the function y_t^V is subject to the constraint

$$\left(\prod_{t=\delta+1}^{n} y_t^V \right)^{\frac{1}{n-\delta}} = 1. \tag{2.3}$$

In the case of the power transformation

$$C_\lambda(z_t) = \begin{cases} \dfrac{z_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \log z_t & \lambda = 0, \end{cases} \tag{2.4}$$

we have

$$\frac{\partial C_\lambda(z_t - y_t^I)}{\partial z_t} = (z_t - y_t^I)^{\lambda - 1}. \tag{2.5}$$

Let G be the geometric mean of $z_t - y_t^I$, for $t = \delta + 1, \ldots, n$. Hence

$$L(\underline{\beta}, \sigma^2; \underline{z}/\underline{z}_0) = J(\underline{\beta}) f(\underline{w}; \underline{\beta}, \sigma^2) = G^{(n-\delta)(\lambda-1)} f(\underline{w}; \underline{\beta}, \sigma^2). \tag{2.6}$$

3. THE INNOVATION SUBMODEL

Since $\{w_t; t \in Z\}$ is a (nonstationary) Gaussian process, the distribution of $\underline{w}$ is multivariate normal with a mean vector denoted by $\underline{M}^w = E(\underline{w}) = (M^w_{\delta+1}, \ldots, M^w_n)^{\mathrm{T}}$, and a variance-covariance matrix $\mathbf{V}$. Its density has the form

$$f(\underline{w}; \underline{\beta}, \sigma^2) = (2\pi)^{-(n-\delta)/2} (\det \mathbf{V})^{-1/2} \exp\left\{-\frac{1}{2}(\underline{w} - \underline{M}^w)^{\mathrm{T}} \mathbf{V}^{-1} (\underline{w} - \underline{M}^w)\right\}. \tag{3.1}$$

In this section, we consider the computation of $\underline{M}^w$ which relies mainly on the innovation submodel. Since $E(a_t) = 0$, we have from (1.13) and (1.12)

$$M^b_t = E\{b_t\} = \alpha_t, \tag{3.2}$$

$$M^w_t = E\{w_t\} = \sum_{i=1}^{p} \phi_{ti} M^w_{t-i} + M^b_t - \sum_{j=1}^{q} \theta_{tj} M^b_{t-j} + \xi_t. \tag{3.3}$$

To determine the initial M^w_t in order to initialize the recurrence (3.3), the following assumptions are made:

H1 $\phi_{ti} = \phi_{\delta+1,i} = \phi_i$ $(i = 1, ..., p)$ for all $t \leq \delta$;
$\theta_{tj} = \theta_{\delta+1,j} = \theta_j$ $(j = 1,..., q)$ for all $t \leq \delta$;

H2(a) $\phi(B) = 1 - \sum_{i=1}^{p} \phi_i B^i$ has all its zeroes outside of the unit circle;

H2(b) $\theta(B) = 1 - \sum_{j=1}^{q} \theta_j B^j$ has all its zeroes outside of the unit circle;

H3 $y^W_t = y^M_t = 0$ for $t \leq \delta$;

H4 $g_t = g_{\delta+1}$, $\mu_t = \mu_{\delta+1}$, and $\mu'_t = \mu'_{\delta+1}$ for $t \leq \delta$.

By H1 and H3, (3.3) implies

$$\phi(B) M^w_t = \theta(B) M^b_t + \theta_0 \tag{3.4}$$

for $t \leq \delta$. In addition, H4 and H2(a) guarantee that the process $\tilde{w}_t = w_t - M^w_t$ has a stationary behavior at least for $t \leq \delta$. By H2(b), the process is a (possibly nonstationary) second order invertible process in the sense of Hallin (1980).

