T E C H N I C A L

R E P O R T

0442

ASYMPTOTIC PROPERTIES OF QUASI-MAXIMUM LIKELIHOOD ESTIMATORS FOR ARMA MODELS WITH TIME-DEPENDENT COEFFICIENTS

AZRAK, R. and G. MELARD

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

INTERUNIVERSITY ATTRACTION POLE

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

ASYMPTOTIC PROPERTIES OF

QUASI-MAXIMUM LIKELIHOOD ESTIMATORS

FOR ARMA MODELS WITH TIME-DEPENDENT COEFFICIENTS

by Rajae AZRAK and Guy MÉLARD[1]

Not to be quoted without the authors agreement. October 21, 2004

Abstract

For about thirty years, time series models with time-dependent coefficients have sometimes been considered as an alternative to models with constant coefficients or non-linear models. Analysis based on models with time-dependent models has long suffered from the absence of an asymptotic theory except in very special cases. The purpose of this paper is to provide such a theory without using a locally stationary spectral representation and time rescaling. We consider autoregressive-moving average (ARMA) models with time-dependent coefficients and a heteroscedastic innovation process. The coefficients and the innovation variance are deterministic functions of time which depend on a finite number of parameters. These parameters are estimated by maximising the Gaussian likelihood function. Deriving conditions for consistency and asymptotic normality and obtaining the asymptotic covariance matrix are done using some assumptions on the functions of time in order to attenuate non-stationarity, mild assumptions for the distribution of the innovations, and also a kind of mixing condition. Theorems from the theory of martingales and mixtingales are used. Some simulation results are given and both theoretical and practical examples are treated.

---

[1] *AMS 1991 subject classification*. Primary 62M15; secondary 62M05.
Key words and phrases. Nonstationary process, time series, time-dependent model.
Correspondence to be sent to the second author, Université Libre de Bruxelles, ECARES and Institut de Statistique et de Recherche Opérationnelle CP 210, Campus Plaine U.L.B., Bd du Triomphe, B-1050 Brussels, Belgium (E-mail address: gmelard@ulb.ac.be). This work has benefited from a grant "Action de Recherche Concertée" 96/01/205 of "Communauté Française Wallonie-Bruxelles" and an IAP-network in Statistics grant, contract P5/24, financially supported by the Belgian Federal Office for Scientific, Technical and Cultural Affairs (OSTC). A large part has been written while the second author was at "Institut National de Statistique et d'Economie Appliquée" (INSEA), Rabat, Morocco, with a grant of "Agence Générale de Coopération au Développement" of Belgium. We thank those who have made comments and suggestions on preliminary versions of this paper, including Denis Bosq, Christian Francq, Marc Hallin, Roch Roy, Benoît Truong Van and Bas Werker, and mainly Rainer Dahlhaus, as well as the referees of a previous version, and two referees of this submission.

1. Introduction

Apart from the recent interest towards non-linear models, most of the literature on time series is concerned with stationary linear models. Time invariance is however difficult to justify over a long period of time but provides a welcome simplification. If non-linear models have been the object of numerous papers in the recent years, it is mainly for theoretical aspects, much less for practical applications. See e.g. Priestley (1988), Tong (1990), Guégan (1994), for reviews. The power of the theory on non-linear models relies on results depending on stationarity and ergodicity. On the contrary, models with time-dependent coefficients have been less studied in the literature and also rarely used in practice. The main reason is their lack of stationarity and ergodicity which has prevented from using standard arguments. Among the first authors who have investigated these models, let us mention Quenouille (1957), Whittle (1965), Abdrabbo and Priestley (1967), Miller (1968, 1969), Subba Rao (1970), Wegman (1974), and Hallin and Mélard (1977). Among the recent references on the subject, let us mention Singh and Peiris (1987), Kowalski and Szynal (1991), Azrak and Mélard (1993) where the extended ARIMA model includes some cases of models with time-dependent coefficients and marginally heteroscedastic models, Grillenzoni (1990) who has used an approach based on recursive estimation, and Dahlhaus (1996 a, b, c, 1997) and Bibi and Francq (2003), whose contributions will be discussed below. For conditionally heteroscedastic covariance stationary models, see Lumsdaine (1996). For ARMA processes with periodic coefficients, see Basawa and Lund (2001).

The asymptotic properties of estimators for models with time-dependent coefficients are not studied in a general framework in the statistical literature although Tjøstheim (1984b) has been close from it with his general theory for linear and nonlinear models. The difficulty to derive conditions for consistency and asymptotic normality, and to obtain the asymptotic covariance matrix is due to three reasons: (a) the observations are not independent; (b) they are not identically distributed; (c) they are not normally distributed.

In this paper, we consider asymptotic properties of quasi-maximum likelihood estimators for a large class of models, the autoregressive-moving average (ARMA) model with time-dependent coefficients and heteroscedastic innovation variance. These coefficients and that variance are assumed to be deterministic functions of time which depend on a finite number of parameters. There are some assumptions on these functions of time in order to attenuate non-stationarity, mild assumptions for the distribution of the innovations, and the estimator which is used is the Gaussian maximum likelihood estimator, maximising the likelihood function as if the process were Gaussian. The advantage of that approach is that the Gaussian likelihood function can be computed exactly, with very efficient algorithms (e.g. Mélard, 1982, Azrak and Mélard, 1998).

Let us consider a stochastic process $w = (w_t, t \in N)$, defined on a probability space (Ω, F, P_β), with values in R, whose distribution depends on a vector $\beta = (\beta_1, \ldots, \beta_r)$ of unknown parameters to be estimated, with β lying in an open set B of an Euclidian space R^r, $r \in N_0$. The true value of β is denoted by β^0. Consider a triangular sequence of observations $w^{(n)} = (w_1^{(n)}, w_2^{(n)}, \ldots, w_n^{(n)})$ of the process w. Let $F_t^{(n)}$ be the σ-field generated by the $(w_u^{(n)}, u \le t)$, and $F_0^{(n)} = \{\emptyset, \Omega\}$. Let $Q_n = Q_n(\beta) = Q_n(\beta; w_1^{(n)}, \ldots, w_n^{(n)})$ be a general penalty function to be minimised, which depends on the observations and β. The estimator is obtained by solving the system of equations

$$\frac{\partial Q_n(\beta)}{\partial \beta_i} = 0 \quad , \quad \text{for} \quad i = 1, \ldots, r.$$

To use the method of maximum likelihood we maximise, with respect to β, the function

$$l_n(\beta; w_1^{(n)}, \ldots, w_n^{(n)}) = \sum_{t=1}^{n} \log f_t(w_t^{(n)}; \beta / F_{t-1}^{(n)}) \quad ,$$

where $f_t(w_t^{(n)}; \beta / F_{t-1}^{(n)})$ is the conditional density of $w_t^{(n)}$ with respect to $F_{t-1}^{(n)}$. For ARMA models, for a reason explained in Section 3, $^{(n)}$ can be omitted for the filtration, leading to F_{t-1}. In what follows, we briefly describe a synthesis of the literature.

Very general non-stationary models have been considered. Silvey (1961) has tried to generalise the statistical properties of maximum likelihood estimators in the non-standard case without specifying a particular model but under assumptions which are difficult to check. Bar-Shalom (1971) shows, in the case of a scalar parameter, convergence in probability and asymptotic efficiency by assuming that, at $\beta = \beta^0$:

$$-E\left(\frac{\partial^2 f_t(w_t^{(n)};\beta/F_{t-1})}{\partial\beta^2}\right) = E\left(\frac{\partial f_t(w_t^{(n)};\beta/F_{t-1})}{\partial\beta}\right)^2 \quad . \tag{1.1}$$

Bhat (1974) extends the results of Bar-Shalom (1971) by showing asymptotic normality of the estimator under the assumption that, at $\beta = \beta^0$,

$$J_t(\beta) = -E\left(\frac{\partial^2 f_t(w_t^{(n)};\beta/F_{t-1})}{\partial\beta^2}/F_{t-1}\right) = E\left(\left(\frac{\partial f_t(w_t^{(n)};\beta/F_{t-1})}{\partial\beta}\right)^2/F_{t-1}\right) \quad , \tag{1.2}$$

and that

$$\frac{1}{n}\sum_{t=1}^{n} J_t(\beta^0) \to J(\beta^0) \quad , \tag{1.3}$$

where $J(\beta^0)$ is a strictly positive constant. In a more general framework, Crowder (1976) gives conditions which guarantee weak consistency of the maximum likelihood estimator. One of these conditions cannot be verified in our case.

Let us now restrict ourselves to more particular models in the class of the ARMA (autoregressive-moving average) processes with time-dependent coefficients. Subba Rao (1970) is interested in AR models using estimators based on the evolutionary spectral analysis of Priestley (1965). Mélard (1977) fits (marginally) heteroscedastic ARMA models called ARMAG models. Mélard and Kiehm (1981) deal with models with time-dependent coefficients and Gaussian maximum likelihood estimation. Mélard (1982) gives an algorithm for computing the exact Gaussian likelihood. These authors have not studied the asymptotic properties of their methods.

Kwoun and Yajima (1986) introduce a first-order AR process with a time-dependent coefficient. They show consistency and asymptotic normality of the least squares estimator, by assuming that the coefficients of the moving average representation are uniformly bounded with respect to t. Hamdoune (1995) extends that approach to autoregressive processes of order p, using also an M-estimateur, and shows strong consistency and asymptotic normality under certain regularity assumptions. The first-order moving average process is also considered.

It is important to notice that the other approaches mentioned before are concerned with homoscedastic innovations. Tyssedal and Tjøstheim (1982), however, consider heteroscedastic autoregressive models. We provide results partly based on the work of Klimko and Nelson (1978), and of Tjøstheim (1984a, 1984b, 1986) which consist in using a Taylor expansion of the penalty function, here the Gaussian likelihood. Tjøstheim (1986) has considered various linear and non-linear autoregressive models, detailed in Tjøstheim (1984a) and Tjøstheim (1984b), respectively, but with a constant innovation variance. Estimation is performed by the maximum likelihood method only for stationary processes, not for processes with time-dependent coefficients for which only the least squares method is used. We shall start from two theorems of Klimko and Nelson (1978) denoted here by KN1 and KN2. KN1 states existence of an estimator, and shows almost sure (a.s.) consistency, whereas KN2 proves asymptotic normality and gives the asymptotic covariance matrix. These theorems are based on a standard technique of Taylor expansion of a general penalty function and on Egorov's theorem. We have not used the versions of Tjøstheim (1986) of KN1 because the assumption on the smallest eigenvalue of a certain matrix is difficult to check when there are more than two parameters. Moreover, the original assumption of KN1 is still necessary for proving KN2.

The difficulty is to check the assumptions of Klimko and Nelson (1978) by stating realistic assumptions on the model with the Gaussian likelihood taken as the penalty function. For the general ARMA case, we give expressions which are generalisations of those given by Kwoun and Yajima (1986) for the AR(1) process. They seem relatively easy to check at least in special cases. For pure autoregressive processes, in a companion paper (Azrak and Mélard, 2004), we propose an alternative mixing condition which assumes that a certain series in the ϕ-mixing coefficients is convergent. Of course, theorems from the theory of martingales are used, essentially a strong law of large numbers (Stout, 1970) and a version of Basawa and Prakasa Rao (1980) of the central limit theorem. For the case where the coefficients depend on time but also on the length of the series, a martingale array is considered and a weak law of large numbers is used to obtain convergence in probability of the estimator instead of almost sure convergence. Also, in order to bound the higher-order term of the Taylor expansion, laws of large numbers for mixtingales are used, either for sequences if the coefficients depend on t, or for arrays if they also depend on n.

More recently, Dahlhaus (1996 a, b, c, 1997) has obtained asymptotic results for a new class of locally stationary processes which includes heteroscedastic ARMA processes with time-dependent coefficients. Either a spectral-based or a maximum likelihood estimation method is used. The asymptotics are based on rescaling time: instead of simply increasing the length n of the series, Dahlhaus assumes that the observation period is fixed, let us say $[0, 1]$, but that the interval between the n observations decreases and tends to zero.

Let us mention another very recent approach by Bibi and Francq (2003). Their assumptions are different, e. g. they assume only finite fourth-order moments but other conditions more difficult to check in our case, their estimation method is quasi-least squares, and their scope of applications favors cyclical ARMA models with non constant periods.

We proceed in two stages. First, we consider the quasi-maximum likelihood estimator of a process which is not necessarily stationary (see Section 2). Then, we particularise to ARMA models with time-dependent coefficients (see Section 3). Section 4 is devoted to five examples where the assumptions of Section 3 can be verified. Some simulation results for examining small sample properties of estimators and their standard errors are given in Section 5 in the case of some examples of Section 4. We end with practical examples on standard time series (Section 6), and some conclusions.

2. Maximum likelihood estimation for a general time series model

In this section, we apply the two theorems KN1 and KN2 of Klimko and Nelson (1978) in the case of a quasi-maximum likelihood. Let us consider $\hat{w}_{t/t-1}(\beta)$, the conditional expectation of w_t given F_{t-1}:

$$\hat{w}_{t/t-1}(\beta) = E_\beta(w_t/F_{t-1}) \quad , \tag{2.1}$$

and similarly the conditional variance:

$$h_{t/t-1}(\beta) = E_\beta((w_t - \hat{w}_{t/t-1}(\beta))^2/F_{t-1}) \quad . \tag{2.2}$$

We take as penalty function, Q_n, minus the logarithm of the quasi-likelihood function $L_n(\beta; w_1, \dots, w_n)$, computed as if the process were Gaussian:

$$L_n(\beta; w_1, \dots, w_n) = (2\pi)^{-n/2}\left(\prod_{t=1}^{n} h_{t/t-1}(\beta)\right)^{-1/2} \exp\left(-\sum_{t=1}^{n} \frac{(w_t - \hat{w}_{t/t-1}(\beta))^2}{2h_{t/t-1}(\beta)}\right) \quad ,$$

hence

$$-Q_n = l_n(\beta) = \log L_n(\beta; w_1, \dots, w_n) = -\frac{1}{2}\sum_{t=1}^{n} \alpha_t(\beta) - \frac{n}{2}\log(2\pi) \quad , \tag{2.3}$$

where

$$\alpha_t(\beta) = \log h_{t/t-1}(\beta) + \frac{(w_t - \hat{w}_{t/t-1}(\beta))^2}{h_{t/t-1}(\beta)} \quad . \tag{2.4}$$

Later we shall consider models where the coefficients depend not only on t but also on n so that it will be necessary to add a superscript (n) to the notations, leading to $\hat{w}^{(n)}_{t/t-1}(\beta)$, $h^{(n)}_{t/t-1}(\beta)$, and $\alpha^{(n)}_t(\beta)$ instead of, respectively, $\hat{w}_{t/t-1}(\beta)$, $h_{t/t-1}(\beta)$, and $\alpha_t(\beta)$. To simplify the presentation, we denote

$$E_{\beta^0}(\cdot(\beta)) = \{E_\beta(\cdot(\beta))\}_{\beta=\beta^0},$$

and similarly $\mathrm{var}_{\beta^0}(\cdot)$ and $\mathrm{cov}_{\beta^0}(\cdot)$. We denote $e_t(\beta) = w_t - \hat{w}_{t/t-1}(\beta)$ and $e^{(n)}_t(\beta) = w_t - \hat{w}^{(n)}_{t/t-1}(\beta)$.

In order to check the assumptions of the theorems KN1 and KN2 of Klimko and Nelson (1978), we need some additional conditions as follows.

Theorem 1

Let the stochastic process $(w_t, t \in \mathrm{N})$ be such that $E_\beta |w_t|^2 < \infty$ for all β and such that $\hat{w}_{t/t-1}(\beta)$ and $h_{t/t-1}(\beta)$ are almost surely (a.s.) twice continuously differentiable in an open subset Θ which contains the true value β^0 of vector β. Suppose there exist two positive constants C_1 and C_2 such that for all $t \geq 1$:

$$H_{1.1} \qquad E_{\beta^0}\left|\frac{\partial\alpha_t(\beta)}{\partial\beta_j}\right|^4 \leq C_1 \quad , \quad j = 1, \ldots, r \quad ; \tag{2.5}$$

$$H_{1.2} \qquad E_{\beta^0}\left|\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j} - E_\beta\left(\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}/F_{t-1}\right)\right|^2 \leq C_2 \quad i, j = 1, \ldots, r \quad . \tag{2.6}$$

Suppose further that

$$H_{1.3} \qquad \lim_{n\to\infty} \frac{1}{2n}\sum_{t=1}^{n} E_{\beta^0}\left(\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}/F_{t-1}\right) = V_{ij}(\beta^0) \quad \text{a.s.,} \quad i, j = 1, \ldots, r \quad , \tag{2.7}$$

where the matrix $V(\beta^0) = (V_{ij}(\beta^0))_{1\leq i,j\leq r}$ is a strictly definite positive matrix of constants;

$$H_{1.4} \quad \lim_{n\to\infty}\sup_{\Delta\downarrow 0}(n\Delta)^{-1}\left|\sum_{t=1}^{n}\left(\left\{\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}\right\}_{\beta=\beta^*} - \left\{\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}\right\}_{\beta=\beta^0}\right)\right| < \infty \quad \text{a.s.,} \quad i, j = 1, \ldots, r \quad , \tag{2.8}$$

where β^* is a point of the straight line joining β^0 to every β, such that $\|\beta - \beta^0\| < \Delta$.

Then,

there exists an estimator $\hat{\beta}_n$ such that $\hat{\beta}_n \to \beta^0$ almost surely, and such that for every $\varepsilon > 0$, there exists an event E in (Ω, F, P_{β^0}) with $P_{\beta^0}(E) > 1 - \varepsilon$ and n_0 such that for $n > n_0$, in E, $\partial l_n(\hat{\beta}_n)/\partial\beta_i = 0$, $i = 1, \ldots, r$, and l_n reaches a relative maximum at the point $\hat{\beta}_n$.

If these conditions are satisfied, as well as the following assumption:

$$H_{1.5} \qquad \frac{1}{n}\sum_{t=1}^{n} E_{\beta^0}\left(\frac{\partial\alpha_t(\beta)}{\partial\beta}\frac{\partial\alpha_t(\beta)}{\partial\beta^T}/F_{t-1}\right) - \frac{1}{n}\sum_{t=1}^{n} E_{\beta^0}\left(\frac{\partial\alpha_t(\beta)}{\partial\beta}\frac{\partial\alpha_t(\beta)}{\partial\beta^T}\right) \to 0 \quad \text{a.s.} \tag{2.9}$$

as $n \to \infty$, where ^{T} denotes transposition, then

$$n^{1/2}(\hat{\beta}_n - \beta^0) \xrightarrow{L} N(0, V(\beta^0)^{-1}W(\beta^0)V(\beta^0)^{-1}) \quad , \tag{2.10}$$

where L indicates convergence in law and $W(\beta^0)$ is defined by:

$$W(\beta^0) = \lim_{n\to\infty}\frac{1}{4n}\sum_{t=1}^{n} E_{\beta^0}\left(\frac{\partial\alpha_t(\beta)}{\partial\beta}\frac{\partial\alpha_t(\beta)}{\partial\beta^T}\right) \quad . \tag{2.11}$$

Remark 1

This theorem is essentially an existence theorem. It is possible to extend it in the same way as Hamdoune (1995) has done in the AR(1) case, by considering M-estimation instead of maximum likelihood estimation. It is more complex here because a term and a factor are introduced in (2.4) for the treatment of heteroscedasticity.

