Ann. Inst. Statist. Math.
Vol. 55, No. 1, 41–68 (2003)

CONSISTENT AND ASYMPTOTICALLY NORMAL ESTIMATORS FOR CYCLICALLY TIME-DEPENDENT LINEAR MODELS

ABDELOUAHAB BIBI[1] AND CHRISTIAN FRANCQ[2]

[1]*Département de Mathématiques, Université Mentouri de Constantine, 25000 Constantine, Algérie*
[2]*Université du Littoral – Côte d'Opale, LMPA Joseph Liouville, Centre Universitaire de la Mi-Voix, 50, rue Ferdinand Buisson – BP 699, 62228 Calais Cedex, France,*
e-mail: Christian.Francq@lmpa.univ-littoral.fr

(Received November 17, 2000; revised June 6, 2002)

Abstract. We consider a general class of time series linear models where parameters switch according to a known fixed calendar. These parameters are estimated by means of quasi-generalized least squares estimators. Conditions for strong consistency and asymptotic normality are given. Applications to cyclical ARMA models with non constant periods are considered.

Key words and phrases: Time varying models, nonstationary processes, quasi-generalized least squares estimator, consistency, asymptotic normality.

1. Introduction

For many real time series, such as seasonal ones that one encounters very often in economics, stationary models are inadequate. The most popular method employed to take non stationarity into account consists in looking for a suitable transformation of the data that generates a new set of data exhibiting no apparent deviation from stationarity. However, this allows for very restrictive types of nonstationarities only. Take, for instance, the simple example of a seasonal daily series x_t of period $s = 7$. The above-mentioned method generally leads to fit an autoregressive moving-average (ARMA) model to the differenced series $y_t = x_t - x_{t-7}$. This implies that the same ARMA model applies for each day of the week. In other words, Sundays and Mondays will be predicted using the same formula. The approach may seem arbitrary. An alternative one consists in considering different ARMA models for each day of the week. This leads to the so-called periodic autoregressive moving-average (PARMA) models which have recently received much attention (see e.g. Adams and Goodwin (1995), Anderson and Vecchia (1983), Bentarzi and Hallin (1993), Lund and Basawa (2000)), Basawa and Lund (2001). PARMA models belong to the wider class of ARMA models with time-varying coefficients. Following the seminal works by Priestley (1965) and Whittle (1965), the probabilistic properties of a wide class of models with time-varying coefficients have been considered by Hallin (1986), Kowalski and Szynal (1991), Singh and Peiris (1987) among others.

However the estimation methods remain less explored for time-varying models than for stationary models. The main reason is certainly that, for such nonstationary models, the convenient asymptotic theory of stationary ergodic processes does not apply, which constitutes a major difficulty. In this paper we study the asymptotic behavior of (quasi-generalized) least squares estimators for a vast class of linear models with time-varying coefficients and heteroskedastic innovations. The quasi-generalized least squares procedure is very simple and, in our framework, its asymptotic efficiency is equivalent to that of the quasi-maximum likelihood method. We attempt to state consistency and asymptotic normality under conditions which are sufficiently explicit to be easily used for the construction of asymptotic confidence intervals and hypothesis tests, when the estimation procedure is applied to simple models of particular interest. We consider in particular models, that we call cyclical time-varying models, with a finite number of recurrent regimes which alternate with a non constant periodicity. The applications we have in mind are, for instance, economics series with different dynamics for worked days and legal holidays.

Among the authors who have studied asymptotic behaviors of estimates of time-varying coefficients, let us mention Azrak and Mélard (2000) who gave conditions for consistency and asymptotic normality of quasi-maximum likelihood estimators for general time-varying ARMA models. For locally stationary processes, Dahlhaus (1997) studied in detail a generalized Whittle estimator. Other relevant references are the following. Tyssedal and Tjøstheim (1982) gave conditions for the consistency of least squares and quasi-generalized least squares estimators of an AR(p) with constant autoregressive coefficients and innovations with time-varying variance. Tjøstheim and Paulsen (1985) showed the asymptotic normality. Kwoun and Yajima (1986) gave conditions for consistency and asymptotic normality of least squares estimators of parameters of an AR(1) with time-varying autoregressive coefficient and innovations with constant variance. The conditions given here are close in spirit to those given by Kwoun and Yajima (1986) but the considered class of models is wider, since it includes ARMA models with time-varying coefficients and time-varying conditional variance.

The paper is organized as follows. Section 2 presents the model and states consistency and asymptotic normality of least squares and quasi-generalized least squares estimators. The latter is used to take into account conditional heteroskedasticity. We make a comparison between the (quasi-generalized) least squares estimators and the quasi-maximum likelihood estimator. In Section 3, the results obtained in Section 2 for general time-varying linear models are applied to simple examples of PARMA and cyclical time-varying models. A numerical illustration is proposed in Section 4. Section 5 contains proofs of theorems stated in Section 2.

2. Main results

Consider the following time-dependent linear model

$$x_0 = \varepsilon_0, \quad x_t = \varepsilon_t + \sum_{i=1}^{t} \phi_{t,i}(\theta_0)\varepsilon_{t-i} \qquad t = 1, 2, \ldots, \tag{2.1}$$

where $\theta_0 = (\theta_0(1), \ldots, \theta_0(d))'$ is an unknown parameter of interest belonging to an open subset Θ of $\mathbb{R}^d$, the $\phi_{t,i}(\cdot)$'s are known functions from $\mathbb{R}^d$ to $\mathbb{R}$. The sequence (ε_t) is supposed to be a heteroskedastic independent white noise. More precisely we assume that $\varepsilon_t = \sigma_t\eta_t$, where $(\eta_t)_{t\geq 0}$ is a sequence of independent and identically distributed (i.i.d.) random variables such that $E\eta_t = E\eta_t^3 = 0$, $E\eta_t^2 = 1$ and $E\eta_t^4 = m_4 < \infty$, the σ_t's are strictly positive numbers.

Now let us introduce a simple example of a time-varying model of form (2.1) to make ideas concrete. Imagine that we suspect d changes in regime at known dates. Let $\Delta(k)$

be the set of the indices corresponding to regime k. Denote by $s_t := \sum_{k=1}^{d} k\mathbb{I}_{\Delta(k)}(t)$ the regime corresponding to index t. We might consider a time-varying MA(1):

$$x_0 = \sigma(s_0)\eta_0, \quad x_t = \sigma(s_t)\eta_t - b(s_t)\sigma(s_{t-1})\eta_{t-1}, \qquad t = 1, 2, \ldots. \tag{2.2}$$

This is clearly a particular form of (2.1), with parameter of interest $\theta_0 = (b(1), \ldots, b(d))$ and nuisance parameter $\beta_0 = (\sigma(1), \ldots, \sigma(d)) \in]0, +\infty[^d$. It is important to note that the coefficients of Model (2.2) are not random (in particular, (2.2) does not belong to the class of threshold moving-average models studied by de Gooijer (1998)). When (s_t) is periodic, (2.2) is a PMA(1) process. For instance, this kind of model could be relevant for daily time series with different regimes for weekdays and weekends (if $t = 0$ corresponds to Sunday, we can set $d = 2$, $\Delta(1) = \{1, 2, 3, 4, 5, 8, 9, \ldots\}$, $\Delta(2) = \{0, 6, 7, 13, 14, \ldots\}$, $(s_t)_{t \geq 0} = (2, 1, 1, 1, 1, 1, 2, 2, 1, \ldots)$). PARMA models are examples of periodically correlated (PC) processes (see Gladyshev (1963), or Alpay *et al.* (2001) for a more recent reference). Extensions which can account for more complex cyclical phenomena have been proposed. In particular, almost periodically correlated (APC) processes were introduced by Gladyshev (1963) and have been discussed by many authors (see e.g. Alpay *et al.* (2000), Dehay and Monsan (1996), Makagon and Miamee (1996) and the references therein). A discrete-time process $(x_t)_{t \in \mathbb{Z}}$ is said to be APC if its covariance function is an almost periodic sequence in the sense of Bohr (for each m, and every $\epsilon > 0$, the set of ϵ-almost periods of the function $k \mapsto R_m(k) := \mathrm{Cov}(x_{m+k}, x_m)$, defined as the natural numbers τ_ϵ such that $|R_m(k+\tau_\epsilon) - R_m(k)| < \epsilon$ for every $k \in \mathbb{Z}$, is relatively dense in $\mathbb{Z}$). Similarly, a continuous-time process $(x_t)_{t \in \mathbb{R}}$ is said to be APC if its covariance function is an almost periodic function on the real line in the sense of Bohr. Examples of APC processes are obtained from contemporaneous aggregation of independent periodic processes with incommensurate periods: for instance, $x_t = y_t \sin(t) + z_t \sin(\pi t)$, $t \in \mathbb{R}$, where (y_t) and (z_t) are independent stationary processes. A sequence on the integers is an almost periodic sequence if and only if it is the restriction on the integers of an almost periodic function on the real line (see e.g. Alpay *et al.* (2000)). It is clear that (2.2) is an APC process when the sequences $(b(s_t))_t$ and $(\sigma(s_t))_t$ are almost periodic. Another interesting situation is when the regimes are recurrent, but alternate with a non constant periodicity. For instance, this could account for legal holidays and worked days in daily economic time series. Such series are cyclical, but they are generally not APC: $\forall \tau \in \mathbb{Z}$, $\exists \epsilon > 0$, $m \in \mathbb{Z}$, $k \in \mathbb{Z}$ such that $|\mathrm{Cov}(x_{m+k+\tau}, x_m) - \mathrm{Cov}(x_{m+k}, x_m)| > \epsilon$ (it suffices to choose m and k so that $m + k$ correspond to a legal holiday and $m + k + \tau$ to a worked day). An example of cyclical model, in which the structural changes are recurrent but not periodic, will be discussed further in Section 3.