Let us denote

$$\psi(B) = \phi^{-1}(B)\theta(B) = \sum_{h=0}^{\infty} \psi_h B^h. \tag{3.5}$$

Hence

$$M^w_t = \psi(B) M^b_t + \phi^{-1}(B)\theta_0. \tag{3.6}$$

By H3 and H4:

$$M_t^b = (\mu'_{\delta+1} + m'_t) g_{\delta+1} \tag{3.7}$$

for $t \le \delta$. But m'_t is a periodic function of t, with period s, which implies that, for $t \le \delta$, M_t^b is also a periodic function with period s, which can be computed for all t.

Noting that the coefficients

$$\Psi_h = \sum_{i=0}^{\infty} \psi_{is+h}, \tag{3.8}$$

$h = 0, 1, ..., s - 1$, can be computed in a finite number of steps, we write

$$\psi(B) M_t^b = \sum_{h=0}^{s-1} \Psi_h M_{t-h}^b, \tag{3.9}$$

and since

$$\phi^{-1}(B)\theta_0 = \frac{\theta_0}{\phi(1)}, \tag{3.10}$$

we obtain

$$M_t^w = \sum_{h=0}^{s-1} \Psi_h M_{t-h}^b + \frac{\theta_0}{\phi(1)} \tag{3.11}$$

for $t = \delta + 1 - p, \ldots, \delta$.

At that point, the recurrence relation (3.3) can be used for $t = \delta + 1, \ldots, n$.

4. THE EXTENDED ARMA SUBMODEL

There remains to evaluate det$\mathbf{V}$ and the quadratic form $(\underline{w} - \underline{M}^w)^{\mathrm{T}} \mathbf{V}^{-1} (\underline{w} - \underline{M}^w)$. In the case of a time-invariant model (e.g., an ARMA model), Schweppe (1965) has shown that the computation can be expressed in terms of the sample normalized process $\{\tilde{a}_t\}$, which are simply the innovations of a zero-mean process $\{\tilde{w}_t\}$ evaluated in terms of the observations, and normalized in such a way that their variance is σ^2. In the original presentation the $\tilde{a}_t$ were obtained by means of the Kalman filter. Ansley (1979) has developed an algorithm based on a Cholesky factorisation for band matrices which is computationally nearly equivalent. For stationary processes, fast Kalman algorithms (also called Chandrasekar-type, Morf *et al.*, 1974) have been described (Pearlman, 1980) and implemented (Mélard, 1984).

If the model is not strictly time-invariant, the arguments of Schweppe can be generalized. The Kalman filter algorithm remains valid (Ansley and Kohn, 1983) provided that some assumptions (like H1 and H2 above) are used (Mélard, 1985a). The Cholesky factorization approach can also be adapted (Mélard, 1982b).

In our problem $\tilde{w}_t = w_t - M_t^w$, $t > \delta$, where the M_t^w have been computed in the previous section.

Two cases are considered.

• Time-dependent process. If the assumptions H1-H4 are fulfilled, $\{\tilde{w}_t; t > \delta\}$ is a zero-mean time-dependent (or evolving) ARMA process with a stationary behavior in the past (before data are available). Note that these assumptions provide a natural way to extend the model in the past but alternative assumptions can be considered. For example, if the coefficients ϕ_{ti} and θ_{tj} are periodic functions of time, other assumptions will appear more natural (Li and Hui, 1988).

• Time-invariant process. If, beside H1-H4, the assumptions
H5- $f_t = g_t = y_t^S = y_t^V = 1$ for all t
H6- $\phi_{ti} = \phi_i$ (i = 1, ..., p) and $\theta_{tj} = \theta_j$ (j = 1,..., q) for all t
are fulfilled, then the $\{\tilde{w}_t; t > \delta\}$ consist in a zero-mean (time-invariant) ARMA process. This case is well known and will not be discussed.

In the time-dependent case, everything is based on the Wold-Cramér decomposition (Cramér, 1961; Rissanen and Barbosa, 1969) or innovation representation of the process for $t \geq \delta$. Since the algorithmic problems have been described in the references, we shall focus on the assumptions and the main ideas by working out an example.