Theorem 1'

Under the same assumptions as in Theorem 1 except that $\hat{w}_{t/t-1}(\beta)$, $h_{t/t-1}(\beta)$, and $\alpha_t(\beta)$ are replaced by $\hat{w}^{(n)}_{t/t-1}(\beta)$, $h^{(n)}_{t/t-1}(\beta)$, and $\alpha^{(n)}_t(\beta)$, respectively, that the assumptions $H_{1.1}$ and $H_{1.2}$ are valid uniformly with respect to n, and almost sure convergence is replaced by convergence in probability in assumptions $H_{1.3}$, $H_{1.4}$ and $H_{1.5}$, then, there exists an estimator $\hat{\beta}_n$ such that $\hat{\beta}_n \rightarrow \beta^0$ in probability. The asymptotic normality result is the same as in Theorem 1.

3. Maximum likelihood estimation for ARMA models with time-dependent coefficients

In this section, we use the assumptions of Theorems 1 and 1' for ARMA models with time-dependent coefficients, based on an innovation process made up of independent random variables with zero mean and a time-dependent variance. We suppose that the autoregressive and moving average coefficients, as well as the innovation variance, are deterministic functions of time which depend on a finite number of parameters which need to be estimated.

While checking the assumptions of Theorem 1, we shall need the following result stated by Hamdoune (1995) which, as noticed by Kwoun and Yajima (1986), follows from a proof in Doob (1953, pp. 492-493, Theorem X.6.2), where it is given for a second order stationary process.

Lemma 1

Let $(w_t, t \in \mathrm{Z})$ be a process with second-order moments which are uniformly bounded with respect to t, and let $\nu > 0$ be such that

$$E\left(\frac{1}{n}\sum_{t=1}^{n} w_t\right)^2 = O\left(\frac{1}{n^{\nu}}\right) \quad .$$

Then, $n^{-1}\sum_{t=1}^{n} w_t$ converges almost surely to zero when n tends to infinity.

Definition 1

The process $(w_t, t \in \mathrm{N})$ is called an autoregressive-moving average process of order (p, q), if and only if it satisfies the equation

$$w_t = \sum_{k=1}^{p} \phi_{tk} w_{t-k} + e_t - \sum_{k=1}^{q} \theta_{tk} e_{t-k} \quad , \tag{3.1}$$

for $t \geq 1$, where the $(e_t, t \in N)$ is an independent white noise process, consisting of independent random variables, not necessarily identically distributed, with zero mean and variance $\sigma_t^2 > 0$. The coefficients ϕ_{tk} and θ_{tk} are the coefficients of the model. The initial values w_t, $t < 1$, and e_t, $t < 1$, which are used in (3.1), are supposed to be equal to zero. In the sequel, we will also use $\phi_{t0} = 1$, $\theta_{t0} = -1$, $\phi_{tk} = 0, k > p$, and $\theta_{tk} = 0, k > q$.

The contents of the vector of parameters β will be discussed soon. We define $e_t(\beta)$ or $e_t^{(n)}(\beta)$ like in Section 2. Of course, $e_t(\beta^0)$ and $e_t^{(n)}(\beta^0) = e_t$.

Thanks to the assumption about initial values and by using (3.1) recurrently, it is possible to write the pure moving average and the pure autoregressive representation of the process. For any β, the pure moving average representation is given by

$$w_t = \sum_{k=0}^{t-1} \psi_{tk}(\beta) e_{t-k}(\beta) \quad , \tag{3.2}$$

where the coefficients, the Green functions (Miller, 1968), $\psi_{tk}(\beta) = \psi_{tk}^{(k)}(\beta), k = 1, \ldots, t-1$, can be obtained from the autoregressive and moving average coefficients using the following recurrences (see Hallin and Mélard, 1977, or Hallin, 1978)

$$\psi_{tj}^{(0)}(\beta) = -\theta_{tj}(\beta), j = 1, \ldots, t-1, \qquad \psi_{t0}^{(0)}(\beta) = 1, \qquad \tilde{\psi}_{tj}^{(0)}(\beta) = \phi_{tj}(\beta), j = 1, \ldots, t-1,$$

$$\psi_{tj}^{(k)}(\beta) = \psi_{tj}^{(k-1)}(\beta) - \tilde{\psi}_{tk}^{(k-1)}(\beta)\theta_{t-k, j-k}(\beta), j = k, \ldots, t-1,$$

$$\tilde{\psi}_{tj}^{(k)}(\beta) = \tilde{\psi}_{tj}^{(k-1)}(\beta) + \tilde{\psi}_{tk}^{(k-1)}(\beta)\phi_{t-k, j-k}(\beta), j = k+1, \ldots, t-1, \quad k = 1, \ldots, t-1.$$

In a Wold-Cramér decomposition (Cramér, 1961) of the process (3.2) would be an infinite series converging in the mean and the e_t would be a weak white noise process. Alternatively, we can use a pure autoregressive representation

$$w_t = \sum_{k=1}^{t-1} \pi_{tk}(\beta) w_{t-k} + e_t(\beta) \quad , \tag{3.3}$$

where the coefficients $\pi_{tk}(\beta) = \pi_{tk}^{(k)}(\beta), k = 1, \ldots, t-1$, can be obtained using the following recurrences (see Hallin and Mélard, 1977, or Hallin, 1978)

$$\pi_{tj}^{(0)}(\beta) = \phi_{tj}(\beta), j = 1, \ldots, t-1, \qquad \tilde{\pi}_{tj}^{(0)}(\beta) = -\theta_{tj}(\beta), j = 1, \ldots, t-1,$$

$$\pi_{tj}^{(k)}(\beta) = \pi_{tj}^{(k-1)}(\beta) - \tilde{\pi}_{tk}^{(k-1)}(\beta)\phi_{t-k, j-k}(\beta), j = k, \ldots, t-1,$$

$$\tilde{\pi}_{tj}^{(k)}(\beta) = \tilde{\pi}_{tj}^{(k-1)}(\beta) + \tilde{\pi}_{tk}^{(k-1)}(\beta)\theta_{t-k, j-k}(\beta), j = k+1, \ldots, t-1, \quad k = 1, \ldots, t-1.$$

Let F_t be the σ-field spanned by $(e_u, u \le t)$. By (3.2), w_t is F_t-measurable, for all t. From (3.3), F_t is also the σ-field spanned by $(w_u, u \le t)$, for all t. If the process were not started at time $t = 1$, it should be necessary to impose a causality and an invertibility condition, for example Hallin and Ingenbleek (1983), Hallin (1986).

In the sequel, (3.1) is used to represent a time series $(w_t, t = 1, \ldots, n)$ where w_t is the observation of the variable analysed at time t, $t = 1, \ldots, n$. The time series is considered as a partial realisation of a stochastic process $(w_t, t \in \mathrm{N})$. The asymptotics are based on a triangular sequence of observations $w^{(n)} = (w_1^{(n)}, w_2^{(n)}, \ldots, w_n^{(n)})$. However, for reasons that will become clear later, the coefficients in (3.1) can also depend on n, the length of the series. In that case, we consider a sequence of processes indexed by n, based on the same innovation sequence $(e_t, t \ge 1)$. Consequently, the sequence of the σ-fields $(F_t, t \ge 1)$ is unique and does not depend on n.

The model depends on the parameters contained in each coefficient $\phi_{tk} = \phi_{tk}(\beta)$ or $\phi_{tk}^{(n)}(\beta)$, $\theta_{tk} = \theta_{tk}(\beta)$ or $\theta_{tk}^{(n)}(\beta)$, and the parameters contained in $\sigma_t^2 = \sigma^2 h_t(\beta) > 0$ or $(\sigma_t^{(n)})^2 = \sigma^2 h_t^{(n)}(\beta)$, where the scale $h_t = h_t(\beta)$ or $h_t^{(n)}(\beta)$ is a deterministic strictly positive function of time. These parameters are estimated from the realisation $w^{(n)} = (w_1^{(n)}, w_2^{(n)}, \ldots, w_n^{(n)})$. In the sequel, we often drop the superscript (n) on the observations. Rather, when the coefficients depend on n, we add a superscript (n) to them but also to all the symbols which depend on them.

Let β be a r-dimensional vector containing all the parameters to be estimated, except for σ^2. We suppose, for the commodity, that the parameters in ϕ_{tk} for $k = 1, \ldots, p$, in θ_{tk} for $k = 1, \ldots, q$, and in h_t are functionaly independent. The vector β in the parameters is thus composed of three sub-vectors ϕ, θ and δ: ϕ is the vector of parameters included in ϕ_{tk}, $k = 1, \ldots, p$, with dimension s_1 ($\beta_i = \phi_i$ for $i = 1, \ldots, s_1$), θ is the vector of parameters included in θ_{tk}, $k = 1, \ldots, q$, with dimension s_2 ($\beta_{i+s_1} = \theta_i$ for $i = 1, \ldots, s_2$), and δ is the vector of parameters included in h_t with dimension $r - s$, where $s = s_1 + s_2$, ($\beta_i = \delta_{i-s}$, $i = s+1, \ldots, r$). Denote κ_{4t} the 4-th order cumulant of the innovation e_t (which does exist and is uniformly bounded by $H_{2.5}$). It may depend on β (and even on n).

We start by giving the explicit expression of $\alpha_t = \alpha_t(\beta)$ or $\alpha_t^{(n)}(\beta)$ in the framework of the model defined by equation (3.1). The conditional expectation $\hat{w}_{t/t-1}(\beta)$ and the conditional variance $h_{t/t-1}(\beta)$ take the following forms

$$\hat{w}_{t/t-1}(\beta) = E_\beta(w_t / F_{t-1}) = \sum_{k=1}^{p} \phi_{tk}(\beta) w_{t-k} - \sum_{k=1}^{q} \theta_{tk}(\beta) e_{t-k}(\beta) \tag{3.4}$$

$$h_{t/t-1}(\beta) = E_\beta((w_t - \hat{w}_{t/t-1}(\beta))^2 / F_{t-1}) = E_\beta(e_t^2(\beta)) = \sigma^2 h_t(\beta) \quad , \tag{3.5}$$

with superscripts (n) if the coefficients depend on n. The expression (2.4) can then be written as

$$\alpha_t(\beta) = \log(\sigma^2 h_t(\beta)) + \frac{(w_t - \hat{w}_{t/t-1}(\beta))^2}{\sigma^2 h_t(\beta)} \quad \text{or} \quad \alpha_t^{(n)}(\beta) = \log(\sigma^2 h_t^{(n)}(\beta)) + \frac{(w_t - \hat{w}_{t/t-1}^{(n)}(\beta))^2}{\sigma^2 h_t^{(n)}(\beta)} \quad ,$$

accordingly.

Now we consider, with possibly superscripts (n),

$$\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i} = \sum_{k=1}^{p} \frac{\partial \phi_{tk}(\beta)}{\partial \beta_i} w_{t-k} - \sum_{k=1}^{q} \frac{\partial \theta_{tk}(\beta)}{\partial \beta_i} e_{t-k}(\beta) - \sum_{k=1}^{q} \theta_{tk}(\beta) \frac{\partial e_{t-k}(\beta)}{\partial \beta_i} \quad , \tag{3.6}$$

$i = 1, \ldots, r$, which is a recurrence equation since $e_t(\beta) = w_t - \hat{w}_{t/t-1}(\beta)$. We can obtain a simpler expression from the pure autoregressive representation (3.3)

$$\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i} = \sum_{k=1}^{t-1} \frac{\partial \pi_{tk}(\beta)}{\partial \beta_i} w_{t-k} \quad ,$$

and similarly

$$\frac{\partial^2 \hat{w}_{t/t-1}(\beta)}{\partial \beta_i \partial \beta_j} = \sum_{k=1}^{t-1} \frac{\partial^2 \pi_{tk}(\beta)}{\partial \beta_i \partial \beta_j} w_{t-k} \quad , \quad \frac{\partial^3 \hat{w}_{t/t-1}(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} = \sum_{k=1}^{t-1} \frac{\partial^3 \pi_{tk}(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} w_{t-k} \quad ,$$

for $i, j, l = 1, \ldots, r$. Alternatively, each of these expressions on the right hand side can be written as a pure moving average

$$\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i} = \sum_{k=1}^{t-1} \psi_{tik}(\beta) e_{t-k}(\beta), \tag{3.7}$$

$$\frac{\partial^2 \hat{w}_{t/t-1}(\beta)}{\partial \beta_i \partial \beta_j} = \sum_{k=1}^{t-1} \psi_{tijk}(\beta) e_{t-k}(\beta), \tag{3.8}$$

$$\frac{\partial^3 \hat{w}_{t/t-1}(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} = \sum_{k=1}^{t-1} \psi_{tijlk}(\beta) e_{t-k}(\beta), \tag{3.9}$$

for $i, j, l = 1, \ldots, r$, where the coefficients $\psi_{tik}(\beta)$, $\psi_{tijk}(\beta)$ and $\psi_{tijlk}(\beta)$ are obtained by the following relations

$$\psi_{tik}(\beta) = \sum_{u=1}^{k} \frac{\partial \pi_{tu}(\beta)}{\partial \beta_i} \psi_{t-u,k-u}(\beta) \quad , \quad \psi_{tijk}(\beta) = \sum_{u=1}^{k} \frac{\partial^2 \pi_{tu}(\beta)}{\partial \beta_i \partial \beta_j} \psi_{t-u,k-u}(\beta) \quad ,$$

$$\psi_{tijlk}(\beta) = \sum_{u=1}^{k} \frac{\partial^3 \pi_{tu}(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} \psi_{t-u,k-u}(\beta) \quad .$$

Example

Let us consider the special case of an ARMA(1,1) process defined by

$$w_t = \phi_t w_{t-1} + e_t - \theta_t e_{t-1} \quad , \tag{3.10}$$

where $\phi_t = \phi_t(\beta)$ and $\theta_t = \theta_t(\beta)$. We shall use the pure moving average representation of the process which has coefficients (to be used in examples 1 to 3 of Section 4)

$$\psi_{tk}(\beta) = \left\{ \prod_{l=0}^{k-2} \phi_{t-l}(\beta) \right\} \{\phi_{t-k+1}(\beta) - \theta_{t-k+1}(\beta)\} \quad ,$$

$k = 1, 2, \ldots, t-1$, where a product for $l = 0$ to -1 is set to one. Similarly, the coefficients of the pure autoregressive form are as follows

$$\pi_{tk}(\beta) = \left\{ \prod_{l=0}^{k-2} \theta_{t-l}(\beta) \right\} \{\phi_{t-k+1}(\beta) - \theta_{t-k+1}(\beta)\} \quad ,$$

so that their derivatives are given by

$$\frac{\partial \pi_{t1}(\beta)}{\partial \beta} = \frac{\partial \phi_t(\beta)}{\partial \beta} - \frac{\partial \theta_t(\beta)}{\partial \beta}$$

$$\frac{\partial \pi_{t2}(\beta)}{\partial \beta} = \frac{\partial \theta_t(\beta)}{\partial \beta} \{\phi_{t-1}(\beta) - \theta_{t-1}(\beta)\} + \theta_t(\beta) \left\{ \frac{\partial \phi_{t-1}(\beta)}{\partial \beta} - \frac{\partial \theta_{t-1}(\beta)}{\partial \beta} \right\}$$

$$\frac{\partial \pi_{t3}(\beta)}{\partial \beta} = \frac{\partial \theta_t(\beta)}{\partial \beta} \theta_{t-1}(\beta) \{\phi_{t-2}(\beta) - \theta_{t-2}(\beta)\} + \theta_t(\beta) \frac{\partial \theta_{t-1}(\beta)}{\partial \beta} \{\phi_{t-2}(\beta) - \theta_{t-2}(\beta)\} + \theta_t(\beta)\theta_{t-1}(\beta) \left\{ \frac{\partial \phi_{t-2}(\beta)}{\partial \beta} - \frac{\partial \theta_{t-2}(\beta)}{\partial \beta} \right\}$$

…

Hence it can be shown that

$$\psi_{tik}(\beta) = \sum_{l=1}^{k} \left(\prod_{h=1}^{k} \chi_{t-h,k,l,h,i}(\beta) \right) \quad , \tag{3.11}$$

where

$$\chi_{tklhi}(\beta) = \begin{cases} \dfrac{\partial \chi_{tklh}(\beta)}{\partial \beta_i} & \text{if } h = l, \\ \chi_{tklh}(\beta) & \text{if } h \neq l, \end{cases} \quad \text{and} \quad \chi_{tklh}(\beta) = \begin{cases} \phi_t(\beta) & \text{if } l \le h < k, \\ \theta_t(\beta) & \text{if } h < l \le k, \\ \phi_t(\beta) - \theta_t(\beta) & \text{if } h = k. \end{cases}$$

We first address the case where the coefficients depend on t but not on n.

Theorem 2

Consider an autoregressive-moving average process defined by (3.1) and suppose that the functions $\phi_{tk}(\beta)$, $\theta_{tk}(\beta)$ and $h_t(\beta)$ are three times continuously differentiable with respect to β, in the open set B containing the true value β^0 of β, that there exist positive constants $\Phi < 1$, $N_1, N_2, N_3, N_4, N_5, K_1, K_2, K_3, m, M$, m_1 and K such that $\forall t$:

$H_{2.1}$ $\sum_{k=\nu}^{t-1} \psi_{tik}^2(\beta^0) < N_1 \Phi^{\nu-1}$, $\sum_{k=\nu}^{t-1} \psi_{tik}^4(\beta^0) < N_2 \Phi^{\nu-1}$, $\sum_{k=\nu}^{t-1} \psi_{tijk}^2(\beta^0) < N_3 \Phi^{\nu-1}$,

$\sum_{k=\nu}^{t-1} \psi_{tijk}^4(\beta^0) < N_4 \Phi^{\nu-1}$, $\sum_{k=1}^{t-1} \psi_{tijlk}^2(\beta^0) < N_5$, $\nu = 1, \ldots, t-1, \quad i, j, l = 1, \ldots, r, \quad t = 1, \ldots, n;$

$H_{2.2}$ $\left| \left\{ \frac{\partial h_t(\beta)}{\partial \beta_i} \right\}_{\beta=\beta^0} \right| \le K_1$, $\left| \left\{ \frac{\partial^2 h_t(\beta)}{\partial \beta_i \partial \beta_j} \right\}_{\beta=\beta^0} \right| \le K_2$, $\left| \left\{ \frac{\partial^3 h_t(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l} \right\}_{\beta=\beta^0} \right| \le K_3 \quad i, j, l = 1, \ldots, r$;

$H_{2.3}$ $0 < m \le h_t(\beta^0) \le m_1$;

$H_{2.4}$ $E(w_t^4) \le M$;

$H_{2.5}$ $(\sigma^8 h_t^4(\beta^0))^{-1} E(e_t^8) \le K$.