We now return to the general specification. Iterating (2.1), ε_t can be written as linear function of $x_t, x_{t-1}, \ldots, x_0$. In other words, the following autoregressive representation holds

$$x_t = \epsilon_t - \sum_{i=1}^{t} \pi_{t,i}(\theta_0) x_{t-i}, \qquad t = 1, 2, \ldots. \tag{2.3}$$

In view of (2.1), ε_t is independent of the σ-field generated by $x_{t-1}, x_{t-2}, \ldots, x_0$. Therefore, (2.3) shows that the best one-step predictor of x_t is $\hat{x}_t(\theta_0) := E_{\theta_0}(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_0) = -\sum_{i=1}^{t} \pi_{t,i}(\theta_0) x_{t-i}$. When θ_0 is replaced by θ, the prediction error is given by

$$
(2.4) \qquad e_t(\theta) := x_t + \sum_{i=1}^{t} \pi_{t,i}(\theta) x_{t-i} = \sum_{i=0}^{t} \psi_{t,i}(\theta,\theta_0)\eta_{t-i}, \qquad t = 1,2,\ldots,
$$

where $\psi_{t,i}(\theta,\theta_0) = \sum_{k=0}^{i} \pi_{t,k}(\theta)\phi_{t-k,i-k}(\theta_0)\sigma_{t-i}(\theta_0)$, $\pi_{t,0}(\theta) = \phi_{t,0}(\theta_0) = 1$. We will refer to $e_t(\theta)$ as a residual. Note that $\epsilon_t = e_t(\theta_0)$ can be interpreted as an innovations process of (x_t), and that the marginal variance of the innovation process is also the conditional variance of the observed process: $\sigma_t^2 = \mathrm{Var}_{\theta_0}(x_t \mid x_{t-1}, x_{t-2}, \ldots, x_0)$. When this conditional variance is not constant over time, i.e. $\sigma_t \not\equiv \sigma$, (x_t) is said to be conditionally heteroskedastic. However, contrary to GARCH-type models, the conditional variance does not depend on past values of (x_t). It would be desirable, in particular for applications to financial time series, to incorporate ARCH effects, but it is beyond our technical capabilities.

Numerous appealing estimation procedures are based on the minimization of weighted sums of squares of residuals (see e.g. Godambe and Heyde (1987) for a general reference and Basawa and Lund (2001) for an application to PARMA models). First consider the (ordinary) least squares estimator (LSE). Let Θ^* be a compact subset of Θ which contains a neighborhood of θ_0. Given a sequence $(x_0, \ldots, x_n)$ of observations, define a LSE as any measurable solution of

$$
(2.5) \qquad \hat{\theta}_n = \arg\min_{\theta\in\Theta^*} Q_n(\theta), \qquad \text{where} \qquad Q_n(\theta) = n^{-1}\sum_{t=1}^{n} e_t^2(\theta).
$$

For convenience, write $\psi_{t,i} = \psi_{t,i}(\theta,\theta_0)$ for $i = 0,\ldots,t$, and $\psi_{t,i} = 0$ for $i > t$. In addition to the previous assumptions, it will be supposed that:

A1. $\forall\theta,\ \theta_0 \in \Theta^*$, if $\theta \neq \theta_0$ then there exists a positive integer q_0 such that $\liminf_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sum_{i=1}^{q_0}\psi_{t,i}^2 > 0$.

A2. $\forall\theta,\ \theta_0 \in \Theta^*$, there exist constants $\gamma(|h|)$ (possibly depending on θ and θ_0 but not on t) such that

$$
\sum_{0\le i,j} |\psi_{t,i}\psi_{t,j}\psi_{t+|h|,|h|+i}\psi_{t+|h|,|h|+j}| \le \gamma(|h|),
$$

and $\sum_{h\ge 0}\gamma(h) < \infty$.

A3. $\forall\theta,\ \theta_0 \in \Theta^*$, there exist constants $\gamma(|h|)$ such that, for all $t_k, h_k, h_k^* \ge 0$, $k = 1,2,3,4$,

$$
\sum_{0\le i_k,j_k} \left| \prod_{k=1}^{4} \psi_{t_k,h_k+i_k}\psi_{t_k,h_k^*+j_k} \right| \le \gamma\left(\max\left\{\max_{k=2,3,4} h_k, \max_{k=2,3,4} h_k^*\right\}\right),
$$

$\sum_{h\ge 0} h\gamma(h) < \infty$ and $\gamma(h) \downarrow 0$ as $h \uparrow \infty$.

A4. $\forall\theta_0 \in \Theta^*$, the functions $\pi_{t,i}(\cdot)$ and $\psi_{t,i}(\cdot,\theta_0)$ admit third order partial derivatives.

A5. $\forall\theta_0 \in \Theta^*$, $\sum_{i\ge 0}\bar{\gamma}_i < \infty$, where

$$
\begin{aligned}
\bar{\gamma}_i = \sup_{t\ge 1}\sup_{\theta\in\Theta^*} \max\{ & |\psi_{t,i}(\theta,\theta_0)|, |\psi_{t,i}^{(1)}(\theta,\theta_0)|, \ldots, |\psi_{t,i}^{(d)}(\theta,\theta_0)|, \\
& |\psi_{t,i}^{(11)}(\theta,\theta_0)|, \ldots, |\psi_{t,i}^{(dd)}(\theta,\theta_0)|, |\psi_{t,i}^{(111)}(\theta,\theta_0)|, \ldots, |\psi_{t,i}^{(ddd)}(\theta,\theta_0)|\},
\end{aligned}
$$

with $\psi_{t,i}^{(k)}(\cdot,\theta_0) = \partial\psi_{t,i}(\cdot,\theta_0)/\partial\theta(k)$, $\psi_{t,i}^{(k_1k_2)}(\cdot,\theta_0) = \partial^2\psi_{t,i}(\cdot,\theta_0)/\partial\theta(k_1)\partial\theta(k_2)$ and $\psi_{t,i}^{(k_1k_2k_3)}(\cdot,\theta_0) = \partial^3\psi_{t,i}(\cdot,\theta_0)/\partial\theta(k_1)\partial\theta(k_2)\partial\theta(k_3)$.

These assumptions will be discussed further in Section 3, and technical remarks will be given in Section 5. Notice that, in view of (2.5), $\hat\theta_n$ can be interpreted as a M-estimator. Under A4, this is also a Z-estimator, because $\hat\theta_n$ is solution to estimating equations

$$\sum_{t=1}^{n} e_t(\hat\theta_n)\sum_{i=1}^{t}\frac{\partial}{\partial\theta(k)}\pi_{t,i}(\hat\theta_n)x_{t-i} = 0, \qquad k = 1,\dots,d. \tag{2.6}$$

We are now in a position to state the first result of this paper.

THEOREM 2.1. *Under* A1, A3–A5, $\hat\theta_n$ *tends almost surely to* θ_0 *as* $n \to \infty$.

Let the gradient vector $\psi_{t,i}^{(\cdot)} = (\psi_{t,i}^{(1)}(\theta_0,\theta_0),\dots,\psi_{t,i}^{(d)}(\theta_0,\theta_0))'$. In order to establish the asymptotic normality of the LSE, consider the following additional assumptions.

A6. $\{\sigma_t\}_t$ is such that $\inf_t \sigma_t > 0$.

A7. There exist a positive integer r_0 such that for all $r \ge r_0$, the matrices

$${}_rI := \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sigma_t^2\sum_{i=1}^{r}\psi_{t,i}^{(\cdot)}\psi_{t,i}^{(\cdot)'}$$

exist.

A8. The matrix

$$J := \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sum_{i=1}^{t}\psi_{t,i}^{(\cdot)}\psi_{t,i}^{(\cdot)'}$$

exists and is strictly positive definite.

It will be shown in Section 5 that, under A1–A8, the matrix

$$I := \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sigma_t^2\sum_{i=1}^{t}\psi_{t,i}^{(\cdot)}\psi_{t,i}^{(\cdot)'}$$

exists and is strictly positive definite. We have the following theorem.