Let us consider the mixed first order autoregressive-first-order moving average time-dependent process, TDARMA(1,1), defined by the equation :

$$\tilde{w}_t = \phi_t \tilde{w}_{t-1} + \tilde{b}_t - \theta_t \tilde{b}_{t-1} \tag{4.1}$$

where the $\tilde{b}_t$ are the innovations, with mean 0 and variance $\sigma_t^2 > 0$, say. Let us assume that δ = 0 to simplify matters and, in accordance with H1, H2 and H4: $\phi_t = \phi_0$, $\theta_t = \theta_0$, and $\sigma_t^2 = \sigma_0^2$ for $t \leq 0$, with $|\phi_0| < 1$, and $|\theta_0| < 1$.

In the time-invariant case, these equalities hold for all t and the autocovariances of the process can be expressed in terms of $\sigma_t^2 = \sigma^2$ and the coefficients of the ARMA model. For example the fast algorithm of Wilson (1979) can be used.

In the time-dependent case, the autocovariances of the stationary process defined by

$$\tilde{w}_t = \phi_0 \tilde{w}_{t-1} + \tilde{b}_t - \theta_0 \tilde{b}_{t-1} \tag{4.2}$$

are computed in the same way. They provide starting values for recurrences with respect to s and t, from which the covariances $\gamma_{ts} = \mathrm{cov}(\hat{w}_t, \hat{w}_s)$ are derived (for more details, see Mélard, 1982b).

Note that the Wold-Cramér decomposition of a time-dependent ARMA process has no direct relation with the rational function $(1 - \theta_t z)/(1 - \phi_t z)$ (Mélard, 1985a; see also Mélard and Herteleer-De Schutter, 1989 for consequences in spectral analysis of nonstationary processes and Hallin, 1986, for related problems).

Because of the autoregressive part of the process, the Wold-Cramér decomposition of $\tilde{w}_t$ makes use of $\tilde{b}_s$ for s = 1, ..., t. To simplify the decomposition, let us consider $\underline{\tilde{w}}^*$ defined by:

$$\begin{cases} \tilde{w}_1^* = \tilde{w}_1 & \text{for} \quad t = 1 \\ \tilde{w}_t^* = \tilde{w}_t - \phi_t \tilde{w}_{t-1} & \text{for} \quad t \geq 2 \end{cases} \tag{4.3}$$

The process $\{\tilde{w}_t^*, t > 0\}$ is nonstationary but note that

(i) it is q-dependent, with q = 1, in the sense that $\gamma_{ts}^* = \mathrm{cov}(\tilde{w}_t^*, \tilde{w}_s^*) = 0$ for $|t - s| > 1$.

(ii) $\gamma_{ts}^* = \mathrm{cov}(\tilde{w}_t^*, \tilde{w}_s^*)$, $|t - s| = 0$ or 1, can be deduced from the γ_{ts};

(iii) the densities $f(\underline{\tilde{w}};\underline{\beta},\sigma^2)$ and $f(\underline{\tilde{w}}^*;\underline{\beta},\sigma^2)$ are equal because the Jacobian of the transformation from $\underline{\tilde{w}}$ to $\underline{\tilde{w}}^*$ is unity.

There remains to evaluate $\underline{\tilde{w}}^{*\mathrm{T}}\mathbf{V}^{*-1}\underline{\tilde{w}}^*$ and $\det\mathbf{V}^*$, where $\mathbf{V}^*$ is composed of the γ_{ts}^*. But, since the process is 1-dependent, we can write its Wold-Cramér decomposition as

$$\begin{cases} \tilde{w}_1^* = h_1\tilde{b}_1 & \text{for} \quad t = 1 \\ \tilde{w}_t^* = h_t\tilde{b}_t - \psi_t h_{t-1}\tilde{b}_{t-1} & \text{for} \quad t = 2, \ldots, n, \end{cases} \tag{4.4}$$

where $h_t\tilde{b}_t$ is the innovation at time t, and $\mathrm{var}(\tilde{b}_t) = \sigma^2$.