Suppose furthermore that

$H_{2.6}$ $$\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \left[\sigma^{-2} E_{\beta^0} \left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i} h_t^{-1}(\beta) \frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j} \right) + \frac{1}{2} \left\{ \frac{\partial h_t(\beta)}{\partial \beta_i} \right\}_{\beta=\beta^0} h_t^{-2}(\beta^0) \left\{ \frac{\partial h_t(\beta)}{\partial \beta_j} \right\}_{\beta=\beta^0} \right] = V_{ij}(\beta^0) ,$$

$i, j = 1, \ldots, r$, where the matrix $V(\beta^0) = (V_{ij}(\beta^0))_{1 \le i,j \le r}$ is a strictly definite positive matrix;

$H_{2.7}$ $$\frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k=1}^{t-1-d} \psi_{tik}(\beta^0) \psi_{tjk}(\beta^0) \psi_{t+d,i,k+d}(\beta^0) \psi_{t+d,j,k+d}(\beta^0) \kappa_{4,t-k} = O\left(\frac{1}{n}\right),$$

$$\frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k=1}^{t-d} \psi_{tik}(\beta^0) \psi_{t+d,i,k+d}(\beta^0) h_{t-k}(\beta^0) = O\left(\frac{1}{n}\right) .$$

Then,

- there exists an estimator $\hat{\beta}_n$ such that $\hat{\beta}_n \to \beta^0$ almost surely and in the sense of Theorem 1;
- $n^{1/2}(\hat{\beta}_n - \beta^0) \xrightarrow{L} N(0, V(\beta^0)^{-1} W(\beta^0) V(\beta^0)^{-1})$ where there exists a matrix $W(\beta^0)$ whose elements are defined by (2.11).

Remark 2

$H_{2.1}$ is satisfied if there exists $\Phi < 1$ such that $\psi_{tik}^2(\beta^0) < \Phi^k$ and similar conditions for the others. $H_{2.7}$ is also verified in that case. $H_{2.7}$ is a sort of mixing condition which will take a more classical form in the case of a pure autoregression (see Azrak and Mélard, 2004). $H_{2.6}$ is probably the most intriguing assumption. It means that each parameter is asymptotically informative either for the coefficients of the ARMA model or for the innovation standard deviation. In the stationary case, the assumption implies that there is no common root in the autoregressive and moving average polynomials. Things are of course much more complex here as will be shown in Example 5 in Section 4. The assumption on existence of 8th-order moments of the innovations can be reduced to the existence of moments of order $4 + \gamma$, where $\gamma > 0$ but this requires strengthening $H_{1.1}$.

Remark 3

As is well known, for a non-Gaussian process, neither $V(\beta^0)^{-1}$, nor $W(\beta^0)^{-1}$, is the asymptotic covariance matrix, see Whittle (1982) (obtained there for independent observations) and Tjøstheim (1986) but well the so-called sandwich estimator $V(\beta^0)^{-1} W(\beta^0) V(\beta^0)^{-1}$.

Remark 4

With the following theorem we shall see that the dependence of the model with respect to n through the coefficients $\phi_{tk} = \phi_{tk}^{(n)}(\beta)$, $\theta_{tk} = \theta_{tk}^{(n)}(\beta)$ and $h_t = h_t^{(n)}(\beta)$ has no substantial effect on the conclusions except that almost sure convergence is replaced by convergence in probability. For convenience, we state the lemma which corresponds to Lemma 1 in the case of triangular arrays of random variables. The proof is immediate since convergence in L_2 norm implies convergence in probability.

Lemma 1'

Let $(w_t^{(n)}, t = 1, \ldots, n)$ be, for each $n \in$ N, a process with second-order moments, and let $\nu > 0$ be such that

$$E\left(\frac{1}{n}\sum_{t=1}^{n} w_t^{(n)}\right)^2 = O\left(\frac{1}{n^\nu}\right) \quad .$$

Then, $n^{-1}\sum_{t=1}^{n} w_t^{(n)}$ converges in probability to zero when n tends to infinity.

Theorem 2'

Consider a sequence of autoregressive-moving average processes based on the same innovation process, defined by (3.1) and indexed by (n), $n \in$ N, and suppose that the functions $\phi_{tk}^{(n)}(\beta)$, $\theta_{tk}^{(n)}(\beta)$ and $h_t^{(n)}(\beta)$ are three times continuously differentiable with respect to β, in the open set B containing the true value β^0 of β, that there exist positive constants $\Phi < 1$, $N_1, N_2, N_3, N_4, N_5, K_1, K_2, K_3, m, M$, m_1 and K, such that $\forall t$ and uniformly with respect to n:

$H_{2'.1}$ $\quad \sum_{k=\nu}^{t-1} \psi_{tik}^{(n)2}(\beta^0) < N_1 \Phi^{\nu-1} \quad , \quad \sum_{k=\nu}^{t-1} \psi_{tik}^{(n)4}(\beta^0) < N_2 \Phi^{\nu-1} \quad , \quad \sum_{k=\nu}^{t-1} \psi_{tijk}^{(n)2}(\beta^0) < N_3 \Phi^{\nu-1} \quad ,$

$\quad \sum_{k=\nu}^{t-1} \psi_{tijk}^{(n)4}(\beta^0) < N_4 \Phi^{\nu-1} \quad , \quad \sum_{k=1}^{t-1} \psi_{tijlk}^{(n)2}(\beta^0) < N_5 \quad , \quad \nu = 1, \ldots, t-1, \quad i, j, l = 1, \ldots, r, \quad t = 1, \ldots, n;$

$H_{2'.2}$ $\quad \left|\left\{\frac{\partial h_t^{(n)}(\beta)}{\partial \beta_i}\right\}_{\beta=\beta^0}\right| \le K_1 \quad , \left|\left\{\frac{\partial^2 h_t^{(n)}(\beta)}{\partial \beta_i \partial \beta_j}\right\}_{\beta=\beta^0}\right| \le K_2 \quad , \left|\left\{\frac{\partial^3 h_t^{(n)}(\beta)}{\partial \beta_i \partial \beta_j \partial \beta_l}\right\}_{\beta=\beta^0}\right| \le K_3 \quad i, j, l = 1, \ldots, r \quad ;$

$H_{2'.3}$ $\quad 0 < m \le h_t^{(n)}(\beta^0) \le m_1 \quad ;$

$H_{2'.4}$ $\quad E(w_t^{(n)4}) \le M \quad ;$

$H_{2'.5}$ $\quad (\sigma^8 h_t^{(n)4}(\beta^0))^{-1} E(e_t^8) \le K \quad .$

Suppose furthermore that

$H_{2'.6}$ $\quad \lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \left[\sigma^{-2} E_{\beta^0}\left(\frac{\partial \hat{w}_{t/t-1}^{(n)}(\beta)}{\partial \beta_i} \{h_t^{(n)}(\beta)\}^{-1} \frac{\partial \hat{w}_{t/t-1}^{(n)}(\beta)}{\partial \beta_j}\right) + \frac{1}{2}\left\{\frac{\partial h_t^{(n)}(\beta)}{\partial \beta_i}\right\}_{\beta=\beta^0} \{h_t^{(n)}(\beta^0)\}^{-2} \left\{\frac{\partial h_t^{(n)}(\beta)}{\partial \beta_j}\right\}_{\beta=\beta^0}\right] = V_{ij}(\beta^0) \quad ,$

$i, j = 1, \ldots, r$, where the matrix $V(\beta^0) = (V_{ij}(\beta^0))_{1 \le i, j \le r}$ is a strictly definite positive matrix;

$H_{2'.7}$ $$\frac{1}{n^2}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\sum_{k=1}^{t-1-d}\psi_{tik}^{(n)}(\beta^0)\psi_{tjk}^{(n)}(\beta^0)\psi_{t+d,i,k+d}^{(n)}(\beta^0)\psi_{t+d,j,k+d}^{(n)}(\beta^0)\kappa_{4,t-k} = O\left(\frac{1}{n}\right),$$

$$\frac{1}{n^2}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\sum_{k=1}^{t-d}\psi_{tik}^{(n)}(\beta^0)\psi_{t+d,i,k+d}^{(n)}(\beta^0)h_{t-k}^{(n)}(\beta^0) = O\left(\frac{1}{n}\right) \quad .$$

Then,

- there exists an estimator $\hat{\beta}_n$ such that $\hat{\beta}_n \to \beta^0$ in probability;
- $n^{1/2}(\hat{\beta}_n - \beta^0) \xrightarrow{L} N(0, V(\beta^0)^{-1}W(\beta^0)V(\beta^0)^{-1})$ where there exists a matrix $W(\beta^0)$ whose elements are defined by (2.11) with $\alpha_t(\beta)$ replaced by $\alpha_t^{(n)}(\beta)$.

Remark 5

It is not obvious that almost sure convergence can be obtained in Theorem 2' using our approach. Indeed, besides the mixtingale array problem mentioned in the proof of Theorem 2, it relies also on Theorem 1' where a strong law of large numbers for martingale difference arrays doesn't seem to exist, and on Lemma 1 for which the proof also makes use of sequence arguments. This contrasts to the approach of Dahlhaus (1997) where almost sure convergence is obtained. Note however that the assumptions of Theorem 2' are weaker: we have no assumption of continuity with respect to time, the possibility of coefficients periodically varying with respect to time t, no condition on the roots of the autoregressive and moving average polynomials considered at each time.

4. Examples

In this section, we consider five examples. The first three of them concern AR(1) processes with a time-dependent coefficient. In the first example, the coefficient is a periodic function of time. In the second example, the coefficient is an exponential function of time. The innovation variance is constant in both cases. In the third example and in part of the simulation results of Section 5, we have a linear function of time for the coefficient and an exponential function of time for the scale factor. The last two examples are for a MA(1) and an ARMA (1,1) processes. The former has a linear function of time for the coefficient and an exponential function of time for the scale factor and is also illustrated in the simulation results of Section 5. The latter is mainly used to investigate the need for $H_{2.6}$.

Example 1

Let us consider the process defined by

$$w_t = \phi_t w_{t-1} + e_t \quad , \tag{4.1}$$

where the coefficient $\phi_t = \beta_{t-s\lfloor t/s\rfloor}$, $s > 1$ is an integer and $\lfloor x\rfloor$ is the largest integer less or equal to x, the innovations have possibly a time-dependent variance which does not depend on parameters and finite 8th-order moment. It is assumed that the product $\beta^* = \beta_0\beta_1\ldots\beta_{s-1} < 1$. Otherwise, (4.1) will not be causal which means it will not have a purely nondeterministic solution. Note that some of the factors can be greater than 1, contrarily to the assumptions of Dahlhaus (1996a). Suppose there are s parameters, let us say $\beta_0, \beta_1, \ldots, \beta_{s-1}$. We now check the assumptions $H_{2.4}$, $H_{2.6}$ and $H_{2.7}$ of Theorem 2, since the other ones are trivially true.

By specializing the results of the example of Section 3, we have

$$\psi_{tk}(\beta) = \prod_{l=0}^{k-1}\phi_{t-l,1}(\beta) \quad , \qquad \psi_{tik}(\beta) = \frac{\partial\phi_{t1}(\beta)}{\partial\beta_i}\prod_{l=1}^{k-1}\phi_{t-l,1}(\beta) \quad .$$

The moving average representation of the process is, for large t,

$$w_t = \sum_{l=0}^{\infty}\beta^{*\lfloor l/s\rfloor}\left(\prod_{j=0}^{l-1-s\lfloor l/s\rfloor}\beta_{t-j-s\lfloor (t-j)/s\rfloor}\right)e_{t-l} \quad .$$

Consequently, assuming that the innovation variance is a constant σ^2:

$$\mathrm{var}(w_t) = \frac{\sigma^2}{1-\beta^{*2}} \sum_{l=0}^{s-1} \left(\prod_{j=0}^{l-1-s\lfloor l/s \rfloor} \beta_{t-j-s\lfloor (t-j)/s \rfloor} \right)^2 \quad . \tag{4.2}$$

An expression for the 4th-order moment can be obtained in the same way. These expressions would be more complex in the marginally heteroscedastic case. Note that the derivatives of $\hat{w}_{t/t-1}(\beta)$ with respect to the parameters are either w_{t-1} or 0, and (4.2) implies that the limit in $H_{2.6}$ exists.

Let us now consider the case where h_t is constant. It is well known (Tiao and Grupe, 1980) that a process with periodic coefficients of period s can be embedded into an s-dimensional stationary autoregressive process. By using Pham and Tran (1985) result, under mild assumptions, the process is therefore strong mixing, at an exponential rate. The conclusions of Azrak and Mélard (2004) apply but, of course, the result is already known for a multivariate stationary process. The result is however new if the innovations have a bounded time-dependent variance. This example is not compatible with the time rescaling approach of Dahlhaus (1997) because of the fixed periodicity.

Example 2

Let us consider again (4.1), where the innovations have a constant variance σ^2 and 8th-order moments, but now $\phi_t = \phi_t^{(n)}(\beta) = \gamma\beta^{t/n}$ where $0 < \gamma < 1$ is a fixed constant. It is assumed that the only parameter of the model is β and that its true value is β^0, $0 < \beta^0 < 1$. The coefficient $\phi_t^{(n)}(\beta)$ is decreasing with t and varies between $\gamma(\beta^0)^{1/n}$ and $\gamma\beta^0$. We shall use the moving average representation of the process which is

$$w_t^{(n)} = \sum_{k=0}^{\infty} \gamma^k \left(\prod_{l=0}^{k-1} \beta^{(t-l)/n} \right) e_{t-k} \quad .$$

Consequently, $E\{(w_t^{(n)})^2\}/\sigma^2$ has an upper bound equal to the sum of a geometric series with rate $\gamma^2\beta^{2/n} < 1$ and a lower bound 1. The existence of $E\left(w_t^{(n)4}\right)$ follows similarly. The expression in $H_{2'.6}$ has the form

$$\lim_{n\to\infty} \frac{\gamma^2}{n\sigma^2} \sum_{t=1}^{n} \left(\frac{t}{n}\right)^2 \beta^{2\left(\frac{t}{n}-1\right)} E(w_{t-1}^{(n)})^2$$

evaluated at $\beta = \beta^0$. There suffices to show convergence of

$$\lim_{n\to\infty} \frac{\gamma^2}{n} \sum_{t=1}^{n} \left(\frac{t}{n}\right)^2 \beta_0^{2\left(\frac{t}{n}-1\right)} = \gamma^2 \lim_{n\to\infty} \frac{1}{n^3} \frac{\beta_0^{2/n-2}(1+\beta_0^{2/n})}{(1-\beta_0^{2/n})^3} = \frac{\gamma^2}{2\beta_0^2(1-\beta_0)^3} \quad .$$

Using Theorem 2', we can try replacing $\psi_{t1k}^{(n)}(\beta)$ by an upper bound γ^k. For example, the second part of $H_{2'.7}$ can be written

$$\frac{\sigma^2}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k_1=1}^{n-t} \pi_{t1k_1}^{(n)}(\beta^0)\pi_{t+d,1,k_1+d}^{(n)}(\beta^0) \le \frac{\sigma^2}{n^2} \sum_{t=1}^{n-2} \sum_{d=1}^{n-t} \sum_{k_1=1}^{n-t} \gamma^{2k_1+d} \le \frac{\sigma^2}{n(1-\gamma^2)(1-\gamma)} \tag{4.3}$$

which is $O(1/n)$. Similarly the first part of $H_{2.7}$ can be replaced by an upper bound using (A1.15) which is a triple sum such as (4.3) but where γ^{2k_1+d} is replaced by γ^{4k_1+2d}. Note that $\partial\phi_t^{(n)}(\beta)/\partial\beta = (t/n)\gamma\beta^{t/n-1}$ and

$$\psi_{t1k}^{(n)}(\beta) = \frac{t}{n}\gamma^k\beta^{t/n-1} \prod_{l=1}^{k-1} \beta^{(t-l)/n} \quad ,$$

decrease exponentially with k, so that $H_{2'.1}$ is verified. The conclusion is similar hence Theorem 2' applies.

In order to conclude Example 2, let us mention that the use of time-dependent coefficients such as $\phi_t^{(n)}(\beta)$, which depend on the length of the series is compatible with the approach of Dahlhaus (1997) mentioned in Section 1.