THEOREM 2.2. *Let the assumptions of Theorem* 2.1 *be satisfied. In addition, assume* A6–A8. *Then* $\sqrt{n}(\hat\theta_n - \theta_0)$ *converges in law to the centered normal distribution with covariance matrix* $\Sigma := J^{-1}IJ^{-1}$.

Detailed proofs of Theorems 2.1 and 2.2 are given in Section 5. An outline of the proofs is the following.

SKETCH OF PROOF OF THEOREMS 2.1 and 2.2. Because $e_t(\theta) - \epsilon_t$ belongs to the σ-field generated by $x_{t-1},\dots,x_0$ and because ϵ_t is centered and independent of this σ-field, $E\epsilon_t\{e_t(\theta) - \epsilon_t\} = 0$ for all t. Therefore, using the identifiability assumption and a strong law of large numbers for independent but non identically distributed random variables, the objective function is shown to be asymptotically minimal at θ_0:

$$Q_n(\theta) - Q_n(\theta_0) = n^{-1}\sum_{t=1}^{n}\{e_t(\theta) - \epsilon_t\}^2 + 2n^{-1}\sum_{t=1}^{n}\epsilon_t\{e_t(\theta) - \epsilon_t\}$$

is asymptotically strictly positive when $\theta \neq \theta_0$. This is not sufficient to show Theorem 2.1, but we can infer Theorem 2.1 from a standard compactness argument.

Using a standard Taylor expansion, we have

$$n^{1/2}(\hat{\theta}_n - \theta_0) = \left\{ -\frac{\partial^2}{\partial\theta\partial\theta'} Q_n(\theta_0) \right\}^{-1} n^{1/2} \frac{\partial}{\partial\theta} Q_n(\theta_0) + o_P(1).$$

The first task is to show that $\{-\partial^2 Q_n(\theta_0)/\partial\theta\partial\theta'\} \to -2J$ in probability. The proof is more demanding than in the standard framework of stationary processes because the ergodic theorem does not apply. Approximating $n^{1/2}\partial Q_n(\theta_0)/\partial\theta = n^{-1/2}\sum_{t=1}^{n} 2\epsilon_t \partial e_t(\theta_0)/\partial\theta$ by standardized empirical means of r-dependent processes and using the Lyapounov central limit theorem, it can be shown that $n^{1/2}\partial Q_n(\theta_0)/\partial\theta$ has a limiting normal distribution with mean 0 and covariance matrix $4I$. The conclusion in Theorem 2.2 follows. □

For the linear regression model, it is well known that the ordinary LSE is not efficient when the errors are heteroskedastic (see e.g. Gouriéroux and Monfort (1995) or Hamilton (1994)). The same problem arises here, in the case of conditionally heteroskedastic models (i.e. when σ_t is not constant). To remedy this problem, the idea is to weight appropriately the residuals. Consider therefore the measurable solutions of

$$\hat{\theta}_n^{(\tau)} = \arg\min_{\theta\in\Theta^*} Q_n^{(\tau)}(\theta), \qquad \text{where} \qquad Q_n^{(\tau)}(\theta) = n^{-1}\sum_{t=1}^{n} \tau_t e_t^2(\theta) \tag{2.7}$$

and $\tau = (\tau_t)_t$ is a sequence of positive weights. Assume that $\tau_t = \tau_t(x_0, \ldots, x_{t-1})$ is measurable with respect to the σ-field generated by the random variables $x_0, \ldots, x_{t-1}$. Denote by A6* the assumption that, almost surely, $\sup_t \tau_t\sigma_t < \infty$ and $\inf_t \tau_t\sigma_t > 0$. Let

$$I^{(\tau)} := \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n} \tau_t^2\sigma_t^2 \sum_{i=1}^{t} \psi_{t,i}^{(\cdot)}\psi_{t,i}^{(\cdot)\prime}, \qquad J^{(\tau)} := \lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sum_{i=1}^{t} \tau_t \psi_{t,i}^{(\cdot)}\psi_{t,i}^{(\cdot)\prime}.$$

Denote by A7* (respectively A8*) the assumption obtained by replacing I by $I^{(\tau)}$ (respectively J by $J^{(\tau)}$) in A7 (respectively A8). We have the following result.

THEOREM 2.3. *Assume* A1–A5 *and* A6*–A8*. *Then the weighted LSE* $\hat{\theta}_n^{(\tau)}$ *converges almost surely to* θ_0 *and* $\sqrt{n}(\hat{\theta}_n^{(\tau)} - \theta_0)$ *converges in law to the centered normal distribution with covariance matrix* $\Sigma^{(\tau)} := (J^{(\tau)})^{-1}I^{(\tau)}(J^{(\tau)})^{-1}$ *as* $n \to \infty$.

The proof of Theorem 2.3 is omitted since it is very similar to that of Theorems 2.1 and 2.2. It is easy to show that the asymptotic variance of the weighted LSE of any linear combination of $\theta(1), \ldots, \theta(d)$ is minimal when $\tau_t = \sigma_t^{-2}$. The proof is given in Lemma 5.10 below. For this asymptotically optimal sequence of weights, the weighted LSE is called the generalized least squares (GLS) estimator and is denoted by $\hat{\theta}_n^G$. Its asymptotic covariance matrix is $\Sigma^G = (J^G)^{-1} := (J^{(\tau_0)})^{-1}$, where $\tau_0 = (\sigma_t^{-2})_t$. In most of the practical situations, as in (2.2) for instance, σ_t is unknown and depends on a nuisance parameter β: $\sigma_t = \sigma_t(\beta)$. When a consistent estimator $\hat{\beta}_n$ of β is available, the quasi-generalized least squares (QLS) estimator $\hat{\theta}_n^Q$, obtained by replacing σ_t by $\sigma_t(\hat{\beta}_n)$

in $\hat{\theta}_n^G$, possesses the same asymptotic behavior as the GLS estimator. More precisely, we have the following result.

THEOREM 2.4. *Suppose that* $\sigma_t = \sigma_t(\beta_0)$ *is continuous in* β_0 *uniformly in* t, *and that* $\hat{\beta}_n$ *converges almost surely to* β_0. *Assume* A1–A5 *and* A6*–A8*, *where* τ *is replaced by* τ_0. *Let* $\hat{\tau}_0 = (\sigma_t^{-2}(\hat{\beta}_n))_t$. *Then the QLS estimator* $\hat{\theta}_n^Q := \hat{\theta}_n^{(\hat{\tau}_0)}$ *converges almost surely to* θ_0 *and* $\sqrt{n}(\hat{\theta}_n^Q - \theta_0)$ *converges in law to the centered normal distribution with covariance matrix* Σ^G *as* $n \to \infty$.

The proof of Theorem 2.4 is given in Section 5.

To discuss the practical implementation of QLS estimators, let us consider the simple model defined by (2.2). Assume that the relative frequency of regime k, $n^{-1}\sum_{t=1}^n \mathbb{I}_{\Delta(k)}(t)$, converges to $\pi(k) \in]0,1[$, for $k = 1,\dots,d$. It is easy to see that a consistent estimator of the conditional variance of regime k is given by

$$\hat{\sigma}_n^2(k) := \frac{1}{\pi(k)n}\sum_{t=1}^n e_t^2(\hat{\theta}_n)\mathbb{I}_{\Delta(k)}(t), \qquad k = 1,\dots,d. \tag{2.8}$$

A QLS estimator is then obtained in the following way:

Step 1. an ordinary LSE is obtained by solving (2.5);
Step 2. an estimate $\hat{\beta} = (\hat{\sigma}_n(1),\dots,\hat{\sigma}_n(d))$ of β is obtained using (2.8);
Step 3. a QLS estimator is obtained by solving (2.7) with $\tau_t = \hat{\sigma}_n^{-2}(s_t)$.

This algorithm is usually attributed to Cochrane and Orcutt (1949). Refined QLS estimators can be found by repeating Steps 2 and 3 several times, replacing, in Step 2, $\hat{\theta}_n$ by the QLS estimator derived in Step 3. All these QLS estimators have the same asymptotic behavior.