Neglecting computational efficiency considerations, we use the Gram-Schmidt orthogonalization procedure on (4.4). Starting with $h_1^2 = \gamma_{1,1}^*/\sigma^2$ and $\tilde{b}_1 = \tilde{w}_1^*/h_1$, we have

$$\begin{cases} \psi_t = -\dfrac{\gamma_{t-1,t}^*}{h_{t-1}^2\sigma^2} \\ h_t^2 = \dfrac{\gamma_{t,t}^*}{\sigma^2} - \psi_t^2 h_{t-1}^2 \\ \tilde{b}_t = \dfrac{\tilde{w}_t^* + \psi_t h_{t-1}\tilde{b}_{t-1}}{h_t} \end{cases} \tag{4.5}$$

for $t > 1$.

We now leave the example and consider the general case where more efficient algorithms are given by Ansley and Kohn (1983) using the Kalman filter, and Mélard (1982b), using the Cholesky factorization (incidentaly the Kalman filter algorithm can be used in the vector time-dependent ARMA case, see Mélard, 1985a). These algorithms require a number of operations (multiplications and divisions) of order $O(n(p + q^2 + 4q))$, which is about less than twice the number of operations for equivalent algorithms in the time-invariant case.

5. MAXIMUM LIKELIHOOD ESTIMATION

Let us now consider the vector of sample normalized innovations $\underline{\tilde{b}} = (\tilde{b}_{\delta+1}, \ldots . \tilde{b}_n)^{\mathrm{T}}$ which are independent normal random variables. Hence, taking care of the Jacobian of the transformation from the $\tilde{w}_t^*$ to the $\tilde{b}_t$, the likelihood function (2.6) is expressed by

$$L(\underline{\beta}, \sigma^2; \underline{\tilde{b}}) = G^{(n-\delta)(\lambda-1)}(2\pi)^{-\frac{n-\delta}{2}}\sigma^{-(n-\delta)}\left(\prod_{t=\delta+1}^{n} h_t^{-1}\right)\exp\left\{-\frac{1}{2\sigma^2}\sum_{t=\delta+1}^{n}\tilde{b}_t^2(\underline{\beta})\right\}, \tag{5.1}$$

where $\tilde{b}_t(\underline{\beta})$ is the value of $\tilde{b}_t$ derived from the $\tilde{w}_t = w_t - M_t^w$, via the $\tilde{w}_t^*$, in function of the parameters $\underline{\beta}$.

There is no analytic expression of the likelihood function. Maximizing it numerically with respect $\underline{\beta}$ and σ is difficult. Since G and the h_t depend on $\underline{\beta}$, the problem is simplified by letting

$$\tilde{b}_{*t}(\underline{\beta}) = \tilde{b}_t(\underline{\beta}) G^{-(\lambda-1)} \left(\prod_{t=\delta+1}^{n} h_t \right)^{\frac{1}{n-\delta}}, \tag{5.2}$$

and

$$\sigma_* = \sigma G^{-(\lambda-1)} \left(\prod_{t=\delta+1}^{n} h_t \right)^{\frac{1}{n-\delta}}. \tag{5.3}$$

It is equivalent to maximize

$$(2\pi)^{-\frac{n-\delta}{2}} \sigma_*^{-(n-\delta)} \exp\left\{ -\frac{1}{2\sigma_*^2} \sum_{t=\delta+1}^{n} \tilde{b}_{*t}^2(\underline{\beta}) \right\}, \tag{5.4}$$

or to minimize

$$\frac{n-\delta}{2} \log(\sigma_*^2) + \frac{1}{2\sigma_*^2} \sum_{t=\delta+1}^{n} \tilde{b}_{*t}^2(\underline{\beta}). \tag{5.5}$$