Example 3

For simplicity, we keep an AR(1) process but with a coefficient which is a linear function of time and the innovation standard deviation which is an exponential function of time, and again innovations with bounded 8-th order moments. More specifically, we suppose that

$$\phi_t^{(n)}(\beta) = \phi' + \frac{1}{n-1}\left(t - \frac{n+1}{2}\right)\phi'' \quad , \qquad h_t^{(n)}(\beta) = \exp\left\{2\frac{\delta}{n-1}\left(t - \frac{n+1}{2}\right)\right\} \quad . \tag{4.4}$$

There will be conditions on the parameters $\beta = (\phi', \phi'', \delta)^T$ which will be discussed below. Note that the scale factor $\{h_t^{(n)}(\beta)\}^{1/2}$ is such that

$$\prod_{t=1}^{n} \{h_t^{(n)}(\beta)\}^{1/2} = e^{\frac{\delta}{n-1}\sum_{t=1}^{n}\left(t - \frac{n+1}{2}\right)} = 1$$

and it varies between $e^{-\delta}$ and e^{δ}, fulfilling $H_{2'.3}$. Let us denote by $\phi^*(\beta)$ the upper bound of $|\phi_t(\beta)|$, $t = 1, \ldots, n$. Then the moving average representation implies that

$$E\left(\{w_t^{(n)}\}^4\right) = \sum_{k=0}^{\infty}\left(\prod_{l=0}^{k-1}\phi_{t-l,1}^{(n)}\right)^4 E(e_{t-k}^4)$$

$$\leq \sum_{k=0}^{\infty}\{\phi^*(\beta)\}^{4k}\left\{\kappa_{4,t-k}^{(n)} + 3\sigma^4\{h_{t-k}^{(n)}(\beta)\}^2\right\}$$

$$= \sum_{k=0}^{\infty}\{\phi^*(\beta)\}^{4k}\kappa_{4,t-k} + 3\sigma^4 e^{\frac{4\delta}{n-1}\left(t - \frac{n+1}{2}\right)}\sum_{k=0}^{\infty}\{\phi^*(\beta)\}^{4k} e^{\frac{-4\delta}{n-1}k} \quad .$$

Hence, if the process were Gaussian, a sufficient condition for the existence of a uniform bound in $H_{2'.4}$, for fixed n, is that $\phi^*(\beta^0)\exp(-\delta^0/(n-1)) < 1$, where δ^0 is the value of δ if $\beta = \beta^0$. The uniform bounds of $H_{2'.1}$ with respect to n are satisfied if $\phi^*(\beta^0) < 1$. This is not a necessary condition. Contrarily to the approach of locally stationary processes of Dahlhaus (1997), there would be no problem, assuming a more complex form for $|\phi_t^{(n)}(\beta^0)|$ than (4.4), if it would be larger than 1 during a finite span of time. Similarly

$$E\left(\{w_t^{(n)}\}^2\right) = \sum_{k=0}^{\infty}\left\{\prod_{l=0}^{k-1}\phi_{t-l}^{(n)}(\beta)\right\}^2 E(e_{t-k}^2) = \sigma^2\sum_{k=0}^{\infty}\left\{\prod_{l=0}^{k-1}\phi_{t-l}^{(n)}(\beta)\right\}^2 h_{t-k}^{(n)}(\beta)$$

$$\leq \sigma^2 e^{\frac{2\delta}{n-1}\left(t - \frac{n+1}{2}\right)}\sum_{k=0}^{\infty}\{\phi^*(\beta)\}^{2k} e^{\frac{-2\delta}{n-1}k} \quad .$$

We have the 3×1 vectors

$$\frac{\partial \hat{w}_{t/t-1}^{(n)}(\beta)}{\partial \beta} = \left(1, \frac{1}{n-1}\left(t - \frac{n+1}{2}\right), 0\right)^T w_{t-1}^{(n)} \quad , \qquad \frac{\partial h_t^{(n)}(\beta)}{\partial \beta} = \left(0, 0, \frac{2h_t^{(n)}(\beta)}{n-1}\left(t - \frac{n+1}{2}\right)\right)^T \quad .$$

The expression in $H_{2'.6}$ takes the form

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial\hat{w}_{t/t-1}^{(n)}(\beta)}{\partial\beta}\left(\frac{\partial\hat{w}_{t/t-1}^{(n)}(\beta)}{\partial\beta}\right)^T\frac{1}{\sigma^2h_t^{(n)}(\beta)}+\frac{1}{2}\frac{\partial h_t^{(n)}(\beta)}{\partial\beta}\left(\frac{\partial h_t^{(n)}(\beta)}{\partial\beta}\right)^T\frac{1}{(h_t^{(n)}(\beta))^2}\right\}_{\beta=\beta^0}$$

$$=\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\left\{S_{tn}\frac{E\left[(w_{t-1}^{(n)})^2\right]}{\sigma^2h_t^{(n)}(\beta^0)}+\frac{1}{2}T_{tn}\right\}\quad,\tag{4.5}$$

where

$$S_{tn}=\begin{pmatrix}1 & t_n & 0\\ t_n & t_n^2 & 0\\ 0 & 0 & 0\end{pmatrix}\quad,\quad T_{tn}=\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & 4t_n^2\end{pmatrix}\quad.\tag{4.6}$$

with $t_n=\{t-(n+1)/2\}/(n-1)$. For example, if $\beta^0=(\phi^0,0,\delta^0)^T$, we have

$$E_{\beta^0}\left[(w_t^{(n)})^2\right]=\sigma^2e^{2\delta^0t_n}\sum_{k=0}^{\infty}(\phi^0)^{2k}e^{\frac{-2\delta^0}{n-1}k}=\frac{\sigma^2h_t^{(n)}(\beta^0)}{1-(\phi^0)^2e^{-2\delta^0/(n-1)}}\quad,$$

hence (4.5) evaluated at $\beta=\beta^0$ is

$$\begin{pmatrix}\left\{1-(\phi^0)^2e^{-\delta^0/(n-1)}\right\}^{-1} & 0 & 0\\ 0 & \left[12\left\{1-(\phi^0)^2e^{-\delta^0/(n-1)}\right\}\right]^{-1} & 0\\ 0 & 0 & 1/6\end{pmatrix}\quad.\tag{4.7}$$

There remains to check the assumptions of Theorem 2’. Here

$$\pi_{t1k}^{(n)}(\beta)=\prod_{l=1}^{k-1}\phi_{t-l}^{(n)}(\beta)\quad,\quad\pi_{t2k}^{(n)}(\beta)=\frac{1}{n-1}\left(t-\frac{n+1}{2}\right)\prod_{l=1}^{k-1}\phi_{t-l}^{(n)}(\beta)\quad,\quad\pi_{t3k}^{(n)}(\beta)=0\quad,$$

which have an upper bound $(\phi^*(\beta))^k$. Hence the second part of $H_{2'.7}$ can be written, for $i,j=1,2,3$,

$$\frac{\sigma^2}{n^2}\sum_{d=2}^{n-1}\sum_{t=1}^{n-d}\sum_{k_1=1}^{n-t}\pi_{tik_1}^{(n)}(\beta^0)\pi_{t+d,j,k_1+d}^{(n)}(\beta^0)h_{t-k_1}^{(n)}(\beta^0)\le\frac{\sigma^2}{n^2}\sum_{t=1}^{n-2}e^{2\delta^0t_n}\sum_{d=1}^{n-t}\sum_{k_1=1}^{n-t}\phi^*(\beta^0)^{2k_1+d}e^{\frac{-2\delta^0}{n-1}k_1}$$

$$=\sigma^2\frac{e^{-\delta^0}}{2n\delta^0(1-(\phi^*(\beta^0))^2)(1-\phi^*(\beta^0))}+O\left(\frac{1}{n^2}\right)$$

which is $O(1/n)$ and similarly for the first part of $H_{2.7}$. The result of Theorem 2’ is again valid.

Let us compute $W(\beta^0)^{-1}$ in order to obtain the asymptotic covariance matrix of $n^{1/2}(\hat{\beta}_n-\beta^0)$. Taking again the case where $\beta^0=(\phi^0,0,\delta^0)^T$, the element (3, 3) of (3.12), for example, is obtained as follows

$$\lim_{n\to\infty}\frac{1}{4n}\sum_{t=1}^{n}E_{\beta^0}\left(\frac{\partial\alpha_t^{(n)}(\beta)}{\partial\beta_3}\right)^2=\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\frac{1}{n(n-1)^2}\left(t-\frac{n+1}{2}\right)^2E_{\beta^0}\left\{1-\frac{(e_t^{(n)}(\beta))^2}{\sigma^2h_t^{(n)}(\beta)}\right\}^2$$

$$=\frac{1}{6}+\lim_{n\to\infty}\frac{1}{n(n-1)^2\sigma^4}\sum_{t=1}^{n}\left(t-\frac{n+1}{2}\right)^2\frac{\kappa_{4,t}^{(n)}}{(h_t^{(n)}(\beta^0))^2}\quad,\tag{4.8}$$

which reduces to 1/6 if the process is Gaussian. Let us suppose this. Then $V(\beta^0)^{-1}=W(\beta^0)^{-1}$ and, consequently, the asymptotic covariance matrix of $n^{1/2}(\hat{\beta}_n-\beta^0)$ is

$$V(\beta^0)^{-1}W(\beta^0)V(\beta^0)^{-1} = \begin{pmatrix} 1-(\phi^0)^2 e^{-2\delta^0/(n-1)} & 0 & 0 \\ 0 & 12\left\{1-(\phi^0)^2 e^{-2\delta^0/(n-1)}\right\} & 0 \\ 0 & 0 & 6 \end{pmatrix} \quad . \tag{4.9}$$

Example 4

This is a pure moving average process defined by

$$w_t^{(n)} = e_t - \theta_t^{(n)} e_{t-1} \quad , \tag{4.10}$$

with a coefficient which is a linear function of time and the innovation standard deviation which is an exponential function of time, and again innovations with bounded 8-th order moments. More specifically, we suppose that

$$\theta_t^{(n)}(\beta) = \theta' + \frac{1}{n-1}\left(t - \frac{n+1}{2}\right)\theta'' \quad , \qquad h_t^{(n)}(\beta) = \exp\left\{2\frac{\delta}{n-1}\left(t - \frac{n+1}{2}\right)\right\} \quad . \tag{4.11}$$

There will be conditions on the parameters $\beta = (\theta', \theta'', \delta)^T$ which will be discussed later. The scale factor varies like in Example 3. Let us denote by $\theta^*(\beta)$ the upper bound of $|\theta_t^{(n)}(\beta)|$, $t = 1, \ldots, n$. There is no condition for the existence of a uniform bound in $H_{2.4}$. Here

$$\psi_{t1k}^{(n)}(\beta) = \prod_{l=1}^{k-1} \theta_{t-l}^{(n)}(\beta) \quad , \quad \psi_{t2k}^{(n)}(\beta) = \frac{1}{n-1}\left(t - \frac{n+1}{2}\right)\prod_{l=1}^{k-1} \theta_{t-l}^{(n)}(\beta) \quad , \quad \psi_{t3k}^{(n)}(\beta) = 0 \quad ,$$

which have an upper bound $(\theta^*(\beta))^k$. The expression in $H_{2'.6}$ takes the form

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n}\left\{S_{tn}.\frac{E\left[e_{t-1}^2 + \{\theta_t^{(n)}(\beta^0)\}^2 e_{t-2}^2 + \{\theta_t^{(n)}(\beta^0)\}^2\{\theta_{t-1}^{(n)}(\beta^0)\}^2 e_{t-2}^2 + \ldots\right]}{\sigma^2 h_t^{(n)}(\beta^0)} + \frac{1}{2}T_{tn}\right\} \quad , \tag{4.12}$$

using again the notations (4.6).

For example, if $\beta^0 = (\theta^0, 0, \delta^0)^T$, we have

$$E_{\beta^0}\left(\{w_t^{(n)}\}^2\right) = \sigma^2 e^{2\delta^0 t_n}\sum_{k=0}^{\infty}(\theta^0)^{2k} e^{\frac{-2\delta^0}{n-1}k} = \frac{\sigma^2 h_t^{(n)}(\beta^0)}{1-(\theta^0)^2 e^{-2\delta^0/(n-1)}} \quad ,$$

with again $t_n = \{t-(n+1)/2\}/(n-1)$. Hence (4.12) evaluated at $\beta = \beta^0$ is

$$\begin{pmatrix} \left\{1-(\theta^0)^2 e^{-2\delta^0/(n-1)}\right\}^{-1} & 0 & 0 \\ 0 & \left[12\left\{1-(\theta^0)^2 e^{-2\delta^0/(n-1)}\right\}\right]^{-1} & 0 \\ 0 & 0 & 1/6 \end{pmatrix} \quad . \tag{4.13}$$

There remains to check the other assumptions of Theorem 2'. The second condition of $H_{2.7}$ can be written, for $i, j = 1, 2, 3$,

$$\frac{\sigma^2}{n^2}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\sum_{k_1=1}^{n-t}\psi_{tik_1}^{(n)}(\beta^0)\psi_{t+d,j,k_1+d}^{(n)}(\beta^0)h_{t-k_1}^{(n)}(\beta^0) \le \frac{\sigma^2}{n^2}\sum_{t=1}^{n-2} e^{2\delta^0 t_n}\sum_{d=1}^{n-t}\sum_{k_1=1}^{n-t}\theta^*(\beta^0)^{2k_1+d} e^{\frac{-2\delta^0}{n-1}k_1}$$

$$= \sigma^2 \frac{e^{-\delta^0}}{2n\delta^0(1-(\theta^*(\beta^0))^2)(1-\theta^*(\beta^0))} + O\left(\frac{1}{n^2}\right)$$

which is $O(1/n)$ and similarly for the first condition of $H_{2'.7}$. The result of Theorem 2' is again valid.

Let us compute $W(\beta^0)^{-1}$ in order to obtain the asymptotic covariance matrix of $n^{1/2}(\hat{\beta}_n - \beta^0)$. Taking again the case where $\beta^0 = (\theta^0, 0, \delta^0)^T$ and assuming that the process is Gaussian, we obtain (4.12). Consequently, in that case, the asymptotic covariance matrix of $n^{1/2}(\hat{\beta}_n - \beta^0)$ is

$$V(\beta^0)^{-1} W(\beta^0) V(\beta^0)^{-1} = \begin{pmatrix} 1-(\theta^0)^2 e^{-2\delta^0/(n-1)} & 0 & 0 \\ 0 & 12\left\{1-(\theta^0)^2 e^{-2\delta^0/(n-1)}\right\} & 0 \\ 0 & 0 & 6 \end{pmatrix} , \qquad (4.14)$$

which is similar to what was obtained for Example 3.

Example 5

We consider a special case of the ARMA(1,1) process considered as an example in Section 3, where the innovations have a constant variance σ^2 and 8th-order moments. We suppose that $\phi_t = \phi_t^{(n)}(\beta) = \phi\lambda^{t/n}$ and $\theta_t = \theta_t^{(n)}(\beta) = \theta\lambda^{t/n}$, where $0 < \lambda < 1$ is a fixed constant, and the two parameters of the model are $\beta = (\beta_1, \beta_2) = (\phi, \theta)$. It is assumed that its true value is $\beta^0 = (\phi^0, \theta^0)$, such that $-1 < \phi^0 < 1$ and $-1 < \theta^0 < 1$. We shall later assume that $\phi^0 \neq \theta^0$ in order to fulfill $H_{2'.6}$. The assumptions $H_{2'.1}$ are clearly satisfied.

Note that $\partial\phi_{t1}^{(n)}/\partial\phi = (\partial\theta_{t1}^{(n)}(\beta))/(\partial\theta) = \lambda^{t/n}$. Then

$$\frac{\partial \hat{w}_{t/t-1}^{(n)}(\beta)}{\partial\phi} = \lambda^{t/n} e_{t-1}^{(n)}(\beta) + \phi\lambda^{t/n}\lambda^{(t-1)/n} e_{t-2}^{(n)}(\beta) + \phi^2\lambda^{t/n}\lambda^{(t-1)/n}\lambda^{(t-2)/n} e_{t-3}^{(n)}(\beta) + \ldots$$

$$\frac{\partial \hat{w}_{t/t-1}^{(n)}(\beta)}{\partial\theta} = -\lambda^{t/n} e_{t-1}^{(n)}(\beta) - \theta\lambda^{t/n}\lambda^{(t-1)/n} e_{t-2}^{(n)}(\beta) - \theta^2\lambda^{t/n}\lambda^{(t-1)/n}\lambda^{(t-2)/n} e_{t-3}^{(n)}(\beta) - \ldots$$

Using (3.11), the expression in $H_{2'.6}$ has the form

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \sum_{k=1}^{t-1} \left[\left\{ \sum_{l=1}^{k} \left(\prod_{h=1}^{k} \chi_{t-h,k,l,h,i}^{(n)} \right) \right\} \left\{ \sum_{l=1}^{k} \left(\prod_{h=1}^{k} \chi_{t-h,k,l,h,j}^{(n)} \right) \right\} \right]$$

to be evaluated at $\beta = \beta^0$ which gives a 2×2 matrix

$$\frac{1}{n} \sum_{t=1}^{n} \sum_{k=1}^{t-1} \left(\prod_{h=1}^{k} \lambda^{(t-h)/n} \right)^2 \begin{pmatrix} (\phi^0)^{2(k-1)} & -(\phi^0)^{k-1}(\theta^0)^{k-1} \\ -(\phi^0)^{k-1}(\theta^0)^{k-1} & (\theta^0)^{2(k-1)} \end{pmatrix} , \qquad (4.15)$$

generally of rank 2 but not if $\phi^0 = \theta^0$. The other conditions are easy to check by proceeding like above.

5. Some simulation results

We consider a limited Monte Carlo study to show that the theory stated above works with finite series. We are interested in the 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. We are also interested in comparing various estimates of the standard errors. Three formulas are used for the evaluation of the asymptotic covariance matrix of the estimator: (a) $V(\beta^0)^{-1}$ which should be correct for a Gaussian process ; (b) $V(\beta^0)^{-1}W(\beta^0)V(\beta^0)^{-1}$; and (c) the standard numerical estimate (called Marquardt default expression in the sequel) based on the sample average of $\{\partial e_t(\beta)/\partial\beta\}\,\{\partial e_t(\beta)/\partial\beta^T\}$. Doing this implies forgetting the term $e_t(\beta)\,\{\partial^2 e_t(\beta)/\partial\beta\partial\beta^T\}$ whose expectation is equal to zero for an ARMA model but not when there is a parameter in $h_t(\beta)$.

The estimation method which is used in the examples of this section has been implemented in Time Series Expert (Mélard and Pasteels, 1994): the exact Gaussian likelihood is computed by the algorithm of Mélard (1982) and the optimum is obtained using a customised variant of Marquardt's (1963) non-linear least-squares optimisation procedure. The algorithm given independently by Dahlhaus (1996a) and Azrak and Mélard (1998), which provides a generalisation of Gardner *et al.* (1980), combined with a good optimisation procedure could have been used instead. Contrarily to Dahlhaus (1997) assertions, the exact maximum likelihood method is not computationally intensive, not very much than for ARMA models with constant coefficients. Indeed, although the number of operations at each time is quadratic with respect to the model orders, the experiments made by Mélard (1982) and Azrak and Mélard (1998) show that the computation times remain reasonable. For example, if the Ansley (1979) Cholesky factorization algorithm is used, the number of operations at each time is barely multiplied by 2, plus the operations needed for computing the coefficients in terms of the parameters, of course. A run of 1000 simulations taken from Table 1, for series of length 400, does not take more than 4 minutes on a computer with a 500 MHz Intel Celeron processor.

Alternatively to the exact maximum likelihood method, the conditional maximum likelihood method can be considered for fitting the models. But, Monte Carlo experiments (e.g. Ansley and Newbold, 1980) have shown that, for stationary MA and ARMA models (not for AR models) and relatively short series of length 50 or 100, exact maximum likelihood estimation is 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 $t \leq m$. It seems therefore plausible that conditional maximum likelihood suffers similarly when dealing with more general time-dependent or non-linear models. Our purpose is also to show that the unconditional quasi-maximum likelihood method is indeed superior to the conditional quasi-maximum likelihood method, at least for MA models.

We consider two sets of simulations. They will address respectively AR(1) and MA(1) processes, but marginally heteroscedastic with an innovation standard deviation which is an exponential function of time. We have generated realizations of these processes and, for each of the series, we have fitted a model with the right specification except that the coefficient is time dependent instead of being constant. Our simulations provide statistical evidence for the test of stability of the coefficient and also compare the empirical results with the asymptotic expressions of Section 4, respectively Examples 3 and 4.