We now briefly compare the QLS and quasi-maximum likelihood (QML) procedures for the general formulation (2.1). The assumptions of Theorem 2.4 are imposed. Parameters θ and β are supposed to be functionally independent. In QML procedures, a likelihood function is used as a vehicle to estimate the parameters, but need not be the correct density. The Gaussian likelihood, referred to as quasi-likelihood, is frequently used to form the estimator. We follow this practice. Thus, given the initial value x_0, the (conditional) log-quasi-likelihood of $(x_1,\dots,x_n)$ is, apart from a constant,

$$\ell_n(\theta,\beta) = -\frac{1}{2}\sum_{t=1}^n \left\{\frac{e_t^2(\theta)}{\sigma_t^2(\beta)} + \ln\sigma_t^2(\beta)\right\}. \tag{2.9}$$

Maximization of (2.9) leads to QML estimators $(\hat{\theta}_n^L, \hat{\beta}_n^L)$. Under suitable regularity conditions, $n^{1/2}\{(\hat{\theta}_n^L, \hat{\beta}_n^L) - (\theta_0,\beta_0)\}$ converges in law to a centered Gaussian distribution with variance involving expectations of first and second order derivatives of $\ell_n(\theta,\beta)$ (the reader is referred to Azrak and Mélard (2000) for a more rigorous statement). Since $\partial\ell_n(\theta,\beta_0)/\partial\theta$ and $\partial^2\ell_n(\theta,\beta_0)/\partial\theta\partial\theta'$ are proportional to $\partial Q_n^{(\tau_0)}(\theta)/\partial\theta$ and $\partial^2 Q_n^{(\tau_0)}(\theta)/\partial\theta\partial\theta'$, it is easily seen that $\hat{\theta}_n^L$ and $\hat{\theta}_n^Q$ have the same first order asymptotic behavior. However, minimization of $Q_n^{(\tau_0)}(\theta)$ is often easier than maximization of (2.9), because the optimization is made over a parameter space of smaller dimension. In some particular cases, the dimension of the QML optimization can be reduced. Indeed, differentiating $\ell_n(\theta,\beta)$ partially with respect to β, $\hat{\theta}_n^L$ minimizes the objective function

$$\ell_n^*(\theta) = n^{-1}\sum_{t=1}^{n}\left[\frac{e_t^2(\theta)}{\sigma_t^2\{\hat{\beta}_n^L(\theta)\}} + \ln\sigma_t^2\{\hat{\beta}_n^L(\theta)\}\right], \tag{2.10}$$

where $\hat{\beta}_n^L(\theta)$ satisfies

$$\sum_{t=1}^{n}\frac{e_t^2(\theta)}{\sigma_t^4\{\hat{\beta}_n^L(\theta)\}}\frac{\partial}{\partial\beta}\sigma_t^2\{\hat{\beta}_n^L(\theta)\} = \sum_{t=1}^{n}\frac{1}{\sigma_t^2\{\hat{\beta}_n^L(\theta)\}}\frac{\partial}{\partial\beta}\sigma_t^2\{\hat{\beta}_n^L(\theta)\}. \tag{2.11}$$

We shall refer to $\ell_n^*(\theta)$ as the reduced likelihood. In the simple case when $\sigma_t^2 \equiv \beta$, Equation (2.11) leads to

$$\hat{\beta}_n^L(\theta) = n^{-1}\sum_{t=1}^{n} e_t^2(\theta)$$

and minimization of (2.10) is equivalent to the LSE minimization:

$$n^{-1}\sum_{t=1}^{n}\left[\frac{e_t^2(\theta)}{\sigma_t^2\{\hat{\beta}_n^L(\theta)\}} + \ln\sigma_t^2\{\hat{\beta}_n^L(\theta)\}\right] = 1 + \ln\left\{n^{-1}\sum_{t=1}^{n}e_t^2(\theta)\right\}.$$

In the general case, the reduced likelihood seems less useful because it is not easy to solve (2.11). Thus, QLS and QML estimators have the same asymptotic accuracy but QLS estimators are generally simpler to implement.

3. Application to seasonal or cyclic time series

In this section we will check Assumptions A1–A8 for very simple examples. The first example is a conditionally homoscedastic PARMA model with two AR(1) regimes. In this very simple case, the LSE of the time-varying coefficients can be obtained from the LSE of a time-constant multivariate model. A condition for the asymptotic normality of the time-varying model LSE follows. This condition is that the product of the AR coefficients over a period is less than unity in absolute value. This is the causality condition given by Vecchia (1985). We show that, in this particular case, Assumptions A1–A8 reduce to this condition. The second example is a conditionally heteroskedastic model with two MA(1) regimes. Cases where the regime switches at regular time intervals (i.e. the PARMA model case) or irregular time intervals (i.e. the cyclical model case) are considered. It will be seen that the QLS estimator may perform much better than the LSE.

3.1 *A seasonal time-varying* AR(1) *model*

Consider a daily time series $x_0, x_1, \ldots$. Suppose that time $t = 0$ corresponds to Monday. We suspect different behaviors on weekends and on weekdays. Thus, we define two sets of indices corresponding to two regimes; $\Delta = \{0,1,2,3,4,7,8,9,\ldots\}$ contains the indices corresponding to the weekday regime (Regime 1), and $\Delta^c = \{5,6,12,13,\ldots\}$ contains the indices corresponding to the weekend regime (Regime 2). The regime corresponding to index t is $s_t = \mathbb{I}_\Delta(t) + 2\mathbb{I}_{\Delta^c}(t)$. We have $s_t = 1$ when $t - 7[t/7] \leq 4$ and

$s_t = 2$ when $t - 7[t/7] > 4$ ($[x]$ denoting the integer part of x). After some relevant transformations (for instance, subtraction of the daily sample means), consider the following time-varying AR(1) model

$$x_0 = \eta_0, \quad x_t = a_t(\theta)x_{t-1} + \eta_t, \qquad t = 1, 2, \dots, \tag{3.1}$$

where $\theta = (a, \tilde{a}) \in \mathbb{R}^2$, and $a_t(\theta) = a\mathbb{I}_\Delta(t) + \tilde{a}\mathbb{I}_{\Delta^c}(t)$. So, the dynamics is that of an AR(1), with parameter $\tilde{a}$ on Saturdays and Sundays, and parameter a for the rest of the week. We do not expect this model will be plausible for numerous real time series. This is only an illustrative example chosen for its simplicity. Figure 1 displays a realization of length 100 of this model.

The LSE $\hat{\theta}_n = (\hat{a}_n, \hat{\tilde{a}}_n)$ is explicitly given by

$$\hat{a}_n = \frac{\sum_{\{t:2\le t\le n\}} x_t x_{t-1}\mathbb{I}_\Delta(t)}{\sum_{\{t:2\le t\le n\}} x_{t-1}^2\mathbb{I}_\Delta(t)}, \qquad \hat{\tilde{a}}_n = \frac{\sum_{\{t:2\le t\le n\}} x_t x_{t-1}\mathbb{I}_{\Delta^c}(t)}{\sum_{\{t:2\le t\le n\}} x_{t-1}^2\mathbb{I}_{\Delta^c}(t)}. \tag{3.2}$$

Because of the non stationarity of the model, the direct study of this estimator is not obvious. With periodic coefficients, it is possible to embed seasons into a multivariate stationary process (see Tiao and Grupe (1980)). More precisely, $\boldsymbol{x}_t := (x_{7t}, x_{7t-1}, \dots, x_{7t-6})'_{t=1,2,\dots}$ is an AR(1) process of the form

$$\boldsymbol{x}_t = \begin{pmatrix} a_0 & 0 & & & \\ 0 & \tilde{a}_0 & \ddots & & \\ & \ddots & \tilde{a}_0 & & \\ & & & a_0 & \\ & & & & \ddots \end{pmatrix} \boldsymbol{x}_{t-1} + \begin{pmatrix} \epsilon_{7t} \\ \epsilon_{7t-1} \\ \\ \vdots \\ \epsilon_{7t-6} \end{pmatrix} := A\boldsymbol{x}_{t-1} + \boldsymbol{\epsilon}_t. \tag{3.3}$$

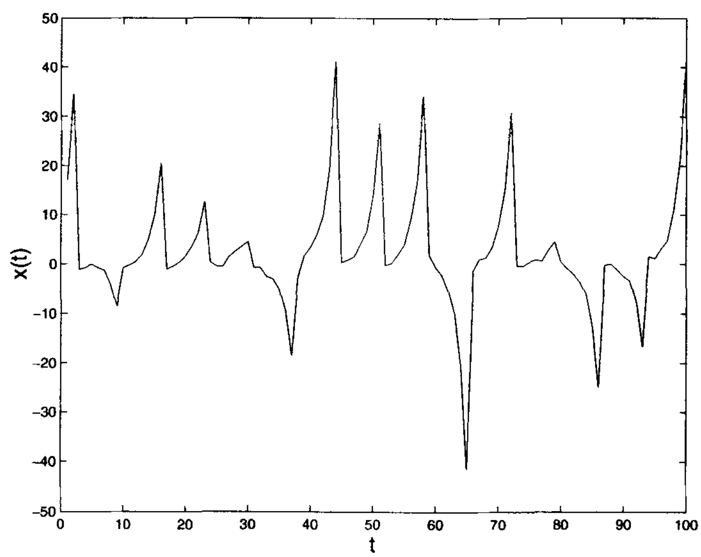

Fig. 1. A simulation of length 100 of Model (3.1) with $a = 2$ and $\tilde{a} = 0$.