It is easy to solve the likelihood equation for σ_*^2

$$-\frac{n-\delta}{2\sigma_*^2} + \frac{1}{2\sigma_*^4} \sum_{t=\delta+1}^{n} \tilde{b}_{*t}^2(\underline{\beta}) = 0, \tag{5.6}$$

whose solution is

$$\hat{\sigma}_*^2 = \frac{1}{n-\delta} \sum_{t=\delta+1}^{n} \tilde{b}_{*t}^2(\underline{\beta}). \tag{5.7}$$

Substituting in (5.5) leads to minimizing

$$\frac{n-\delta}{2} \log\left(\frac{1}{n-\delta} \sum_{t=\delta+1}^{n} \tilde{b}_{*t}^2(\underline{\beta}) \right) \tag{5.8}$$

or the sum of squares

$$\left(\prod_{t=\delta+1}^{n} h_t^2 \right)^{\frac{1}{n-\delta}} G^{-2(\lambda-1)} \sum_{t=\delta+1}^{n} \tilde{b}_t^2(\underline{\beta}). \tag{5.9}$$

Very often, the procedure to minimize the sum of squares is the Marquardt's method (1963) but other procedures can be used. Let $\underline{\hat{\beta}}$ be the maximum likelihood estimator for $\underline{\beta}$. As a consequence, the maximum likelihood estimator of σ^2 is

$$\frac{1}{(n-\delta)} \sum_{t=\delta+1}^{n} \tilde{b}_t^2(\underline{\hat{\beta}}). \tag{5.10}$$

References

ABDRABBO, N. A. and PRIESTLEY, M. B. (1967) On the prediction of non-stationary processes, *J. Roy. Statist. Soc. Ser. B* **29**, 570-585.

ABRAHAM, B. and BOX, G. E. P. (1978) Deterministic and forecast-adaptative time-dependent models, *J. Roy. Statist. Soc. Ser. C Appl. Statist.* **27**, 120-130.

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.

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

BOUGEROL, P. (1992) Kalman filtering with random coefficients and contractions, *Siam J. Contr. Optim.* (to appear).

BOX, G. E. P. and COX, D. R. (1964) An analysis of transformations, *J. Roy. Statist. Soc. Ser. B* **26**, 211-243.

BOX, G. E. P. and JENKINS, G. M. (1976) *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco (revised edition).

BOX, G. E. P. and TIAO, G. C. (1975) Intervention analysis with applications to economic and environmental problems, *J. Amer. Statist. Assoc.* **70**, 70-79.

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.

FOX, A. J. (1972) Outliers in time series, *J. Roy. Statist. Soc. Ser. B* 34**,** 350-363.

GRANGER, C. W. J. and JOYEUX, R. (1980) An introduction to long-memory time series models and fractional differencing, *J. Time Series Anal.* **1**, 15-29.

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

HALLIN, M. (1980) Invertibility and generalized invertibility of time series models, *J. Roy. Statist. Soc. Ser. B* **42**, 210-212; **43**, 103**.**

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.

HERBST, L. J. (1963) A test for variance heterogeneity in the residuals of a Gaussian moving average, *J. Roy. Statist. Soc. Ser. B* **25**, 451-454.

HIPEL, K. W. and McLEOD, A. I. (1980) Perspectives in stochastic hydrology, in O. D. ANDERSON (ed.) *Time Series*, North-Holland, Amsterdam, pp. 73-102.

KOHN, R. and ANSLEY, C.F. (1982) A note on obtaining the theoretical autocovariances of an ARMA process, *J. Statist. Comput. Simul.* **15**, 273-283.

LI, W. K. and HUI, Y. V. (1988) An algorithm for the exact likelihood of periodic autoregressive moving average models, *Commun. Statist. - Simula.* **17**, 1483-1494.

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

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**, 1977, 285-295.