An AR(1) process with a time dependent coefficient, defined by (4.1) with the same specification as in Example 3, has been simulated, with $\phi' = 0.5$, $\phi'' = 0$ and $\delta \neq 0$, and using a normal or double exponential distribution for the e_t. The length n of the series varies from 25 to 400. Note that $\delta/(n-1)$ is used as a parameter in (4.4) with the same δ for each n. This is a way to implement the condition $H_{2'.3}$ of Theorem 2'. The same stream of sequences of 400 uniform pseudo-random numbers has been used to simulate the innovations for the two distributions, using the inverse distribution function procedure. A number of 1000 series of length 400 have been generated using a method which doesn't require warming. For series of length $n < 400$, the first n observations have been used. The purpose is to reduce randomness among the experiments. The results are given in Tables 1 and 2 in the case of a normal distribution, and in Tables 1 and 3 for the double exponential distribution. Note that for short n (n = 25, 50 and even 100), the entire estimation procedure has not been carried on successfully for a few simulated series. This is due to the fact that the estimate of V obtained by finite second order derivatives of the log-likelihood is not always positive definite. Results denoted by [(a)] and [(b)] in Tables 1-3 are based on the remaining series.

As shown in Table 1, for the three parameters, the bias tends to zero when n increases. Also, there is not much difference when the innovation distribution is compatible with the law used in the quasi-maximum likelihood, i.e. normal[(a)], or when it is double exponential[(b)]. In Table 2, as far as standard errors are concerned, $V^{-1(a)}$ gives satisfactory result in the normal case, as expected. The empirical standard deviations of the estimates agree with the theoretical values deduced from (4.9). Note, however, that the standard errors provided by the Marquardt optimisation procedure[(c)] are bad for δ, a fact that has been conjectured by Mélard (1985). The number of replications for which the

Table 1. Estimated parameter for ϕ' (true value: 0.5), ϕ'' (true value: 0), and $\delta/(n-1)$ (true value as shown) for the AR(1) model with [(a)]*normal* or [(b)]*double exponential* innovation distribution, obtained by quasi-maximum likelihood methods, for n = 25, 50, 100, 200, 400; 1000 replications, except [(a)]819, [(b)] 802 for n = 25, and [(a)]980, [(b)] 962 for n = 50.

Length n	25	50	100	200	400
Distribution ϕ' = 0.5					
normal[(a)]	0.451	0.476	0.489	0.497	0.498
double exponential[(b)]	0.455	0.475	0.489	0.497	0.498
Distribution ϕ'' = 0.0					
normal[(a)]	-0.00048	-0.00150	-0.00032	-0.00005	-0.00004
double exponential[(b)]	0.00030	-0.00115	-0.00037	-0.00004	-0.00004
Distribution true $\delta/(n-1)$	0.0480	0.0240	0.0120	0.00600	0.00300
normal[(a)]	0.0408	0.0228	0.0116	0.00593	0.00299
double exponential[(b)]	0.0384	0.0226	0.0116	0.00595	0.00301

Marquardt optimisation procedure[(c)] gives a satisfactory covariance matrix is larger than with the more correct procedure[(a)]. The averages of the estimates for case [(c)] are not shown here but are very close to those of Table 1 except for n = 25. For example, 0,460 is obtained for ϕ' instead of 0,451. When n = 25, the averages of the estimated standard errors are often unreliable so that the average has no meaning and have been replaced by dashes. This has however never occurred in our simulations for n = 50 or larger.

Table 2. Theoretical, empirical and estimated standard errors for the AR(1) model with *normal* innovation distribution obtained by the unconditional quasi-maximum likelihood methods, for n = 25, 50, 100, 200, 400; 1000 replications, except [(a)]819, [(c)]951 for n = 25 and [(a)]980, [(c)]991 for n = 50. Standard errors are computed [(a)]using V^{-1}; [(c)]using Marquardt default; a —— indicates an unreliable result.

Length n	25	50	100	200	400
ϕ' = 0.5					
theoretical	0.173	0.106	0.0866	0.0603	0.0433
empirical[(a)]	0.188	0.129	0.0887	0.0603	0.0421
estimated[(a)](avg)	0.217	0.135	0.0904	0.0625	0.0437
estimated[(c)](avg)	——	0.130	0.0895	0.0622	0.0437
ϕ'' = 0.0					
theoretical	0.0240	0.00734	0.00300	0.00106	0.000375
empirical[(a)]	0.0275	0.00956	0.00314	0.00110	0.000381
estimated[(a)](avg)	0.0347	0.01008	0.00325	0.00110	0.000383
estimated[(c)](avg)	——	0.00917	0.00315	0.00109	0.000380
true $\delta/(n-1)$	0.048	0.024	0.012	0.006	0.003
theoretical	0.0196	0.00693	0.00245	0.000866	0.000306
empirical[(a)]	0.0229	0.00763	0.00256	0.000871	0.000311
estimated[(a)] (avg)	0.0239	0.00745	0.00254	0.000882	0.000309
estimated[(c)] (avg)	——	0.01044	0.00357	0.001245	0.000436

With double exponential innovations, we didn't try to evaluate the theoretical standard errors because of the difficulty to compute (4.8). As shown in Table 3, using $V^{-1(d)}$ seems to agree with empirical standard deviations except for $\delta/(n-1)$, and this whatever the sample size. In the latter case, as expected, $V^{-1}WV^{-1}$ proves to be necessary for evaluating the standard errors.

Table 3. Empirical and estimated standard deviation for the AR(1) model with *double exponential* innovation distribution obtained by the unconditional quasi-maximum likelihood method, for n = 25, 50, 100, 200, 400; 1000 replications, except $^{(b)}$802, $^{(d)}$815 for n = 25, $^{(b)(d)}$962 for n = 50, and $^{(d)}$998 for n = 100. Standard errors are computed $^{(b)}$using $V^{-1}WV^{-1}$; $^{(d)}$using V^{-1}; a —— indicates an unreliable result.

Length n	25	50	100	200	400
ϕ' = 0.5					
empirical$^{(b)}$	0.182	0.127	0.0877	0.0606	0.0423
estimated$^{(b)}$ (avg)	——	——	0.0858	0.0608	0.0430
estimated$^{(d)}$ (avg)	0.211	0.135	0.0904	0.0625	0.0438
ϕ'' = 0.0					
empirical$^{(b)}$	0.0260	0.00937	0.00310	0.00109	0.000378
estimated$^{(b)}$ (avg)	——	——	0.00305	0.00106	0.000376
estimated$^{(d)}$ (avg)	0.0333	0.01016	0.00327	0.00111	0.000384
true $\delta/(n-1)$	0.048	0.024	0.012	0.006	0.003
empirical$^{(b)}$	0.0320	0.0112	0.00381	0.00134	0.000491
estimated$^{(b)}$ (avg)	——	——	0.00364	0.00136	0.000498
estimated$^{(d)}$ (avg)	0.0235	0.0075	0.00256	0.00089	0.000310

The second simulation experiment is based on a MA(1) process with marginally heteroscedastic innovations defined by (4.10) and (4.11), using θ' = 0.9, θ'' = 0, δ as in Table 4, and a normal or double exponential distribution for the e_t. The length of the series varies from 25 to 400. A number of 1000 series have been generated. The results are given in Table 4 and Table 6 in the case of a normal distribution, in Table 5 and Table 7, in the case of a double exponential distribution. In Table 4 and Table 5, it can be seen that 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. Tables 6 and 7 are concerned with the standard errors for the moving average series produced with normal and double exponential innovations, respectively. They are provided here only when the maximum quasi-likelihood method is used. The following are compared: the theoretical value (only in the normal case), the Monte Carlo standard deviation, and the results of V^{-1}, $V^{-1}WV^{-1}$ as well as the standard errors provided by the Marquardt optimisation procedure. The results are similar to those for the AR(1) model. In Table 6, the empirical values largely agree with the theoretical values deduced from (4.14) in the normal case.

Table 4. Estimated parameters for the MA(1) model with *normal* innovation distribution obtained by the conditional or unconditional quasi-maximum likelihood methods, n = 25, 50, 100, 200, 400; 1000 replications, except [(a)]996, [(b)]954 for n = 25, [(a)]997, [(b)]979 for n = 50, and [(a)]999, [(b)]998 for n = 100.

Length n	25	50	100	200	400
Method $\theta' = 0.5$					
conditional[(a)]	0.552	0.515	0.509	0.500	0.501
unconditional[(b)]	0.516	0.513	0.510	0.501	0.501
Method $\theta'' = 0.0$					
conditional[(a)]	0.00937	0.00089	0.00011	-0.00003	0.00003
unconditional[(b)]	0.00978	0.00119	0.00007	-0.00006	0.00002
Method true $\delta/(n-1)$	0.048	0.0240	0.0120	0.0060	0.0030
conditional[(a)]	0.040	0.0233	0.0118	0.0060	0.0030
unconditional[(b)]	0.043	0.0236	0.0119	0.0060	0.0030

Table 5. Estimated parameters for the MA(1) model with *double exponential* innovation distribution obtained by the conditional or unconditional quasi-maximum likelihood methods, n = 25, 50, 100, 200, 400; 1000 replications except [(a)]929, [(b)]781 for n = 25, [(a)]991, [(b)]912 for n = 50, and [(a)]998 [(b)]983 for n = 100.

Length n	25	50	100	200	400
Method $\theta' = 0.5$					
conditional[(a)]	0.501	0.514	0.508	0.500	0.501
unconditional[(b)]	0.496	0.503	0.508	0.501	0.501
Method $\theta'' = 0.0$					
conditional[(a)]	0.00434	0.00058	0.00012	-0.00003	0.00003
unconditional[(b)]	0.01175	0.00182	0.00016	-0.00006	0.00002
Method true $\delta/(n-1)$	0.048	0.0240	0.0120	0.0060	0.0030
conditional[(a)]	0.042	0.0233	0.0118	0.0060	0.0030
unconditional[(b)]	0.041	0.0235	0.0118	0.0060	0.0030

Table 6. Theoretical, empirical and estimated standard errors for the MA(1) model with *normal* innovation distribution obtained by the unconditional quasi-maximum likelihood methods, for n = 25, 50, 100, 200, 400; 999 or 1000 replications, except [(b)]760, [(c)]954 for n = 25, [(b)]898, [(c)]979 for n = 50, and [(b)]984, [(c)]998 for n = 100. Standard errors are computed [(b)]using V^{-1}; [(c)]using Marquardt default; a —— indicates an unreliable result.

Length n	25	50	100	200	400
θ’ = 0.5					
theoretical	0.173	0.106	0.0866	0.0603	0.0433
empirical[(b)]	0.241	0.156	0.0962	0.0650	0.0437
estimated[(b)](avg)	0.216	0.139	0.0926	0.0636	0.0441
estimated[(c)](avg)	0.210	0.127	0.0881	0.0622	0.0437
θ’’ = 0.0					
theoretical	0.0240	0.00734	0.00300	0.00106	0.000375
empirical[(b)]	0.0418	0.01201	0.00362	0.00119	0.000402
estimated[(b)](avg)	0.0338	0.01049	0.00341	0.00114	0.000389
estimated[(c)](avg)	0.0359	0.00998	0.00320	0.00110	0.000382
true $\delta/(n-1)$	0.048	0.024	0.012	0.006	0.003
theoretical	0.0196	0.00693	0.00245	0.000866	0.000306
empirical[(b)]	0.0252	0.00791	0.00259	0.000871	0.000312
estimated[(b)] (avg)	0.0237	0.00751	0.00255	0.000883	0.000309
estimated[(c)] (avg)	0.0325	0.01051	0.00358	0.001245	0.000436

Table 7. Empirical and estimated standard errors for the MA(1) model with *double exponential* innovation distribution obtained by the unconditional quasi-maximum likelihood method, for n = 25, 50, 100, 200, 400; 1000 replications, except [(b)]783, [(c)]781 for n = 25 and [(b)(c)]912 for n = 50, [(b)(c)]983 for n = 100. Standard errors are computed [(b)]using $V^{-1}WV^{-1}$; [(c)]using V^{-1}; a —— indicates an unreliable result.

Length n	25	50	100	200	400
θ’ = 0.5					
empirical[(b)]	0.226	0.147	0.0937	0.0649	0.0439
estimated[(b)] (avg)	——	——	0.0940	0.0638	0.0442
estimated[(c)] (avg)	0.217	0.139	0.0925	0.0635	0.0441
θ’’ = 0.0					
empirical[(b)]	0.0343	0.01076	0.00341	0.00119	0.000397
estimated[(b)] (avg)	——	——	0.00342	0.00114	0.000390
estimated[(c)] (avg)	0.0344	0.01053	0.00341	0.00114	0.000390
true $\delta/(n-1)$	0.048	0.024	0.012	0.006	0.003
empirical[(b)]	0.0320	0.0112	0.00382	0.00135	0.000491
estimated[(b)] (avg)	——	——	0.00364	0.00136	0.000499
estimated[(c)] (avg)	0.0242	0.0076	0.00257	0.00089	0.000310

The conclusions of these experiments are
- the results of the asymptotic theory can be expected to work even on short series;
- for MA processes, exact pseudo-maximum likelihood is better than using a conditional least-squares approach;

- standard errors provided by the Marquardt optimization procedure are reliable except for short series or when the parameter is used to measure heteroscedasticity;

- standard errors provided by the sandwich estimator $V^{-1}WV^{-1}$ are of course recommended in non normal situations but may be difficult to obtain for short series.

6. Practical time series

We start with three renowned examples from Box and Jenkins (1976). The first one, based on the so-called Series A (measurements of the concentration of a chemical process once every two hours, n = 197), is aimed at illustrating the use of an ARMA model with a time dependent coefficient. For Series A, Box and Jenkins (1976) have proposed an ARIMA(0,1,1) model with equation

$$\nabla w_t = e_t - \theta e_{t-1} \quad . \tag{6.1}$$

We modify that model by assuming that the moving average coefficient varies slightly, writing

$$\nabla w_t = e_t - \theta_t e_{t-1} \quad , \tag{6.2}$$

where $\theta_t = \theta' + t\theta''$ is a smooth function of time. To facilitate non-linear estimation, it has been seen that the following parametrisation is better:

$$\theta_t = \theta' + (t-(n+1)/2)\theta'' \tag{6.3}$$

The results shown in Table 8 reveal that the Student statistic for $\hat{\theta}''$ (denoted TDMA1) is equal to -2.5, thus the hypothesis of a model with a time-invariant model is rejected at the 5% level.

Table 8. Series A. Estimation results for the model (6.1-3) where θ' is denoted by "MA1" and $\hat{\theta}''$, by "TDMA1".

```
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
      (ALGORITHM FOR TIME DEPENDENT ARMA MODELS)
=== MODEL DESCRIPTION                    FORM      DEGREE/ORD PARAMETERS NUMBER
- DIFFERENCE                           REGULAR             1
- ADDITIVE CONSTANT                    AUTOMATIC
- ARMA MODEL
      MOVING AVERAGE POLYNOMIAL        REGULAR             1      MA  nn       1
- MA TIME DEPENDENT COEFFICIENTS       LINEAR              1      MATDnn       1
      PIVOTAL TIME FOR ARTD/MATD               99.0
FINAL VALUES OF THE PARAMETERS                        WITH 95% CONFIDENCE LIMITS
             NAME      VALUE         STD ERROR   T-VALUE  LOWER          UPPER
     1      MA   1    .60155       5.87912E-02    10.2     .49          .72
     2      TDMA 1  -2.35026E-03   9.27707E-04    -2.5   -4.17E-03   -5.32E-04
=== SUMMARY MEASURES
TOTAL  NUMBER  OF  PARAMETERS =  3  STANDARD DEVIATION =  .317066
INFORMATION CRITERIA :    AIC =  112.656          SBIC =  125.855
```

The second example is Series B (closing IBM stock prices, $n = 395$) which will show the treatment of marginal heteroscedasticity (although it can also be shown for illustrating conditional heteroscedasticity, as for many stock prices series, e. g. Bollerslev, 1986). When the model (6.1) is fitted, $\hat{\theta} = -0.085$ is obtained. Inspection of the residuals reveals that the volatility is higher at the end of the estimation period. Furthermore, the fit is not very good. For example, the Ljung-Box test statistic with 36 lags has a probability of significance equal to 0.022. A much better model is obtained by using (6.1-4.11) with ∇w_t instead of w_t:

$$\nabla w_t = e_t - \theta e_{t-1} \quad . \tag{6.4}$$

As can be seen from Table 9, the estimate $\hat{\delta}$ is significantly different from zero at any usual level and the fit is better, with a probability of significance of 0.370 for the Ljung-Box test. Furthermore, the Schwarz Bayesian information criterion SBIC = 2489 is much lower than for the previous model, SBIC = 2524.

Table 9. Series B. Estimation results for the model (6.1-4.11) where θ is denoted by "MA1" and δ, by "VART1".

```
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
      (ALGORITHM FOR TIME DEPENDENT ARMA MODELS)
=== MODEL DESCRIPTION                  FORM      DEGREE/ORD PARAMETERS NUMBER
- DIFFERENCE                         REGULAR             1
- ADDITIVE CONSTANT                  AUTOMATIC
- ARMA MODEL
      MOVING AVERAGE POLYNOMIAL      REGULAR             1     MA  nn      1
- TIME DEPENDENT INNOVATION STD DEV. EXPONENTIAL               VART 1      1
FINAL VALUES OF THE PARAMETERS                    WITH 95% CONFIDENCE LIMITS
             NAME      VALUE         STD ERROR  T-VALUE  LOWER          UPPER
     1      MA   1   -.14529       5.21067E-02    -2.8    -.25       -4.32E-02
     2      VART 1   2.36966E-03   3.57617E-04     6.6    1.67E-03    3.07E-03
=== SUMMARY MEASURES
TOTAL  NUMBER  OF  PARAMETERS =  3  STANDARD DEVIATION =  6.86460
INFORMATION CRITERIA :    AIC =  2473.84         SBIC =  2489.53
=== RESIDUAL ANALYSIS WITH 368 RESIDUALS, BEGINNING AT TIME       2===
LJUNG-BOX PORTMANTEAU TEST STATISTICS ON RESIDUAL AUTOCORRELATIONS
 ORDER D.F. STATISTIC  SIGNIFICANCE
 48     47   49.60        .370
```

The third example is Series G from Box and Jenkins (1976), the airline passengers data. It is well known that the series should be treated in logarithms (multiplied here by the geometric mean). The usual model has the following form

$$\nabla\nabla_{12}\log(w_t) = (1-\theta B)(1-\Theta B^{12})e_t \quad , \tag{6.5}$$

where B is the lag operator. We have replaced the constant innovation standard deviation by an exponential function of time (4.4) with δ instead of $\delta/(n-1)$. The results of the fitted model are displayed in Table 10. Note that the estimate $\hat{\delta}$ is again significant. Table 11 shows a similar model but with (4.4) replaced by the linear function

$$h_t = 1 + \delta(t-(n+1)/2) \quad . \tag{6.6}$$

It is clear that the results are close to those in Table 10. There is a trend in the innovation standard deviation but small enough so that a linear or an exponential specification are alike.