Therefore, estimates (3.2) can be obtained by estimating the VAR(1) model (3.3) subject to linear constraints (coming from the special form of A). Neglecting the fact that the initial distribution is not the invariant distribution (because of the initial value $x_0 = \eta_0$), (x_t) is stationary whenever $|\det A| = |\tilde{a}_0^2 a_0^5|$ is strictly less than one. It is well known that $|\det A| < 1$ is a sufficient condition for the consistency and asymptotic normality of the LSE of A (in addition to standard assumptions such as the existence of fourth-order moments for the innovations). It is also well known that the LSE asymptotic distribution is not standard for unit-root models (see e.g. Dickey and Fuller (1979)) and for explosive models in which $|\det A| > 1$ (see e.g. White (1958)). Although a formal proof appears quite challenging, it seems reasonable to conjecture that

$$|\tilde{a}_0^2 a_0^5| < 1 \tag{3.4}$$

is a necessary and sufficient condition for the LSE asymptotic distribution stated in Theorem 2.2 to hold. This condition is that the product of the AR coefficients over a period is less than unity in absolute value. This is the causality condition given by Vecchia (1985).

Our aim in this section is therefore to see whether Assumptions A1–A8 reduce to (3.4) or not. Denote $\theta_0 = (a_0, \tilde{a}_0)$ the true value of the unknown parameter. When t corresponds to Monday, it is easy to see that $(\psi_{t,i}, i = 1, 2, \ldots) = (a_0 - a, (a_0 - a)\tilde{a}_0, (a_0 - a)\tilde{a}_0^2, (a_0 - a)\tilde{a}_0^2 a_0, (a_0 - a)\tilde{a}_0^2 a_0^2, \ldots, (a_0 - a)\tilde{a}_0^2 a_0^5, (a_0 - a)\tilde{a}_0^3 a_0^5, (a_0 - a)\tilde{a}_0^4 a_0^5, (a_0 - a)\tilde{a}_0^4 a_0^6, \ldots, (a_0 - a)\tilde{a}_0^4 a_0^{10}, \ldots)$. Assume that (3.4) holds. We have then

$$\lim_{t\to\infty, t-7[t/7]=0} \sum_{i=1}^{t} \psi_{t,i}^2$$
$$= (a_0 - a)^2 \left\{ \frac{1}{1 - \tilde{a}_0^4 a_0^{10}} (1 + \tilde{a}_0^2 + \tilde{a}_0^4 + \tilde{a}_0^4 a_0^2 + \tilde{a}_0^4 a_0^4 + \tilde{a}_0^4 a_0^6 + \tilde{a}_0^4 a_0^8 + \tilde{a}_0^4 a_0^{10}) \right\}.$$

Similar results hold for the other weekdays. Using Cesaro sums, it is now easy to see that

$$\frac{1}{n} \sum_{t=1}^{n} \sum_{i=1}^{t} \psi_{t,i}^2$$

converges to a finite positive number. It is clear that A1 holds since this number is strictly positive if and only if $a_t(\theta) = a_t(\theta_0)$ for all t, which is equivalent to $(a, \tilde{a}) = (a_0, \tilde{a}_0)$. Let δ a small positive number such that $\theta_0 \in \Theta^* = \Theta_\delta := \{\theta = (a, \tilde{a}) : |a^5 \tilde{a}^2| \le \rho := (1 - \delta) < 1, |a| \le \delta^{-1}, |\tilde{a}| \le \delta^{-1}\}$. We have $\sup_t \sup_{\theta \in \Theta_\delta} |\psi_{t,i}| \le 2\delta^{-7} \rho^{[i/7]}$. Therefore A2 holds with

$$\gamma(h) = 8\delta^{-28} \left(\frac{1}{1 - \rho^{2/7}} \right)^2 \rho^{2h/7}.$$

Similarly it is easy to check A3. Assumption A4 is straightforwardly satisfied. On Θ_δ, $|\psi_{t,i}^{(k)}| \le \delta^{-7} \rho^{[(i-1)/7]}$. Moreover, second and third order partial derivatives $\psi_{t,i}^{(k_1 k_2)}(\theta, \theta_0)$ and $\psi_{t,i}^{(k_1 k_2 k_3)}(\theta, \theta_0)$ are equal to zero. Therefore A5 is satisfied. Assumption A6 is straightforwardly satisfied. Tedious computations show that $I = J$ is diagonal, with diagonal elements

$$I(1,1) = \frac{1}{7(1 - a_0^{10} \tilde{a}_0^4)} \{5 + \tilde{a}_0^2 + \tilde{a}_0^4 + a_0^2 (4 + \tilde{a}_0^2 + 2\tilde{a}_0^4)$$

$$+ a_0^4(3+\tilde{a}_0^2+3\tilde{a}_0^4) + a_0^6(2+\tilde{a}_0^2+4\tilde{a}_0^4) + a_0^8(1+\tilde{a}_0^2+5\tilde{a}_0^4)\},$$

$$I(2,2) = \frac{1}{7(1-a_0^{10}\tilde{a}_0^4)}\left\{\frac{(1+\tilde{a}_0^2)(1-a_0^{12})}{1-a_0^2} + 1 + a_0^{10}\tilde{a}_0^2\right\}.$$

Assumptions A7 and A8 follow. We have verified that, as expected, the assumptions of Theorem 2.2 are satisfied when (3.4) holds. Note that, for this PAR(1) model, an alternative way to compute I is to apply Theorem 3.1 in Lund *et al.* (2001).

A few comments on the matrix I are the following. It can be shown that nI is asymptotically equivalent to the Fisher information matrix of the parameters. Thus, $nI(1,1)$ and $nI(2,2)$ can be interpreted as the information that a series of length n contains about a_0 and $\tilde{a}_0$, respectively. Note that when $\tilde{a}_0 = a_0$, we obtain $I(2,2) = 2/\{7(1-a_0^2)\}$ and $I(1,1) = 5/\{7(1-a_0^2)\}$. Recall that, for a time-constant AR(1) realization of length n, the information about the parameter a_0 is $n(1-a_0^2)^{-1}$. Thus, when $\tilde{a}_0 = a_0$, the information about the weekend and weekday parameters is proportional to the weekends and weekdays frequencies. This result is not valid when $\tilde{a}_0 \neq a_0$. If, for instance, $a_0 = 2$ and $\tilde{a}_0 = 0$, then we have $I(2,2) = 1366/7$. In this case the information given by the time-varying AR series about $\tilde{a}_0$ is much more important than that given by the corresponding time-constant AR of same length (i.e. $2/7$). In other words, weekdays contain useful information about the weekend behavior.

3.2 *A time-varying conditionally heteroskedastic* MA(1) *model*

Consider a time-varying MA(1) model of the form (2.2) with two regimes:

$$x_0 = \sigma_0\eta_0, \quad x_t = \sigma_t\eta_t - b_t(\theta_0)\sigma_{t-1}\eta_{t-1}, \qquad t = 1,2,\dots, \tag{3.5}$$

where the parameter of interest $\theta = (b,\tilde{b})$, $b_t(\theta) = b\mathbb{I}_\Delta(t) + \tilde{b}\mathbb{I}_{\Delta^c}(t)$, the conditional variance $\sigma_t^2 = \sigma^2\mathbb{I}_\Delta(t) + \tilde{\sigma}^2\mathbb{I}_{\Delta^c}(t)$ depends on the nuisance parameter $\beta = (\sigma,\tilde{\sigma})$, Δ denotes a subset of integers and Δ^c denotes its complement. An important difference with Example (3.1) is that it is not supposed that the process switches at regular time intervals. It is important to relax this assumption when, for instance, one wants to take into account weekends, legal holidays and strike days. The term σ_t has been added to allow a different mean square error of prediction in the two regimes. We have,

$$\psi_{t,0}(\theta,\theta_0) = \sigma_t, \quad \psi_{t,1}(\theta,\theta_0) = \{b_t(\theta) - b_t(\theta_0)\}\sigma_{t-1},$$
$$\psi_{t,i}(\theta,\theta_0) = b_t(\theta)\cdots b_{t-i+2}(\theta)\{b_{t-i+1}(\theta) - b_{t-i+1}(\theta_0)\}\sigma_{t-i}, \qquad 2 \le i \le t.$$

Assume that the relative frequency of the first regime, $n^{-1}\sum_{t=1}^n \mathbb{I}_\Delta(t)$, converges to some number $\pi \in]0,1[$, as $n \to \infty$. Then A1 holds with $q_0 = 1$, since

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^n \psi_{t,1}^2 &\ge \frac{1}{n}\sum_{t=1}^n \{b_t(\theta) - b_t(\theta_0)\}^2 \min\{\sigma_0^2,\tilde{\sigma}_0^2\} \\ &\to \min\{\sigma_0^2,\tilde{\sigma}_0^2\}\{\pi(b-b_0)^2 + (1-\pi)(\tilde{b}-\tilde{b}_0)^2\}.\end{aligned}$$

Now suppose that

$$\text{for some } \rho \in]0,1[,\ \theta_0 \in \Theta^* := \Theta_\rho = \{(b,\tilde{b}) : \max\{|b|,|\tilde{b}|\} \le \rho\}. \tag{3.6}$$

Then we have

$$|\psi_{t,i}(\theta,\theta_0)| \le \rho^{i-1}2\rho\max\{\sigma_0,\tilde{\sigma}_0\}.$$

Similar bounds hold for $|\psi_{t,i}^{(k)}(\theta,\theta_0)|$, $|\psi_{t,i}^{(k_1k_2)}(\theta,\theta_0)|$ and $|\psi_{t,i}^{(k_1k_2k_3)}(\theta,\theta_0)|$. This entails A2, A3 and A5. Therefore Theorem 2.1 applies.