MELARD, G. (1981a) On an alternative model for intervention analysis, in O. D. ANDERSON and M. R. PERRYMAN (eds) *Time Series Analysis*, North-Holland, Amsterdam, 1981, pp. 345-354.

MELARD, G. (1981b) On a deterministic sub-model for the innovation process in ARIMA model, *1981 Proceedings of the Business and Economic Statistics Section*, American Statistical Association, Washington D.C., 1981, pp. 329-333.

MELARD, G. (1982a) Software for time series analysis, in H. CAUSSINUS, P. ETTINGER and R. TOMASSONE (eds), *COMPSTAT 1982 Part 1: Proceedings in Computational Statistics*, Physica-Verlag, Vienna (Austria), 1982, pp. 336-341.

MELARD, G. (1982b) 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. (1985a) *Analyse de données chronologiques*, Coll. Séminaire de mathématiques supérieures - Séminaire scientifique OTAN (NATO Advanced Study Institute) n°89, Presses de l'Université de Montréal, Montréal.

MELARD, G. (1985b) Illustration of the use of a general time series model, in O.D. ANDERSON (ed.) *Time Series Analysis: Theory and Practice 7, Proceedings of the (general interest) International Conference held at Toronto, Canada, 18-21 August 1983*, North-Holland, Amsterdam, pp. 53-75.

MELARD, G. and HERTELEER-DE SCHUTTER, A. (1989) Contributions to the evolutionary spectral theory, *J. Time Series Anal.* **10**, 41-63.

MELARD, G. and KIEHM, J.-L. (1981) ARIMA models with time-dependent coefficients for economic time series, in O.D. ANDERSON and M.R. PERRYMAN (eds), *Time Series Analysis*, North-Holland, Amsterdam, 1981, pp. 355-363.

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

MILLER, K. S. (1968) *Linear Difference Equations,* Benjamin, New York, 1968.

MILLER, K. S. (1969) Nonstationary autoregressive processes, *I.E.E.E. Transactions Information Theory* **IT-15**, 1969, 315-316.

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.

PRIESTLEY, M. B. (1981) *Spectral Analysis and Time Series,* Volume 2, Academic Press, New York.

PRIESTLEY, M. B. (1988) *Non-Linear and Non-Stationary Time Series Analysis*, Academic Press, New York.

RISSANEN, J. and BARBOSA, L. (1969) Properties of infinite covariance matrices and stability of optimum predictors, *Information Sci.* **1**, 221-236.

QUENOUILLE, M. H. (1957) *The Analysis of Multiple Time Series,* Griffin, London.

SCHWEPPE, F. C. (1965) Evaluation of likelihood functions of Gaussian signals, *I.E.E.E. Transactions Information Theory* **IT-11**, 61-70.

SUBBA RAO, T. (1970) The fitting of non-stationary time-series models with time dependent parameters, *J. Roy. Statist. Soc. Ser. B* **32**, 1970, 312-322.

SUBBA RAO, T. (1981) On the theory of bilinear models, *J. Roy. Statist. Soc. Ser. B* **42**, 245-292.

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

TSAY, R. S. (1988) Outliers, level shifts, and variance changes in time series, *J. of Forecasting* **7**, 1-20.

TYSSEDAL, J. S. and TJØSTHEIM, D. (1982) Autoregressive processes with a time-dependent variance, *J. Time Series Anal.* **3**, 209-217.

YOUNG, P. C. (1984) *Recursive Estimation and Time Series Analysis,* Springer-Verlag, Berlin.

WHITTLE, P. (1965) Recursive relations for predictors of non-stationary processes, *J. Roy. Statist. Soc. Ser. B* **27**, 523-532.

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

WICHERN, D. W., MILLER, R. B. and HSU, D. A. (1976) Changes of variance in first-order autoregressive time series models - with an application, *J. Roy. Statist. Soc. Ser. C Appl. Statist.* **25**, 248-256.