Table 10. Series G. Estimation results for the model (6.5-4.4) where θ, Θ, and δ are denoted by "MA1", "SMA1", and "VART1", respectively.

```
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
      (ALGORITHM FOR TIME DEPENDENT ARMA MODELS)
=== MODEL DESCRIPTION                  FORM      DEGREE/ORD PARAMETERS NUMBER
- SEASONAL PERIOD                              12
- NORMALIZED BOX COX TRANSFORMATION  LOGARITHMS                BOXC 1      0
- DIFFERENCE                         REGULAR             1
- DIFFERENCE                         SEASONAL            1
- ARMA MODEL
      MOVING AVERAGE POLYNOMIAL      REGULAR             1     MA  nn      1
      MOVING AVERAGE POLYNOMIAL      SEASONAL            1     SMA nn      1
- TIME DEPENDENT INNOVATION STD DEV. EXPONENTIAL               VART 1      1
FINAL VALUES OF THE PARAMETERS                    WITH 95% CONFIDENCE LIMITS
             NAME      VALUE         STD ERROR  T-VALUE  LOWER          UPPER
     1      MA   1    .31340       9.18528E-02     3.4     .13         .50
     2      SMA  1    .50222       7.77547E-02     6.5     .35         .66
     3      VART 1  -6.45416E-03   1.94479E-03    -3.3   -1.03E-02   -2.57E-03
=== SUMMARY MEASURES  <V>
TOTAL  NUMBER  OF  PARAMETERS =  3  STANDARD DEVIATION =  8.51820
INFORMATION CRITERIA :    AIC =  945.645         SBIC =  958.436
```

Table 11. Series G. Estimation results for the model (6.5-6.6) where θ, Θ, and δ are denoted by "MA1", "SMA1", and "VART1", respectively.

```
=== ESTIMATION BY MAXIMIZATION OF THE EXACT (LOG)LIKELIHOOD
      (ALGORITHM FOR TIME DEPENDENT ARMA MODELS)
=== MODEL DESCRIPTION                   FORM      DEGREE/ORD PARAMETERS NUMBER
- SEASONAL PERIOD                                12
- NORMALIZED BOX COX TRANSFORMATION  LOGARITHMS                 BOXC 1      0
- DIFFERENCE                         REGULAR             1
- DIFFERENCE                         SEASONAL            1
- ARMA MODEL
      MOVING AVERAGE POLYNOMIAL      REGULAR             1      MA  nn      1
      MOVING AVERAGE POLYNOMIAL      SEASONAL            1      SMA nn      1
- TIME DEPENDENT INNOVATION STD DEV. LINEAR                     VART 1      1
FINAL VALUES OF THE PARAMETERS                    WITH 95% CONFIDENCE LIMITS
             NAME      VALUE         STD ERROR   T-VALUE  LOWER         UPPER
     1      MA   1    .31646       9.18827E-02     3.4      .13          .50
     2      SMA  1    .50787       7.68019E-02     6.6      .35          .66
     3      VART 1  -5.87685E-03   1.66481E-03    -3.5   -9.20E-03   -2.55E-03
=== SUMMARY MEASURES  <V>
TOTAL  NUMBER  OF  PARAMETERS =  3  STANDARD DEVIATION =  8.70774
INFORMATION CRITERIA :    AIC =  951.455          SBIC =  964.246
```

7. Conclusions

We have shown that under suitable conditions, quasi-maximum likelihood estimators of a large class of marginally heteroscedastic ARMA models with time-dependent coefficients do exist, converge almost surely or in probability, and are asymptotically normally distributed with a known covariance matrix. The conditions are partly similar to those of Dahlhaus (1997) but the class of models seems more general and the asymptotics are apparently different.

The empirical results, albeit partial, show that the approach can be used in practice even for series of moderate length. Note that results of that nature have never appeared in the literature, although the subject has been studied for a long time.

In principle, the approach can be extended to more general estimators, such as M-estimators, and to more general models, including multivariate models and non-linear models. It is possible to add a mean to the model. As indicated by Azrak and Mélard (1993), the simplest way is to replace (3.1) by

$$w_t - \mu_t = \sum_{k=1}^{p} \phi_{tk}(w_{t-k} - \mu_{t-k}) + e_t - \sum_{k=1}^{q} \theta_{tk} e_{t-k} \quad ,$$

where μ_t is a deterministic function of time (and perhaps on the length of the series) and depends on a finite number of parameters. Under assumptions of uniform boundedness and a Noether condition, similar results can be obtained. For all these models, there remains to work on specification procedures because it is not feasible to introduce too many parameters.

Appendix 1

The following are the proofs of all the theorems stated in the paper.

Proof of Theorem 1

We check the assumptions of KN1 for $Q_n(\beta) = -l_n(\beta)$. For the third assumption, we have to show that

$$\frac{1}{n}\left\{\frac{\partial l_n(\beta)}{\partial \beta_i}\right\}_{\beta=\beta^0} \to 0 \quad \text{a.s.} \quad , \quad i = 1, \ldots, r. \tag{A1.1}$$

Note that $\{\partial\alpha_t(\beta)/\partial\beta_i, F_t\}$ is a martingale difference sequence since it belongs to F_t and is integrable and

$$\frac{\partial\alpha_t(\beta)}{\partial\beta_i} = \frac{1}{h_{t/t-1}(\beta)}\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_i} - \frac{2(w_t - \hat{w}_{t/t-1}(\beta))}{h_{t/t-1}(\beta)}\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial\beta_i} - \frac{(w_t - \hat{w}_{t/t-1}(\beta))^2}{(h_{t/t-1}(\beta))^2}\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_i} \quad . \qquad \text{(A1.2)}$$

In order to invoke the strong law of large numbers for martingales (Stout, 1974, p. 154, Theorem 3.3.8), we need to suppose that the martingale difference sequence $\{\partial\alpha_t(\beta)/\partial\beta_i, F_t\}$ satisfies the condition

$$\sum_{t=1}^{\infty}\frac{E_{\beta^0}|\partial\alpha_t(\beta)/\partial\beta_i|^p}{t^{1+p/2}} < \infty$$

for $p \geq 2$. But, taking $p = 4$, $\beta = \beta^0$ and using $H_{1.1}$, we have

$$\sum_{t=1}^{\infty}\frac{E_{\beta^0}|\partial\alpha_t(\beta)/\partial\beta_i|^4}{t^3} \leq C_1\sum_{t=1}^{\infty}\frac{1}{t^3} < \infty.$$

We conclude that (A1.1) holds.

To check the second assumption of KN1, we begin by evaluating the second derivative of the function α_t defined by (2.4):

$$\begin{aligned}\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j} =& h_{t/t-1}^{-1}(\beta)\frac{\partial^2 h_{t/t-1}(\beta)}{\partial\beta_i\partial\beta_j} - h_{t/t-1}^{-2}(\beta)\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_i}\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_j} - 2\frac{\partial^2\hat{w}_{t/t-1}(\beta)}{\partial\beta_i\partial\beta_j}h_{t/t-1}^{-1}(\beta)(w_t - \hat{w}_{t/t-1}(\beta))\\ &+ 2\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_i}h_{t/t-1}^{-1}(\beta)\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_j} + 2\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_i}(w_t - \hat{w}_{t/t-1}(\beta))h_{t/t-1}^{-2}(\beta)\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_j}\\ &+ 2\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_j}h_{t/t-1}^{-2}(\beta)\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_i}(w_t - \hat{w}_{t/t-1}(\beta))\\ &+2(w_t - \hat{w}_{t/t-1}(\beta))^2 h_{t/t-1}^{-3}(\beta)\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_i}\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_j} - (w_t - \hat{w}_{t/t-1}(\beta))^2 h_{t/t-1}^{-2}(\beta)\frac{\partial^2 h_{t/t-1}(\beta)}{\partial\beta_i\partial\beta_j}. \qquad \text{(A1.3)}\end{aligned}$$

It is clear that

$$\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j} - E_\beta\left(\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}/F_{t-1}\right)$$

defines a martingale difference sequence. Using again the strong law of large numbers for martingales (Stout, 1974), given $H_{1.2}$, with $p = 2$,

$$\sum_{t=1}^{\infty}\frac{E_{\beta^0}\left|\partial^2\alpha_t(\beta)/\partial\beta_i\partial\beta_j - E_{\beta^0}(\partial^2\alpha_t(\beta)/\partial\beta_i\partial\beta_j/F_{t-1})\right|^2}{t^2} < \sum_{t=1}^{\infty}\frac{C_2}{t^2} < \infty \quad ,$$

we can deduce the following convergence

$$\frac{1}{n}\sum_{t=1}^{n}\left\{\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}\right\}_{\beta=\beta^0} - \frac{1}{n}\sum_{t=1}^{n}E_{\beta^0}\left(\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}/F_{t-1}\right) \to 0 \quad \text{a.s.}$$

But, from $H_{1.3}$

$$\lim_{n\to\infty}\frac{1}{2n}\sum_{t=1}^{n}E_{\beta^0}\left(\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}/F_{t-1}\right) = V_{ij}(\beta^0)$$

where $V(\beta^0) = (V_{ij}(\beta^0))_{1 \le i,j \le r}$ is a strictly definite positive matrix of constants. The second assumption of KN1 is thus satisfied. Since we have taken $Q_n(\beta) = -l_n(w_1, \ldots, w_n;\beta)$, $H_{1.4}$ and the first condition of KN1 are identical. The three assumptions of that theorem are thus proved. Consequently, there exists $\hat{\beta}_n$ such that $\hat{\beta}_n \to \beta^0$ in the sense given in the statement of the theorem.

To prove asymptotic normality of the estimator $\hat{\beta}_n$, there suffices to prove that the first assumption of KN2 is satisfied. We shall make use of a central limit theorem for martingales (Basawa et Prakasa Rao, 1980, p. 388). Let us first show that

$$n^{-1/2}\left\{\frac{\partial l_n(\beta)}{\partial \beta}\right\}_{\beta=\beta^0} \xrightarrow{L} N(0,W) \quad . \tag{A1.4}$$

For that part of the proof, we use the Cramér-Wold device. Let λ be a constant vector of dimension $r \times 1$ and ξ_t a random variable defined by $\xi_t = \lambda^T\{\partial\alpha_t(\beta)/\partial\beta\}_{\beta=\beta^0}$. Our purpose is to show asymptotic normality of $\sum_{t=1}^{n} \xi_t$ for all λ. We start by checking the assumptions of the theorem of Basawa and Prakasa Rao (1980) for the variable ξ_t. We have already shown that $E_{\beta^0}(\xi_t/F_{t-1}) = 0$.

Let us now check that there exists $\Delta_1 > 0$ and $M(\lambda)$, such that $E|\xi_t|^{2+\Delta_1} \le M(\lambda) < \infty, \forall t \ge 1$. For $\Delta_1 = 2$,

$$E_{\beta^0}(|\xi_t|^4) = E_{\beta^0}\left(\left|\sum_{i=1}^{r}\lambda_i \frac{\partial\alpha_t(\beta)}{\partial\beta_i}\right|^4\right) \le r\|\lambda\|^4 \sum_{i=1}^{r} E_{\beta^0}\left(\frac{\partial\alpha_t(\beta)}{\partial\beta_i}\right)^4 \le r\|\lambda\|^4 C_1 < \infty \quad ,$$

using Cauchy inequality and $H_{1.1}$, where $\|.\|$ is the Euclidian norm such that $\|\lambda\| = (\lambda^T\lambda)^{1/2}$. Hence, the condition is satisfied. There remains to prove that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} E_{\beta^0}(\xi_t^2/F_{t-1}) = \gamma^2(\lambda) \quad , \tag{A1.5}$$

where $\gamma^2(\lambda) \ge 0$. From $H_{1.5}$, the left hand side of (A1.5) is equal to

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} E_{\beta^0}\left(\lambda^T \frac{\partial\alpha_t(\beta)}{\partial\beta}\frac{\partial\alpha_t(\beta)}{\partial\beta^T}\lambda/F_{t-1}\right) = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} E_{\beta^0}\left(\lambda^T \frac{\partial\alpha_t(\beta)}{\partial\beta}\frac{\partial\alpha_t(\beta)}{\partial\beta^T}\lambda\right) = 4\lambda^T W\lambda \ge 0 \quad .$$

We take that expression as $\gamma^2(\lambda)$. All the assumptions of the theorem of Basawa and Prakasa Rao (1980) are then verified. Consequently, we have

$$n^{-1/2}\sum_{t=1}^{n}\xi_t \xrightarrow{L} N(0,\gamma^2(\lambda)) \quad \text{or} \quad n^{-1/2}\lambda^T\left\{\frac{\partial l_n(\beta)}{\partial\beta}\right\}_{\beta=\beta^0} \xrightarrow{L} \mathrm{N}(0,\lambda^T W\lambda) \quad ,$$

and since it is true for all λ, we deduce (A1.4). To conclude, Theorem KN2 leads to Theorem 1.

Proof of Theorem 1'

We use a weak version of theorems KN1 and KN2 of Klimko and Nelson (1978) (e.g. Bar Shalom, 1971, Bhat, 1974), where almost sure convergence is replaced everywhere by convergence in probability. We simply indicate the modifications to the proof of Theorem 1.

Since the coefficients can now depend on n, when increasing n, all the preceeding terms in the log-likelihood are changed. Therefore, the log-likelihood is not a martingale. However $\{\partial\alpha_t^{(n)}(\beta)/\partial\beta_i, F_t\}$ is a martingale difference array in the sense that $\{\partial\alpha_t^{(n)}(\beta)/\partial\beta_i, F_t, t = 1, \ldots, n\}$ is a martingale difference sequence for each n. Instead of a strong law of large numbers for martingales, we use a weak law of large numbers for a martingale difference array (Chow, 1971, Davidson, 1994, p. 299, Theorem 19.7 with $k_n = n$). More specifically, we take $\xi_{nt} = \{\partial\alpha_t^{(n)}(\beta)/\partial\beta_i\}_{\beta=\beta^0}$ and want to show that $n^{-1}\{\partial l_n(\beta)/\partial\beta_i\}_{\beta=\beta^0} = n^{-1}\sum_{t=1}^{n}\xi_{nt} \overset{\pi}{\rightarrow} 0$ as $n \rightarrow \infty$, where $\overset{\pi}{\rightarrow}$ denotes convergence in L_π norm and $1 \le \pi \le 2$, implying convergence in probability. We take $c_{nt} = 1/n$ and $\pi = 2$ so that all the conditions are satisfied: (i) $\{|\xi_{nt}/c_{nt}|^\pi\}$ is uniformy integrable since $\{|\xi_{nt}/c_{nt}|^2\}^2 = \left(|\partial\alpha_t^{(n)}(\beta)/\partial\beta_i|_{\beta=\beta^0}\right)^4$ is bounded with respect to t and n, by $H_{1.1}$; (ii) $\limsup_{n\rightarrow\infty}\sum_{t=1}^{n} c_{nt} = 1 < \infty$; (iii) $\lim_{n\rightarrow\infty}\sum_{t=1}^{n} c_{nt}^2 = 0$.

To check the second assumption of KN1 in the sense of convergence in probability, we need to apply a weak law of large numbers on the martingale difference array

$$\left\{\frac{\partial^2\alpha_t^{(n)}(\beta)}{\partial\beta_i\partial\beta_j} - E_\beta\left(\frac{\partial^2\alpha_t^{(n)}(\beta)}{\partial\beta_i\partial\beta_j}/F_{t-1}\right)\right\}_{\beta=\beta^0} \quad . \tag{A1.6}$$

We proceed as above but with $\pi = 1$ and condition (i) results from $H_{1.2}$. The remaining part for checking the second assumption of KN1 is identical to the proof of Theorem 1, implying existence of $\hat{\beta}_n$ such that $\hat{\beta}_n \rightarrow \beta^0$ in probability.

Asymptotic normality of $\hat{\beta}_n$ comes again by checking the first assumption of KN2 which is now obtained by using a central limit theorem for martingale difference arrays. Indeed $\xi_{nt} = \lambda^T\{\partial\alpha_t^{(n)}(\beta)/\partial\beta\}_{\beta=\beta^0}$ is a martingale difference sequence for all n. We can use Hall and Heyde (1980, Theorem 2.23 p. 44 and Corollary 3.1, p. 58) where we have replaced the conditional Lindeberg condition by the stronger Lyapounov condition which has already been checked in Theorem 1. There is no substantial change in the rest of the proof, leading to the same conclusion.

Proof of Theorem 2

We refer to Theorem 1 and want to show that its assumptions $H_{1.1}$ - $H_{1.5}$ are fulfilled.