To obtain the asymptotic distribution of the LSE, additional assumptions on Δ are required. Indeed, as we will see later, I depends on Δ not only through π. For instance, consider the case when $\{\mathbb{I}_\Delta(t)\}_{t=0,1,\ldots}$ constitutes a realization of an i.i.d. sequence of Bernoulli distribution with parameter π. We assume that this sequence of Bernoulli random variables is independent of the noise (η_t). In this case, Δ is referred to as a set of Bernoulli shifts with parameter π. Applying the ergodic theorem to the stationary processes $\{b_t(\theta_0)\cdots b_{t-r+2}(\theta_0)\mathbb{I}_\Delta(t-r+1)\sigma_{t-r}\}_{t\geq r}$, $r\geq 2$, standard computations show that

$$I=\frac{\sigma_0^2\pi+\tilde{\sigma}_0^2(1-\pi)}{1-\{b_0^2\pi+\tilde{b}_0^2(1-\pi)\}}\times\begin{pmatrix}\pi\{\sigma_0^2-\tilde{b}_0^2(1-\pi)(\sigma_0^2-\tilde{\sigma}_0^2)\} & 0\\ 0 & (1-\pi)\{\tilde{\sigma}_0^2+b_0^2\pi(\sigma_0^2-\tilde{\sigma}_0^2)\}\end{pmatrix},$$

$$J=\frac{\sigma_0^2\pi+\tilde{\sigma}_0^2(1-\pi)}{1-\{b_0^2\pi+\tilde{b}_0^2(1-\pi)\}}\begin{pmatrix}\pi & 0\\ 0 & 1-\pi\end{pmatrix},$$

and

$$\Sigma=\frac{1-\{b_0^2\pi+\tilde{b}_0^2(1-\pi)\}}{\sigma_0^2\pi+\tilde{\sigma}_0^2(1-\pi)}\times\begin{pmatrix}\{\sigma_0^2-\tilde{b}_0^2(1-\pi)(\sigma_0^2-\tilde{\sigma}_0^2)\}/\pi & 0\\ 0 & \{\tilde{\sigma}_0^2+b_0^2\pi(\sigma_0^2-\tilde{\sigma}_0^2)\}/(1-\pi)\end{pmatrix},$$

for almost all observed sequences $\{\mathbb{I}_\Delta(t)\}_{t=0,1,\ldots}$ of shifts, provided $b_0^2\pi+\tilde{b}_0^2(1-\pi)<1$. Similar calculations show that the asymptotic variance of the QLS estimators is

$$\Sigma^G=\frac{1-\{b_0^2\pi+\tilde{b}_0^2(1-\pi)\}}{\sigma_0^2\pi+\tilde{\sigma}_0^2(1-\pi)}\times\begin{pmatrix}\pi^{-1}\{\sigma_0^{-2}-\tilde{b}_0^2(1-\pi)(\sigma_0^{-2}-\tilde{\sigma}_0^{-2})\}^{-1} & 0\\ 0 & (1-\pi)^{-1}\{\tilde{\sigma}_0^{-2}+b_0^2\pi(\sigma_0^{-2}-\tilde{\sigma}_0^{-2})\}^{-1}\end{pmatrix}.$$

According to the general theory, the QLS estimators outperform the ordinary least squares estimator ($\Sigma-\Sigma^G$ is always a semi-positive definite matrix). The efficient gain can be significant. If, for instance, $\pi=1/2$, $b_0=0.7$, $\tilde{b}_0=-0.7$, $\sigma_0=1$ and $\tilde{\sigma}_0=100$ then

$$\Sigma=\begin{pmatrix}0.4999 & 0\\ 0 & 1.5401\end{pmatrix}\quad\text{and}\quad\Sigma^G=\begin{pmatrix}0.0003 & 0\\ 0 & 0.0008\end{pmatrix}.$$

In order to show that the distribution for the shifts is material for the LSE and QLS asymptotic distribution, consider the simple case when $\Delta=\{0,2,4,\ldots\}$. This set is referred to as a set of alternating shifts. In this case, we have $\pi=1/2$ and, when

$b_0^2\tilde{b}_0^2 < 1$,

$$(3.7)\qquad \begin{cases} I = \dfrac{1}{2(1-b_0^2\tilde{b}_0^2)} \begin{pmatrix} \tilde{\sigma}_0^2(\sigma_0^2+\tilde{b}_0^2\tilde{\sigma}_0^2) & 0 \\ 0 & \sigma_0^2(\tilde{\sigma}_0^2+b_0^2\sigma_0^2) \end{pmatrix}, \\ J = \frac{1}{2(1-b_0^2\tilde{b}_0^2)} \begin{pmatrix} \tilde{\sigma}_0^2(1+\tilde{b}_0^2) & 0 \\ 0 & \sigma_0^2(1+b_0^2) \end{pmatrix}, \\ \Sigma = 2(1-b_0^2\tilde{b}_0^2) \begin{pmatrix} \dfrac{\sigma_0^2+\tilde{b}_0^2\tilde{\sigma}_0^2}{(1+\tilde{b}_0^2)^2\tilde{\sigma}_0^2} & 0 \\ 0 & \dfrac{\tilde{\sigma}_0^2+b_0^2\sigma_0^2}{(1+b_0^2)^2\sigma_0^2} \end{pmatrix}. \end{cases}$$

The above Σ is not the same as when the shifts are independent Bernoulli variables, although π takes the same value 1/2. The same remark can be made for the asymptotic variance of the QLS estimator, which is now

$$\Sigma^G = 2(1-b_0^2\tilde{b}_0^2) \begin{pmatrix} \dfrac{\sigma_0^2}{\tilde{\sigma}_0^2+\tilde{b}_0^2\sigma_0^2} & 0 \\ 0 & \dfrac{\tilde{\sigma}_0^2}{\sigma_0^2+b_0^2\tilde{\sigma}_0^2} \end{pmatrix}.$$

Once again, the QLS estimators outperform the LSE. If, for instance, $b_0 = 0.7$, $\tilde{b}_0 = -0.7$, $\sigma_0 = 1$ and $\tilde{\sigma}_0 = 100$ then

$$\Sigma = \begin{pmatrix} 0.3355 & 0 \\ 0 & 6845.9730 \end{pmatrix} \quad \text{and} \quad \Sigma^G = \begin{pmatrix} 0.0002 & 0 \\ 0 & 3.1010 \end{pmatrix}.$$

4. Numerical illustration

In this section, the previous asymptotic results are illustrated by means of simple Monte Carlo simulation experiments. We consider the time-varying MA(1) model introduced in Subsection 3.2:

$$(4.1)\qquad x_0 = \sigma_0\eta_0, \quad x_t = \sigma_t\eta_t - b_t(\theta)\sigma_{t-1}\eta_{t-1}, \qquad t = 1, 2, \ldots,$$

where $b_t(\theta) = b\mathbb{I}_\Delta(t) + \tilde{b}\mathbb{I}_{\Delta^c}(t)$, $\sigma_t^2 = \sigma^2\mathbb{I}_\Delta(t) + \tilde{\sigma}^2\mathbb{I}_{\Delta^c}(t)$, (η_t) i.i.d. $\mathcal{N}(0,1)$, and Δ is a given subset of integers. In a first set of experiments, displayed in Tables 1–2, Δ constitutes a set of Bernoulli shifts with parameter $\pi = 1/2$, as described in Subsection 3.2. The sequence of Bernoulli random variables was generated independently of the noise (η_t). In a second set of experiments, displayed in Tables 3–4, we take the set of alternating shifts $\Delta = \{0, 2, 4, \ldots\}$. For these two set of experiments, five hundred independent trajectories of size n of model (4.1) have been simulated. For each trajectory, b and $\tilde{b}$ have been estimated by LS and QLS estimators. In the algorithm presented in Section 2, Steps 2 and 3 have been repeated until the QLS estimator stabilizes (two repetitions are generally sufficient). The nuisance parameters σ and $\tilde{\sigma}$ have been estimated in the same way, but the results are not reported here for reasons of space. Replacing the unknown parameters b and $\tilde{b}$ by their LS estimates $\hat{b}_n$ and $\hat{\tilde{b}}_n$, we obtain an estimate $\hat{\Sigma}$ of the LSE asymptotic covariance matrix Σ defined in Subsection 3.2. We denote by $\mathrm{Var}_{as}(\hat{b}_n)^{1/2} := n^{-1/2}\hat{\Sigma}(1,1)^{1/2}$ the estimate of the standard deviation of $\hat{b}_n$. In order to demonstrate that this estimate, although based on the asymptotic theory, can

Table 1. LS and QLS estimates of the parameters of Model (4.1) in case of Bernoulli shifts with parameter $\pi = 1/2$.

Parameter	Method	$n=100$ MEAN	$n=100$ RMSE	$n=100$ RMSE*	$n=250$ MEAN	$n=250$ RMSE	$n=250$ RMSE*	$n=500$ MEAN	$n=500$ RMSE	$n=500$ RMSE*
		Design 1: $b=0.75$, $\tilde{b}=0.25$, $\sigma=1$, $\tilde{\sigma}=10$.								
b	LS	0.7475	0.0415	0.0377	0.7483	0.0231	0.0219	0.7496	0.0156	0.0152
	QLS	0.7493	0.0183	0.0165	0.7494	0.0114	0.0105	0.7496	0.0079	0.0075
$\tilde{b}$	LS	0.2634	0.1664	0.1383	0.2515	0.0916	0.0884	0.2509	0.0636	0.0626
	QLS	0.2509	0.0352	0.0303	0.2515	0.0188	0.0194	0.2506	0.0140	0.0137
		Design 2: $b=1.2$, $\tilde{b}=0.0$, $\sigma=1$, $\tilde{\sigma}=1$.								
b	LS	1.2139	0.0995	0.0675	1.2051	0.0568	0.0460	1.2038	0.0340	0.0329
	QLS	1.2138	0.0995	0.0675	1.2051	0.0569	0.0460	1.2038	0.0340	0.0329
$\tilde{b}$	LS	0.0051	0.0938	0.0681	0.0028	0.0514	0.0461	0.0018	0.0336	0.0329
	QLS	0.0062	0.0936	0.0674	0.0028	0.0516	0.0459	0.0017	0.0336	0.0329
		Design 3: $b=-0.5$, $\tilde{b}=-0.5$, $\sigma=1$, $\tilde{\sigma}=2$.								
b	LS	−0.5084	0.0978	0.0903	−0.5046	0.0548	0.0574	−0.5021	0.0404	0.0406
	QLS	−0.5120	0.0885	0.0799	−0.5069	0.0496	0.0511	−0.5039	0.0365	0.0362
$\tilde{b}$	LS	−0.5128	0.1675	0.1433	−0.5040	0.0973	0.0923	−0.5044	0.0667	0.0656
	QLS	−0.5035	0.1521	0.1278	−0.5018	0.0889	0.0824	−0.5049	0.0587	0.0586

MEAN corresponds to the mean of 500 estimates obtained from 500 independent realizations of the model
RMSE corresponds to the root mean squared error of estimation over the 500 replications
RMSE* corresponds to the mean of estimates of the RMSE (based on the asymptotic theory)

be successfully applied to finite samples of reasonable size, the mean of $\text{Var}_{as}(\hat{b}_n)^{1/2}$ over the 500 replications, denoted by RMSE*, has been compared to the root of the mean of $(b - \hat{b}_n)^2$ over the 500 replications, denoted by RMSE. Similar comparisons are made for the other estimators. Now, let us consider the hypothesis $H_0^{(1)} : b = 0$, $H_0^{(2)} : \tilde{b} = 0$ and $H_0^{(3)} : b = \tilde{b}$. With the LSE, $H_0^{(1)}$ (respectively $H_0^{(2)}$) is rejected when, in absolute value, $\hat{b}_n$ (respectively $\hat{\tilde{b}}_n$) is greater than 1.96 times its estimated standard deviation $\text{Var}_{as}(\hat{b}_n)^{1/2}$ (respectively $\text{Var}_{as}(\hat{\tilde{b}}_n)^{1/2}$). If the asymptotic theory applies for such sample sizes then the error of first kind should be approximately 5%. A Wald-type test is used for $H_0^{(3)}$. More precisely, $H_0^{(3)}$ is rejected when $n(\hat{b}_n - \hat{\tilde{b}}_n)^2(\hat{\Sigma}(1,1)+\hat{\Sigma}(2,2))^{-1}$ is greater than the 95%-quantile of the χ_1^2 distribution. The same tests were run with the QLS estimator.

The results reported in Tables 1–2 are in accordance with the asymptotic theory. The QLS estimator clearly outperforms the LSE when $\sigma \neq \tilde{\sigma}$ (i.e. in Designs 1 and 3). For the first and third parameter values, $(b, \tilde{b}, \sigma, \tilde{\sigma}) = (0.75, 0.25, 1, 10)$ and $(b, \tilde{b}, \sigma, \tilde{\sigma}) = (-0.5, -0.5, 1, 2)$, Table 1 shows that the RMSE's are smaller for the QLS estimator, and Table 2 shows that the tests based on the QLS estimator are more powerful that those based on the LSE. For the second parameter value, $(b, \tilde{b}, \sigma, \tilde{\sigma}) = (1.2, 0, 1, 1)$, LS and QLS estimators have very similar behaviors. This is not surprising because when $\sigma = \tilde{\sigma}$ they should have the same asymptotic behavior. Note however that, for $H_0^{(3)}$ with the third parameter value and for $H_0^{(2)}$ with the second parameter value, the rejection relative frequencies of both tests are significantly greater than the theoretical 5% for small sample sizes (for a theoretical 5% size, the empirical size on 500 independent

Table 2. Relative frequency of rejection of 5% level tests for Model (4.1) in case of Bernoulli shifts with parameter $\pi = 1/2$.

Null hypothesis	True/False	Method	$n = 100$	$n = 250$	$n = 500$
Design 1: $b = 0.75$, $\tilde{b} = 0.25$, $\sigma = 1$, $\tilde{\sigma} = 10$.					
$H_0^{(1)} : b = 0$	False	LS	100.0%	100.0%	100.0%
		QLS	100.0%	100.0%	100.0%
$H_0^{(2)} : \tilde{b} = 0$	False	LS	49.4%	79.8%	97.8%
		QLS	100.0%	100.0%	100.0%
$H_0^{(3)} : b = \tilde{b}$	False	LS	90.4%	90.4%	97.8%
		QLS	100.0%	100.0%	100.0%
Design 2: $b = 1.2$, $\tilde{b} = 0.0$, $\sigma = 1$, $\tilde{\sigma} = 1$.					
$H_0^{(1)} : b = 0$	False	LS	100.0%	100.0%	100.0%
		QLS	100.0%	100.0%	100.0%
$H_0^{(2)} : \tilde{b} = 0$	True	LS	17.6%	9.6%	5.8%
		QLS	17.4%	9.4%	5.8%
$H_0^{(3)} : b = \tilde{b}$	False	LS	100.0%	100.0%	100.0%
		QLS	100.0%	100.0%	100.0%
Design 3: $b = -0.5$, $\tilde{b} = -0.5$, $\sigma = 1$, $\tilde{\sigma} = 2$.					
$H_0^{(1)} : b = 0$	False	LS	100.0%	100.0%	100.0%
		QLS	100.0%	100.0%	100.0%
$H_0^{(2)} : \tilde{b} = 0$	False	LS	90.2%	100.0%	100.0%
		QLS	95.2%	100.0%	100.0%
$H_0^{(3)} : b = \tilde{b}$	True	LS	7.2%	6.2%	5.2%
		QLS	8.6%	5.2%	5.4%

samples is between 3% and 7% with probability 0.95). This is due to the fact that the standard deviation of the estimators are slightly underestimated by the $\mathrm{Var}_{as}(\cdot)^{1/2}$'s. The results reported in Tables 3–4 also point out the superiority of the QLS estimator over the LSE, for finite samples. The parameter estimation for alternating shifts seems however much more difficult than for Bernoulli shifts. This was to be expected because, for the three parameter values, the asymptotic covariance matrices of the estimators are greater for the alternating shifts than for the Bernoulli shifts. Even for large samples, the inference made from the asymptotic theory is not completely satisfactory. It can be seen in Table 4 that, in Design 1, the rejection relative frequency of $H_0^{(2)}$ is only 47.8% for $n = 5000$ with the LSE. In Design 2, the rejection relative frequencies of $H_0^{(2)}$ with both estimators are far from the theoretical 5%, even for $n = 500$. More surprisingly, several results worsen when the sample size increases. In Design 1, with the LSE the rejection relative frequency of $H_0^{(2)}$ decreases when n increases from $n = 100$ to $n = 500$. In Design 2, with both estimators the rejection relative frequency of $H_0^{(2)}$ deteriorates when n increases from $n = 250$ to $n = 500$. This is due to the fact that the standard deviation of the estimators are not well estimated by the $\mathrm{Var}_{as}(\cdot)^{1/2}$'s, even for large samples. To solve the problem, alternative estimators of Σ and Σ^G, such as bootstrap

Table 3. LS and QLS estimates of the parameters of Model (4.1) in case of alternating shifts.