Proof of $H_{1.1}$. Let us show there is a positive constant C_1 such that (2.5) holds. From (A1.2),

$$\left(\frac{\partial\alpha_t(\beta)}{\partial\beta_i}\right)^4 \le 8\left\{\frac{1}{h_t(\beta)^4}\left(\frac{\partial h_t(\beta)}{\partial\beta_i}\right)^4\left(1-\frac{e_t^2(\beta)}{\sigma^2 h_t(\beta)}\right)^4 + \frac{2}{(\sigma^2 h_t(\beta))^4}e_t^4(\beta)\left(\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_i}\right)^4\right\} \quad .$$

The mathematical expectation of the absolute value of that quantity at $\beta = \beta^0$ can be bounded above by 8 times

$$E_{\beta^0}\left|\frac{1}{h_t(\beta)^4}\left(\frac{\partial h_t(\beta)}{\partial\beta_i}\right)^4\left(1-\frac{e_t^2(\beta)}{\sigma^2 h_t(\beta)}\right)^4\right| \tag{A1.7}$$

$$+E_{\beta^0}\left|\frac{2}{(\sigma^2 h_t(\beta))^4}e_t^4(\beta)\left(\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_i}\right)^4\right| \quad . \tag{A1.8}$$

We show separately that the expressions (A1.7-8) are bounded. The term (A1.7) is bounded from above by

$$\frac{1}{h_t(\beta^0)^4}\left(\left\{\frac{\partial h_t(\beta)}{\partial\beta_i}\right\}_{\beta=\beta^0}\right)^4 E_{\beta^0}\left|\left(1-\frac{e_t^2(\beta)}{\sigma^2 h_t(\beta)}\right)^4\right| \le \frac{8}{h_t(\beta^0)^4}\left(\left\{\frac{\partial h_t(\beta)}{\partial\beta_i}\right\}_{\beta=\beta^0}\right)^4\left(1+\frac{E|e_t^8|}{(\sigma^2 h_t(\beta^0))^4}\right)$$

$\le C_1^{(1)} = 8K_1^4(1+K)/m^4$, say, using $H_{2.2}$ - $H_{2.3}$. Since the e_t are independent random variables, (A1.8) equals

$$\frac{2}{(\sigma^2 h_t(\beta^0))^4} E(e_t^4) E_{\beta^0}\left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i}\right)^4 \quad .$$

From $H_{2.5}$ and Cauchy-Schwarz inequality, we deduce that $\{\sigma^2 h_t(\beta^0)\}^{-2} E(e_t^4) \le K^{1/2}$. There remains to show that $E_{\beta^0}(\partial \hat{w}_{t/t-1}(\beta)/\partial \beta_i)^4$ is bounded. Using (3.7), it is equal to

$$\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4} \psi_{tik_1}(\beta^0)\psi_{tik_2}(\beta^0)\psi_{tik_3}(\beta^0)\psi_{tik_4}(\beta^0)\,\{E(e_{t-k_1}e_{t-k_2}e_{t-k_3}e_{t-k_4})\} \quad ,$$

where the sums over $k_u, u = 1, \ldots, 4$ are from 1 to $t-1$. We have

$$E_{\beta^0}(\partial \hat{w}_{t/t-1}(\beta)/\partial \beta_i)^4 = \sum_k \psi_{tik}^4(\beta^0) E(e_{t-k}^4) + 3\sigma^4 \sum_{\substack{k_1,k_2 \\ k_1 \ne k_2}}\sum \psi_{tik_1}^2(\beta^0)\psi_{tjk_2}^2(\beta^0) h_{t-k_1}(\beta^0) h_{t-k_2}(\beta^0)$$

$$\le m_1^2(N_2 K^{1/2} + 3N_1)\sigma^4 \qquad \text{(A1.9)}$$

Hence (A1.8) is smaller than

$$C_1^{(2)} = 2\frac{m_1^2 K^{1/2}}{m^2}(N_2 K^{1/2} + 3N_1) \quad .$$

Consequently, the upper bound (2.5) is checked by taking the constant C_1 equal to $8\max(C_1^{(1)}, C_1^{(2)})$.

Proof of $H_{1.2}$. To check that assumption, we need to show there is a constant C_2 such that (2.6) holds. From (A1.3) and its conditional expectation at $\beta = \beta^0$, we deduce that

$$E_{\beta^0}\left|\frac{\partial^2 \alpha_t(\beta)}{\partial \beta_i \partial \beta_j} - E_{\beta^0}\left(\frac{\partial^2 \alpha_t(\beta)}{\partial \beta_i \partial \beta_j}/F_{t-1}\right)\right|^2 = E_{\beta^0}\left| -2(\sigma^2 h_t(\beta))^{-1} e_t(\beta)\frac{\partial^2 \hat{w}_{t/t-1}(\beta)}{\partial \beta_i \partial \beta_j}\right.$$

$$+ \frac{1}{h_t(\beta)}\left(1 - \frac{e_t^2(\beta)}{(\sigma^2 h_t(\beta))}\right)\left(\frac{\partial^2 h_t(\beta)}{\partial \beta_i \partial \beta_j} - \frac{2}{(h_t(\beta))}\frac{\partial h_t(\beta)}{\partial \beta_i}\frac{\partial h_t(\beta)}{\partial \beta_j}\right)$$

$$\left. + 2e_t(\beta)\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}(\sigma h_t(\beta))^{-2}\frac{\partial h_t(\beta)}{\partial \beta_i}\right|^2 \quad , \forall i, j = 1, \ldots, r. \qquad \text{(A1.10)}$$

Since that expression is not easy to bound, we proceed in three stages.

Stage 1

Let us now verify that for any pair $i, j = 1, \ldots, s$, there is a positive constant C_2 such that (2.6) is verified. For example, $\beta_i = \phi_i$ and $\beta_j = \theta_{j-s_1}$. Since $h_t(\beta)$ doesn't depend on the autoregressive and moving average parameters, and $\partial \hat{w}_{t/t-1}(\beta)/\partial \beta_i$ doesn't depend on the parameters of the variance, (A1.10) becomes

$$E_{\beta^0}\left| -2(\sigma^2 h_t(\beta))^{-1} e_t(\beta)\frac{\partial^2 \hat{w}_{t/t-1}(\beta)}{\partial \beta_i \partial \beta_j}\right|^2 = \frac{4}{h_t(\beta^0)}\sum_{k=1}^{t-1}\psi_{tijk}^2(\beta^0) h_{t-k}(\beta^0) \le \frac{4N^3 m_1}{m},$$

which is the value of the constant C_2 in (2.6).

Stage 2

Let us now consider each pair $\beta_i = \delta_{i-s}$ and $\beta_j = \delta_{j-s}$, for $i, j = s+1, \ldots r$. The left hand side of (2.6) equals

$$E_{\beta^0}\left| \frac{1}{h_t(\beta)}\left(1 - \frac{e_t^2(\beta)}{(\sigma^2 h_t(\beta))}\right)\left(\frac{\partial^2 h_t(\beta)}{\partial\beta_i\partial\beta_j} - \frac{2}{(h_t(\beta))}\frac{\partial h_t(\beta)}{\partial\beta_i}\frac{\partial h_t(\beta)}{\partial\beta_j}\right)\right|^2$$

which can be bounded by

$$\frac{1}{h_t^2(\beta^0)}\left|\left\{\frac{\partial^2 h_t(\beta)}{\partial\beta_i\partial\beta_j}\right\}_{\beta=\beta^0} - \frac{2}{h_t(\beta^0)}\left\{\frac{\partial h_t(\beta)}{\partial\beta_i}\right\}_{\beta=\beta^0}\left\{\frac{\partial h_t(\beta)}{\partial\beta_j}\right\}_{\beta=\beta^0}\right|^2 E_{\beta^0}\left(1 - \frac{e_t^2(\beta)}{\sigma^2 h_t(\beta)}\right)^2$$

$$\leq \frac{2}{m^2}\left(K_2^2 + \frac{4K_1^2}{m}\right)(3 + K^{1/2}) \quad ,$$

and defines the constant C_2 in this case.

Stage 3

Let us now consider each pair for $i = 1, \ldots, s$ and $j = s+1, \ldots, r$, i. e. $\beta_i = \phi_i$ and $\beta_j = \delta_{j-s}$. Similarly to the derivation leading to (A1.9)

$$E_{\beta^0}(\partial\hat{w}_{t/t-1}(\beta)/\partial\beta_i)^2 = \sum_{k=1}^{t-1}\psi_{tik}^2(\beta^0)E(e_{t-k}^2) \leq m_1 N_1 K^{1/4} \quad . \qquad \text{(A1.11)}$$

The left hand side of (2.6) equals

$$E_{\beta^0}\left| 2e_t(\beta)\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_i}(\sigma h_t(\beta))^{-2}\frac{\partial h_t(\beta)}{\partial\beta_j}\right|^2 \leq 4\sigma^{-2}(h_t(\beta^0))^{-3}\left|\left\{\frac{\partial h_t(\beta)}{\partial\beta_j}\right\}_{\beta=\beta^0}\right|^2 E_{\beta^0}\left|\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_i}\right|^2$$

and is bounded by

$$C_2 = \frac{4K_1^2 m_1 N_1 K^{1/4}}{m^3} \quad .$$

Proof of $H_{1.3}$. Let us show (2.7) where $V(\beta^0) = (V_{ij}(\beta^0))_{1\leq i,j\leq r}$ is a positive definite matrix of constants. We have

$$\frac{1}{2n}\sum_{t=1}^{n} E_{\beta^0}\left(\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}/F_{t-1}\right) = \frac{1}{n\sigma^2}\sum_{t=1}^{n}\left\{\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_i}\right\}_{\beta=\beta^0} h_t^{-1}(\beta^0)\left\{\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_j}\right\}_{\beta=\beta^0}$$

$$+\frac{1}{2n}\sum_{t=1}^{n}\left\{\frac{\partial h_t(\beta)}{\partial\beta_i}\right\}_{\beta=\beta^0} h_t^{-2}(\beta^0)\left\{\frac{\partial h_t(\beta)}{\partial\beta_j}\right\}_{\beta=\beta^0} . \qquad \text{(A1.12)}$$

The almost sure limit of (A1.12) for $n \to \infty$ will exist and be equal to

$$\lim_{n\to\infty}\frac{1}{n\sigma^2}\sum_{t=1}^{n} E_{\beta^0}\left(\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_i}h_t^{-1}(\beta)\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_j}\right) + \lim_{n\to\infty}\frac{1}{2n}\sum_{t=1}^{n}\left\{\frac{\partial h_t(\beta)}{\partial\beta_i}\right\}_{\beta=\beta^0} h_t^{-2}(\beta^0)\left\{\frac{\partial h_t(\beta)}{\partial\beta_j}\right\}_{\beta=\beta^0} ,$$

if the two conditions of Lemma 1, applied to the process $Z_t^{ij}(\beta)$ defined by

$$Z_t^{ij}(\beta) = \frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_i}h_t^{-1}(\beta)\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_j} - E\left(\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_i}h_t^{-1}(\beta)\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_j}\right)$$

are satisfied for $\beta=\beta^0$. To prove that the first condition of Lemma 1 is fulfilled, there suffices to show that the expression

$$E_{\beta^0}(Z_t^{ij}(\beta))^2 \le E_{\beta^0}\left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i} h_t^{-1}(\beta) \frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right)^2$$

is uniformly bounded for all t. Using Cauchy-Schwarz inequality, $H_{2.3}$ and inequality (A1.9), the expression is bounded by a constant. The first condition is thus checked.

To prove the second condition, let us show that

$$E_{\beta^0}\left(\frac{1}{n}\sum_{t=1}^{n} Z_t^{ij}(\beta)\right)^2 = \mathrm{var}_{\beta^0}\left(\frac{1}{n}\sum_{t=1}^{n}\left(\left\{\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i}\right\} h_t^{-1}(\beta)\left\{\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right\}\right)\right) = O\left(\frac{1}{n}\right).$$

The left hand side of the last expression equals

$$\frac{1}{n^2}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\mathrm{cov}_{\beta^0}\left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i} h_t^{-1}(\beta)\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}, \frac{\partial \hat{w}_{t+d/t+d-1}(\beta)}{\partial \beta_i} h_{t+d}^{-1}(\beta)\frac{\partial \hat{w}_{t+d/t+d-1}(\beta)}{\partial \beta_j}\right) \quad \text{(A1.13)}$$

The covariance can be written as

$$\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4}\psi_{tik_1}(\beta^0)\psi_{tjk_2}(\beta^0)\psi_{t+d,i,k_3}(\beta^0)\psi_{t+d,j,k_4}(\beta^0)$$

$$\{E(e_{t-k_1}e_{t-k_2}e_{t+d-k_3}e_{t+d-k_4}) - E(e_{t-k_1}e_{t-k_2})E(e_{t+d-k_3}e_{t+d-k_4})\} \quad , \qquad \text{(A1.14)}$$

where the sums over $k_u, u=1,\ldots,4$ are from 1 to $t-1$ (if u = 1 or 2) or $t+d-1$ (if u = 3 or 4). In the sequel, the sums which don't contain any term because of the constraints imposed to the summation indices are set equal to zero. Like in the proof of $H_{1.1}$, (A1.14) becomes

$$\sum_{k_1}\psi_{tik_1}(\beta^0)\psi_{tjk_1}(\beta^0)\psi_{t+d,i,k_1+d}(\beta^0)\psi_{t+d,j,k_1+d}(\beta^0)E\left(e_{t-k_1}^4\right)$$

$$+\sigma^4\left\{\sum_{\substack{k_1,k_3\\k_1\ne k_3-d}}\sum \psi_{tik_1}(\beta^0)\psi_{tjk_1}(\beta^0)\psi_{t+d,i,k_3}(\beta^0)\psi_{t+d,j,k_3}(\beta^0)h_{t-k_1}(\beta^0)h_{t+d-k_3}(\beta^0)\right.$$

$$+\sum_{\substack{k_1,k_2\\k_1\ne k_2}}\sum \psi_{tik_1}(\beta^0)\psi_{tjk_2}(\beta^0)\psi_{t+d,i,k_1+d}(\beta^0)\psi_{t+d,j,k_2+d}(\beta^0)h_{t-k_1}(\beta^0)h_{t-k_2}(\beta^0)$$

$$\left.+\sum_{\substack{k_1,k_2\\k_1\ne k_2}}\sum \psi_{tik_1}(\beta^0)\psi_{tjk_2}(\beta^0)\psi_{t+d,i,k_2+d}(\beta^0)\psi_{t+d,j,k_1+d}(\beta^0)h_{t-k_1}(\beta^0)h_{t-k_2}(\beta^0)\right\}$$

$$-\left\{\sum_{k_1}\sum_{k_2}\psi_{tik_1}(\beta^0)\psi_{tjk_2}(\beta^0)E(e_{t-k_1}e_{t-k_2})\right\}\left\{\sum_{k_3}\sum_{k_4}\psi_{t+d,i,k_3}(\beta^0)\psi_{t+d,j,k_4}(\beta^0)E(e_{t+d-k_3}e_{t+d-k_4})\right\}.$$

Hence, by completing the sums,

$$\sum_{k}\psi_{tik}(\beta^0)\psi_{tjk}(\beta^0)\psi_{t+d,i,k+d}(\beta^0)\psi_{t+d,j,k+d}(\beta^0)\kappa_{4,t-k}$$

$$+\sigma^4\left[\left\{\sum_k \psi_{tik}(\beta^0)\psi_{t+d,i,k+d}(\beta^0)h_{t-k}(\beta^0)\right\}\left\{\sum_k \psi_{tjk}(\beta^0)\psi_{t+d,j,k+d}(\beta^0)h_{t-k}(\beta^0)\right\}\right.$$

$$\left.+\left\{\sum_k \psi_{tik}(\beta^0)\psi_{t+d,j,k+d}(\beta^0)h_{t-k}(\beta^0)\right\}\left\{\sum_k \psi_{tjk}(\beta^0)\psi_{t+d,i,k+d}(\beta^0)h_{t-k}(\beta^0)\right\}\right] \quad . \qquad \text{(A1.15)}$$

By $H_{2.7}$, we have

$$\frac{1}{n^2}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\text{(A1.15)} = O\left(\frac{1}{n}\right) \quad .$$

This completes the proof of the two conditions of Lemma 1. Hence, we have the following convergence

$$\frac{1}{n}\sum_{t=1}^{n}\left(\left\{\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i}\right\}_{\beta=\beta^0} h_t^{-1}(\beta^0)\left\{\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right\}_{\beta=\beta^0} - E_{\beta^0}\left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i} h_t^{-1}(\beta)\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right)\right) \xrightarrow{a.s.} 0.$$

That result implies that

$$\lim_{n\to\infty}\frac{1}{2n\sigma^2}\sum_{t=1}^{n} E_{\beta^0}\left(\frac{\partial^2 \alpha_t(\beta)}{\partial\beta_i\partial\beta_j}/F_{t-1}\right) = \lim_{n\to\infty}\frac{1}{n}\sigma^2\sum_{t=1}^{n} E_{\beta^0}\left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i} h_t^{-1}(\beta)\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right)$$

$$+\lim_{n\to\infty}\frac{1}{2n}\sum_{t=1}^{n}\left\{\frac{\partial h_t(\beta)}{\partial\beta_i}\right\}_{\beta=\beta^0} h_t^{-2}(\beta^0)\left\{\frac{\partial h_t(\beta)}{\partial\beta_j}\right\}_{\beta=\beta^0} = V_{ij}(\beta^0) \quad . \qquad \text{(A1.16)}$$

From $H_{2.6}$, the matrix $V(\beta^0) = (V_{ij}(\beta^0))_{1\le i,j\le r}$ is a strictly definite positive matrix of constants, so $H_{1.3}$ is satisfied.

Proof of $H_{1.4}$. The mean value theorem allows us to write

$$\left\{\sum_{t=1}^{n}\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}\right\}_{\beta=\beta^*} - \left\{\sum_{t=1}^{n}\frac{\partial^2\alpha_t(\beta)}{\partial\beta_i\partial\beta_j}\right\}_{\beta=\beta^0} = \sum_l(\beta_l^* - \beta_l^0)\left\{\sum_{t=1}^{n}\frac{\partial^3\alpha_t(\beta)}{\partial\beta_i\partial\beta_j\partial\beta_l}\right\}_{\beta=\beta^{**}} \quad ,\forall \quad 1\le i,j\le r \quad ,$$

β^{**} being a point on the line joining β^* and β^0, with $\|\beta^* - \beta^0\| < \Delta$. Consequently, the left hand side of (2.8) is bounded by

$$\lim_{n\to\infty}\sup_{\Delta\downarrow 0}(n\Delta)^{-1}\|\beta^*-\beta^0\|\left|\left\{\sum_{t=1}^{n}\frac{\partial^3\alpha_t(\beta)}{\partial\beta_i\partial\beta_j\partial\beta_l}\right\}_{\beta=\beta^{**}}\right| \le \lim_{n\to\infty}\sup_{\Delta\downarrow 0}\frac{1}{n}\left|\left\{\sum_{t=1}^{n}\frac{\partial^3\alpha_t(\beta)}{\partial\beta_i\partial\beta_j\partial\beta_l}\right\}_{\beta=\beta^{**}}\right| \quad , \qquad \text{(A1.17)}$$

by continuity when $\|\beta^* - \beta^0\| \to 0$, hence, to check (2.8), there suffices to show that

$$\lim_{n\to\infty}\frac{1}{n}\left|\sum_{t=1}^{n}\left\{\frac{\partial^3\alpha_t(\beta)}{\partial\beta_i\partial\beta_j\partial\beta_l}\right\}_{\beta=\beta^0}\right| < \infty \quad \text{a.s.} \quad \text{for} \quad i,j,l = 1,\ldots,r. \qquad \text{(A1.18)}$$

By considering the derivative of (A1.3) with respect to β_l, and taking (3.4-5) and (3.7-9) into account, it is possible to show that the expression which is involved in (A1.18) takes the following form

$$\frac{1}{n}\sum_{t=1}^{n}\left|\frac{\partial^3\alpha_t(\beta)}{\partial\beta_i\partial\beta_j\partial\beta_l}\right| \le \frac{1}{n}\sum_{t=1}^{n}\{A_{1t}(\beta) + A_{2t}(\beta)e_t^2(\beta) + \sum_{k=1}^{t-1}A_{3kt}(\beta)\,|\,e_t(\beta)e_{t-k}(\beta)| \; +$$

$$+ B_{1t}(\beta) + B_{2t}(\beta)\}, \qquad \text{(A1.19)}$$

where the coefficients $A_{1t}(\beta)$, $A_{2t}(\beta)$, and $A_{3kt}(\beta), k = 1,\ldots,t-1$, evaluated at the point β^0 are bounded functions of time, by $H_{2.1}$-$H_{2.3}$, and the random variables $B_{1t}(\beta)$ and $B_{2t}(\beta)$ are defined below. Denote $\tilde{A}_1$, $\tilde{A}_2$ and $\tilde{A}_{3k}$ (for all k), respectively, the upper bounds of $A_{1t}(\beta^0)$, $A_{2t}(\beta^0)$ and $A_{3kt}(\beta^0)$. For example

$$|A_{1t}(\beta^0)| = \left|2h_{t/t-1}^{-3}(\beta^0)\left\{\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_i}\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_j}\frac{\partial h_{t/t-1}(\beta)}{\partial\beta_l}\right\}_{\beta=\beta^0}\right| \le \tilde{A}_1 = \frac{2K_1^3}{m^3},$$

with more complex expressions for the other terms: $B_{1t}(\beta)$ contains terms like $(\partial e_t(\beta)/\partial\beta_{i_1})\,(\partial e_t(\beta)/\partial\beta_{i_2})$, $B_{2t}(\beta)$ contains terms proportional to $\left(\partial^2 e_t(\beta)/\partial\beta_{i_1}\partial\beta_{i_2}\right)$, where $i_1, i_2 = i, j, l$.
The expression (A1.19) at the point β^0 can then be bounded by

$$\tilde{A}_1 + \frac{1}{n}\tilde{A}_2 \sum_{t=1}^{n} e_t^2 + \frac{1}{n}\sum_{t=1}^{n}\sum_{k=1}^{t-1}\tilde{A}_{3k}|\,e_t e_{t-k}| + \frac{1}{n}\sum_{t=1}^{n}\tilde{B}_1 + \frac{1}{n}\sum_{t=1}^{n}\tilde{B}_2, \qquad \text{(A1.20)}$$

where the random variable $\tilde{B}_1$ and $\tilde{B}_2$ still need to be determined. For the second and third terms, there suffices to show that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} e_t^2 < \infty \qquad \text{a.s.} \quad , \qquad \text{(A1.21)}$$

since, by Cauchy-Schwarz inequality, we have

$$\frac{1}{n}\sum_{t=1}^{n}|\,e_t e_{t-k}| \le \left[\left(\frac{1}{n}\sum_{t=1}^{n} e_t^2\right)\left(\frac{1}{n}\sum_{t=1}^{n} e_{t-k}^2\right)\right]^{1/2}$$

for all finite k. For the limit in (A1.21), we use the strong law of large numbers for an appropriate martingale difference sequence $Y_t = e_t^2 - E(e_t^2)$. It is immediate that $E_{\beta^0}(Y_t/F_{t-1}) = 0$ and that $E_{\beta^0}(Y_t^2)$ is uniformly bounded, using $\boldsymbol{H_{1.5}}$ and Cauchy-Schwarz inequality. We can then conclude that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} Y_t = \lim_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n} E_{\beta^0}(Y_t) < \infty \quad \text{a.s.} \quad , \qquad \text{(A1.22)}$$

The determination of $\tilde{B}_1$ and $\tilde{B}_2$ is more difficult. Let us consider, for example, $(\partial e_t(\beta)/\partial\beta_i)\,(\partial e_t(\beta)/\partial\beta_j)$ and show that

$$X_t = \left.\frac{\partial e_t(\beta)}{\partial\beta_i}\frac{\partial e_t(\beta)}{\partial\beta_j}\right|_{\beta=\beta^0} - E_{\beta^0}\left\{\frac{\partial e_t(\beta)}{\partial\beta_i}\frac{\partial e_t(\beta)}{\partial\beta_j}\right\}$$

$$= \sum_{k_1=1}^{t-1}\sum_{k_2=1}^{t-1}\psi_{t+1,i,k_1}(\beta^0)\psi_{t+1,j,k_2}(\beta^0)\,\{e_{t-k_1}e_{t-k_2} - E(e_{t-k_1}e_{t-k_2})\}\,.$$

is a L_2-mixtingale (Andrews, 1988). Let $\nu \ge 1$ be an integer. Since

$$E_{\beta^0}(X_t\,|\,F_{t-\nu}) = \sum_{k_1=\nu+1}^{t-1}\sum_{k_2=\nu+1}^{t-1}\psi_{t+1,i,k_1}(\beta^0)\psi_{t+1,j,k_2}(\beta^0)\,\{e_{t-k_1}e_{t-k_2} - E(e_{t-k_1}e_{t-k_2})\}\,,$$

and, using the L_2-norm $\|\quad\|_2$

$$\|E_{\beta^0}(X_t\,|\,F_{t-\nu})\|_2^2 = E\left\{E_{\beta^0}(X_t\,|\,F_{t-\nu})^2\right\}$$

$$= \sum_{k_1=\nu+1}^{t-1}\sum_{k_2=\nu+1}^{t-1}\sum_{k_3=\nu+1}^{t-1}\sum_{k_4=\nu+1}^{t-1}\psi_{tik_1}(\beta^0)\psi_{tjk_2}(\beta^0)\psi_{tik_3}(\beta^0)\psi_{tjk_4}(\beta^0)$$

$$E[\{e_{t-k_1}e_{t-k_2} - E(e_{t-k_1}e_{t-k_2})\}\;\{e_{t-k_3}e_{t-k_4} - E(e_{t-k_3}e_{t-k_4})\}]$$

$$= \sum_{k=\nu+1}^{t-1} \psi_{tik}(\beta^0)^2 \psi_{tjk}(\beta^0)^2 \mathrm{var}(e_{t-k}^2) + \sum\sum_{k_1,k_2=\nu+1,k_1\neq k_2}^{t-1} \psi_{tik_1}(\beta^0)^2 \psi_{tjk_2}(\beta^0)^2 E\left(e_{t-k_1}^2\right) E\left(e_{t-k_2}^2\right)$$

$$+ \sum\sum_{k_1,k_2=\nu+1,k_1\neq k_2}^{t-1} \psi_{tik_1}(\beta^0)\psi_{tjk_2}(\beta^0)\psi_{tik_2}(\beta^0)\psi_{tjk_1}(\beta^0) E\left(e_{t-k_1}^2\right) E\left(e_{t-k_2}^2\right) \quad ,$$

hence $\| E_{\beta^0}(X_t \mid F_{t-\nu})\|_2$ is bounded by a constant times Φ^ν using $H_{2.1}$, $H_{2.3}$ and $H_{2.5}$. The second condition, concerning $\| X_t - E_{\beta^0}(X_t \mid F_{t+\nu})\|_2$ is trivially true. Consequently (X_t, F_t) is a L_2-mixtingale, with the sequence $\psi_\nu = \Phi^\nu \to 0$, exponentially as $\nu \to \infty$, since $0 \le \Phi < 1$, and the sequence of constants c_t which doesn't vary with t. By using a strong law of large numbers for mixtingales (e. g. Hall and Heyde, 1980, p. 41, Theorem 2.21), we have that $n^{-1}\sum_{t=1}^{n} X_t \to 0$, and conclude like before for $\tilde{B}_1$. The determination of $\tilde{B}_2$ is similar so that $H_{2.4}$ is checked.

To conclude, the first four assumptions of Theorem 1 are satisfied. Consequently, there exists an estimator $\hat{\beta}_n$ such that $\hat{\beta}_n \to \beta^0$ a.s, and which maximises l_n in the sense of the statement of Theorem 1.

Proof of $H_{1.5}$. Let us determine the explicit form of (2.9) for all $1 \le i, j \le r$. From (A1.2), it is equal to

$$\frac{1}{n}\sum_{t=1}^{n} \frac{4}{\sigma^2}\left(\left\{\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i}\right\}_{\beta=\beta^0} h_t^{-1}(\beta^0)\left\{\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right\}_{\beta=\beta^0} - E_{\beta^0}\left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_i} h_t^{-1}(\beta^0)\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right)\right)$$

$$+\frac{1}{n}\sum_{t=1}^{n} 4\frac{E_{\beta^0}\{e_t^3(\beta)\}}{\sigma^4(h_t(\beta^0))^3}\left\{\frac{\partial h_t(\beta)}{\partial \beta_i}\right\}_{\beta=\beta^0}\left(\left\{\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right\}_{\beta=\beta^0} - E_{\beta^0}\left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right)\right). \tag{A1.23}$$

While checking $H_{1.3}$, we have shown that the first term of (A1.23) tends to zero. There remains to prove that the second term also tends to zero. Let

$$\tilde{Z}_t^{ij}(\beta) = K_t^i(\beta)\left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j} - E\left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right)\right),$$

where

$$K_t^i(\beta) = 4\frac{E\{e_t^3(\beta)\}}{\sigma^4(h_t(\beta))^3}\left\{\frac{\partial h_t(\beta)}{\partial \beta_i}\right\}.$$

We now show that the two conditions of Lemma 1 are satisfied for $\tilde{Z}_t^{ij}(\beta)$ for $i, j = 1, \ldots, r$. Let us first remark that the expectation in

$$E_{\beta^0}\{\tilde{Z}_t^{ij}(\beta)\}^2 \le (K_t^i(\beta^0))^2 E_{\beta^0}\left(\frac{\partial \hat{w}_{t/t-1}(\beta)}{\partial \beta_j}\right)^2$$

is uniformly bounded thanks to (A1.11). Also, by Cauchy-Schwarz inequality, $H_{2.5}$ and the arguments developed for proving $H_{1.2}$, we deduce that $E\{e_t^3\} \le \sigma^2 K^{1/4} m_1$. Using $H_{2.2}$ and $H_{2.3}$, we have

$$|K_t^i(\beta^0)| < \frac{4K^{1/4} m_1 K_1}{\sigma^2 m^3}. \tag{A1.24}$$

There remains to verify the second condition, in an manner analogous to while checking $H_{1.3}$:

$$E_{\beta^0}\left(\frac{1}{n}\sum_{t=1}^{n}\tilde{Z}_t^{ij}(\beta)\right)^2 = \frac{1}{n^2}\sum_{d=1}^{n-1}\sum_{t=1}^{n-d}\operatorname{cov}_{\beta^0}\left(K_t^i(\beta)\frac{\partial\hat{w}_{t/t-1}(\beta)}{\partial\beta_j}, K_{t+d}^i(\beta)\frac{\partial\hat{w}_{t+d/t+d-1}(\beta)}{\partial\beta_j}\right) \quad .$$

The covariance can be written as

$$\sigma^2 K_t^i(\beta^0)K_{t+d}^i(\beta^0)\sum_k \psi_{tjk}(\beta^0)\psi_{t+d,j,k+d}(\beta^0)h_{t-k}(\beta^0) \quad ,$$

where (A1.24) implies that

$$|\sigma^2 K_t^i(\beta^0)K_{t+d}^i(\beta^0)| < \frac{16K^{1/2}m_1^2K_1^2}{\sigma^2 m^6},$$

and again we have to verify that n^{-2} times the sum over d and the sum over t is $O(1/n)$. This is true according to $H_{2.7}$. Consequently, $H_{1.5}$ is checked.

As a conclusion, the asymptotic convergence of the estimator $\hat{\beta}_n$ towards the normal distribution is ensured and the proof of Theorem 2 is achieved.

Proof of Theorem 2’

We refer now to Theorem 1’ and keep most of the proof of Theorem 2, but where almost sure convergence is replaced everywhere by convergence in probability. Therefore, we simply indicate the modifications in the check of $H_{1.3}$, $H_{1.4}$ and $H_{1.5}$.

For $H_{1.3}$ and $H_{1.5}$, we need to consider two martingale difference arrays, respectively

$$Z_t^{(n)ij}(\beta) = \frac{\partial\hat{w}_{t/t-1}^{(n)}(\beta)}{\partial\beta_i}(h_t^{(n)}(\beta))^{-1}\frac{\partial\hat{w}_{t/t-1}^{(n)}(\beta)}{\partial\beta_j} - E\left\{\frac{\partial\hat{w}_{t/t-1}^{(n)}(\beta)}{\partial\beta_i}(h_t^{(n)}(\beta))^{-1}\frac{\partial\hat{w}_{t/t-1}^{(n)}(\beta)}{\partial\beta_j}\right\}$$

and

$$\tilde{Z}_t^{(n)ij}(\beta) = K_t^{(n)i}(\beta)\left(\frac{\partial\hat{w}_{t/t-1}^{(n)}(\beta)}{\partial\beta_j} - E\left\{\frac{\partial\hat{w}_{t/t-1}^{(n)}(\beta)}{\partial\beta_j}\right\}\right) \quad \text{, where} \quad K_t^{(n)i}(\beta) = 4\frac{E\{e_t^{(n)}(\beta)\}^3}{\sigma^4(h_t^{(n)}(\beta))^3}\left\{\frac{\partial h_t^{(n)}(\beta)}{\partial\beta_i}\right\}.$$

$t = 1, \ldots, n$, for each $n \in N$, instead of martingale sequences $Z_t^{ij}(\beta)$ and $\tilde{Z}_t^{ij}(\beta)$. The unique condition of Lemma 1’ is then verified such as in the proof of Theorem 1.

While checking $H_{1.4}$ we have used a strong law of large numbers for a mixtingale sequence. Such a strong law doesn’t exist for a mixingale array. We revert therefore to a weak law of large numbers (for example Andrews, 1988, p. 461, Theorem 2 with $k_n = n$ and sequence $c_t^{(n)} = 1$) which provides the requested result but with convergence in probability (and even in L_1-norm).

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.

ANDREWS, D. W. K. (1988) Laws of large numbers for dependent non-identically distributed random variables, *Econom. Theory* **4**, 458-467.

ANSLEY, C. F. (1979) An algorithm for the exact likelihood of a mixted autoregressive-moving average process, *Biometrika* **66**, 59-65.

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

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. and MELARD, G. (1998) The exact quasi-likelihood of time-dependent ARMA models, *J. Statist. Plann. Inference*, 68, 31-45.

AZRAK, R. and MELARD, G. (2004) Autoregressive models with time-dependent coefficients - A comparison with Dahlhaus' approach, ECARES working paper, Université Libre de Bruxelles.

BAR-SHALOM, Y. (1971) On the asymptotic properties of the maximum-likelihood estimate obtained from dependent observations, *J. Roy. Statist. Soc. Ser. B* **33**, 72-77.

BASAWA, I. V. and LUND, R. L. S. (2001) Large sample properties of parameter estimates for periodic ARMA models, *J. Time Ser. Anal.* **22**, 651-663.

BASAWA, I. V. and PRAKASA RAO, B. L. S. (1980) *Statistical Inference for Stochastic Processes*, Academic Press, London New York.

BHAT, B. R. (1974), On the method of maximum-likelihood for dependent observations, *J. Roy. Statist. Soc. Ser. B* **36**, 48-53.

BIBI, A. and FRANCQ, C. (2003), Consistent and asymptotically normal estimators for cyclically time-dependent linear models, *Ann. Inst. Statist. Math.* **55**, 41-68.

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

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

CHOW (1971) On the L_p convergence for $n^{-1/p}S_n$, $0 < p < 2$, *Ann. Math. Statist.* **36**, 393-394.

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.

CROWDER, J. (1976) Maximum-likelihood estimation for dependent observations, *J. Roy. Statist. Soc. Ser. B* **38**, 45-53.

DAHLHAUS, R. (1996a) Maximum likelihood estimation and model selection for locally stationary processes, *J. Nonparametr. Statist.* **6**, 171-191.

DAHLHAUS, R. (1996b) On the Kullback-Leibler information divergence of locally stationary processes, *Stochastic Process. Appl.* **62**, 139-168.

DAHLHAUS, R. (1996c) Asymptotic statistical inference for nonstationary processes with evolutionary spectra, In *Athens Conference on Applied Probability an Time Series Analysis*, (P. M. Robinson and M. Rosenblatt, eds) **2**. Springer, New York, pp. 145-159.

DAHLHAUS, R. (1997) Fitting time series models to nonstationary processes, *Ann. Statist.* **25**, 1-37.

DAVIDSON, J. E. H. (1994) *Stochastic Limit Theorems. An Introduction for Econometricians,* Oxford University Press, Oxford.

DOOB, J. L. (1953), *Stochastic Processes*, Wiley, New York.

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.

GUEGAN, D. (1994) *Séries chronologiques non linéaires à temps discret*, Economica, Paris.

HALL, P. and HEYDE, C. C. (1980) *Martingale Limit Theory and its Application,* Academic Press, New York.

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. Probab.* **18**, 170-210.

HALLIN, M. and INGENBLEEK, J.F. (1983) Nonstationary Yule-Walker equations, *Statist. Probab. Lett.* **1**, 189-195.

HALLIN, M. et MELARD, G. (1977) Indéterminabilité pure et inversibilité des processus autorégressifs moyenne mobile à coefficients dépendant du temps, *Cahiers du Centre d’Etude de Recherche Opérationnelle* **19**, 385-392.

HAMDOUNE, S. (1995) Etude des problèmes d’estimation de certains modèles ARMA évolutifs, Thesis presented at Université Henri Poincaré, Nancy 1.

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

KOWALESKI, A. and SZYNAL, D. (1991) On a characterization of optimal predictors for nonstationary ARMA processes, *Stochastic Process. Appl.* **37**, 71-80.

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

LUMSDAINE, R. L. (1996) Consistency and asymptotic normality of the quasi-maximum estimator in IGARCH(1, 1) and covariance stationary GARCH(1, 1) models, *Econometrica* **64**, 575-596.

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, *Cahier 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. (1985) *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. 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, 355-363.

MELARD, G. et PASTEELS, J-M. (1994) Manuel d'utilisation de Time Series Expert (TSE version 2.2), Institut de Statistique, Université Libre de Bruxelles.

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

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

PHAM, D. and TRAN, T. (1985) Some mixing properties of time series models, *Stochastic Process. Appl.* **19**, 297-303.

PRIESTLEY, M. B. (1965) Evolutionary spectra and non-stationary processes, *J. Roy. Statist. Soc. Ser. B* **27**, 204-237.

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

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

SILVEY, D. (1961) A note on maximum-likelihood in the case of dependent random variables, *J. Roy. Statist. Soc. Ser. B* **23**, 444-452.

SINGH, N. and PEIRIS, M. S. (1987) A note on the properties of some nonstationary ARMA processes, *Stochastic Process. Appl.* **24**, 151-155.

STOUT, W. F. (1974) *Almost Sure Convergence*, Academic Press, New York.

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

TIAO, G. C., and GRUPE, M. R. (1980) Hidden periodic autoregressive-moving average models in time series data, *Biometrika* **67**, 365-373.

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 Process. Appl.* **21**, 251-273.

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

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

WEGMAN, E. J. (1974) Some results on non stationary first order autoregression, *Technometrics* **16**, 321-322.

WHITE, H. (1982) Maximum likelihood estimation of misspecified models, *Econometrica* **50**, 1-25.

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