Parameter	Method	$n=100$ MEAN	$n=100$ RMSE	$n=100$ RMSE*	$n=500$ MEAN	$n=500$ RMSE	$n=500$ RMSE*	$n=5000$ MEAN	$n=5000$ RMSE	$n=5000$ RMSE*
		Design 1: $b=0.75$, $\tilde{b}=0.25$, $\sigma=1$, $\tilde{\sigma}=10$.								
b	LS	0.7656	0.0606	0.0416	0.7514	0.0222	0.0184	0.7500	0.0052	0.0048
	QLS	0.7504	0.0148	0.0137	0.7498	0.0058	0.0062	0.7499	0.0019	0.0020
$\tilde{b}$	LS	0.2738	0.7103	0.5548	0.2324	0.4194	0.3530	0.2433	0.1263	0.1251
	QLS	0.2740	0.2283	0.1797	0.2550	0.0865	0.0819	0.2500	0.0255	0.0260
		Design 2: $b=1.2$, $\tilde{b}=0.0$, $\sigma=1$, $\tilde{\sigma}=1$.								
b	LS	1.2081	0.1478	0.1415	1.1997	0.0584	0.0632	1.1994	0.0196	0.0199
	QLS	1.2083	0.1486	0.1415	1.1997	0.0589	0.0632	1.1994	0.0197	0.0199
$\tilde{b}$	LS	0.0076	0.1064	0.0903	−0.0002	0.0427	0.0405	−0.0008	0.0126	0.0128
	QLS	0.0062	0.1082	0.0892	−0.0002	0.0428	0.0405	−0.0008	0.0126	0.0128
		Design 3: $b=-0.5$, $\tilde{b}=-0.5$, $\sigma=1$, $\tilde{\sigma}=2$.								
b	LS	−0.5102	0.0844	0.0767	−0.5028	0.0335	0.0346	−0.5005	0.0109	0.0109
	QLS	−0.5075	0.0708	0.0663	−0.5028	0.0281	0.0296	−0.5006	0.0093	0.0094
$\tilde{b}$	LS	−0.5118	0.2350	0.2182	−0.5107	0.0972	0.1004	−0.5036	0.0322	0.0320
	QLS	−0.5189	0.2061	0.1898	−0.5093	0.0858	0.0862	−0.5023	0.0270	0.0274

MEAN corresponds to the mean of 500 estimates obtained from 500 independent realizations of the model
RMSE corresponds to the root mean squared error of estimation over the 500 replications
RMSE* corresponds to the mean of estimates of the RMSE (based on the asymptotic theory)

estimators, could be investigated. This is left for future research.

5. Lemmas and proofs

First we give some remarks about Assumptions A1–A8. It is easy to show that A3 is stronger than A2. The latter is given because it is sufficient for intermediate results given below. For $\theta=\theta_0$ we have $\psi_{t,i}=0$, $\forall i\geq 1$. Therefore A1 can be interpreted as an identifiability assumption. Assumption A3 holds for instance when, for each θ and θ_0, $\psi_{t,i}(\theta,\theta_0)$ tends to zero at an exponential rate uniformly in t, as $i\to\infty$. Assumption A5 holds when the convergence of the previous sequence and of $\{\psi_{t,i}^{(k)}(\theta,\theta_0)\}_i$ holds uniformly in θ. Since $\psi_{t,0}=\sigma_t$, A5 implies that $\{\sigma_t\}_t$ is bounded. Lemma 5.1 below shows that A8 holds when

$$\exists r_0:\quad \forall r\geq r_0,\quad {}_rJ:=\lim_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sum_{i=1}^{r}\psi_{t,i}^{(\cdot)}\psi_{t,i}^{(\cdot)\prime} \tag{5.1}$$

exists and is a strictly positive definite matrix. Since for all $\lambda\in\mathbb{R}^d$ and $r>r_0$,

$$\lambda'{}_rJ\lambda=\lambda'{}_{r_0}J\lambda+\lim_{n\to\infty}n^{-1}\sum_{t=1}^{n}\sum_{i=r_0+1}^{r}\lambda'\psi_{t,i}^{(\cdot)}\psi_{t,i}^{(\cdot)\prime}\lambda\geq\lambda'{}_{r_0}J\lambda,$$

it suffices to check the strict positive-definiteness of ${}_{r_0}J$.

Now we give proofs of the Section 2 results. In the rest of this section, the letter K will be used to denote positive constants whose values are unimportant and may vary. The proofs are broken up into a series of lemmas.

Table 4. Relative frequency of rejection of 5% level tests for Model (4.1) in case of alternating shifts.

Null hypothesis	True/False	Method	$n = 100$	$n = 250$	$n = 500$	$n = 5000$
		Design 1: $b = 0.75$, $\tilde{b} = 0.25$, $\sigma = 1$, $\tilde{\sigma} = 10$.				
$H_0^{(1)} : b = 0$	False	LS	100.0%	100.0%	100.0%	100.0%
		QLS	100.0%	100.0%	100.0%	100.0%
$H_0^{(2)} : \tilde{b} = 0$	False	LS	32.8%	24.8%	19.8%	47.8%
		QLS	33.6%	57.6%	86.0%	100.0%
$H_0^{(3)} : b = \tilde{b}$	False	LS	25.0%	24.8%	19.8%	47.8%
		QLS	74.4%	57.6%	86.0%	100.0%
		Design 2: $b = 1.2$, $\tilde{b} = 0.0$, $\sigma = 1$, $\tilde{\sigma} = 1$.				
$H_0^{(1)} : b = 0$	False	LS	100.0%	100.0%	100.0%	100.0%
		QLS	100.0%	100.0%	100.0%	100.0%
$H_0^{(2)} : \tilde{b} = 0$	True	LS	11.0%	6.8%	7.4%	4.6%
		QLS	11.4%	6.8%	7.6%	4.6%
$H_0^{(3)} : b = \tilde{b}$	False	LS	100.0%	100.0%	100.0%	100.0%
		QLS	100.0%	100.0%	100.0%	100.0%
		Design 3: $b = -0.5$, $\tilde{b} = -0.5$, $\sigma = 1$, $\tilde{\sigma} = 2$.				
$H_0^{(1)} : b = 0$	False	LS	100.0%	100.0%	100.0%	100.0%
		QLS	100.0%	100.0%	100.0%	100.0%
$H_0^{(2)} : \tilde{b} = 0$	False	LS	62.4%	96.4%	99.6%	100.0%
		QLS	75.8%	99.2%	100.0%	100.0%
$H_0^{(3)} : b = \tilde{b}$	True	LS	8.4%	3.8%	5.2%	5.6%
		QLS	8.4%	4.8%	5.4%	5.6%

LEMMA 5.1. *Under* A5, (5.1) *entails* A8. *Under* A5–A8, *the matrix*

$$I = \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \sigma_t^2 \sum_{i=1}^{t} \psi_{t,i}^{(\cdot)} \psi_{t,i}^{(\cdot)\prime} = \lim_{r\to\infty} {}_rI$$

exists and is strictly positive definite.

PROOF. Let $J_n = n^{-1} \sum_{t=1}^{n} \sum_{i\geq 1} \psi_{t,i}^{(\cdot)} \psi_{t,i}^{(\cdot)\prime}$, ${}_rJ_n = n^{-1} \sum_{t=1}^{n} \sum_{i=1}^{r} \psi_{t,i}^{(\cdot)} \psi_{t,i}^{(\cdot)\prime}$ and ${}^rJ_n = J_n - {}_rJ_n$. Assumption A5 implies that $\sup_n \|{}^rJ_n\| \to 0$ as $r \to \infty$. Therefore, $({}_rJ)_r$ is a Cauchy sequence which converges to some limit J_∞. For every $\epsilon > 0$, there exists a sufficiently large integer $r = r(\epsilon) \geq r_0$ such that $\|{}_rJ - J_\infty\| < \epsilon/3$ and $\sup_n \|{}_rJ_n - J_n\| < \epsilon/3$. For n large enough,

$$\|J_n - J_\infty\| \leq \|J_n - {}_rJ_n\| + \|{}_rJ_n - {}_rJ\| + \|{}_rJ - J_\infty\| \leq \epsilon.$$

Since ϵ is chosen arbitrarily, $\lim_{n\to\infty} \|J_n - J_\infty\| = 0$. The existence of I is shown by the same arguments. Since $\lambda' J \lambda \geq (\sup_t \sigma_t^2)^{-1} \lambda' I \lambda$ and $\lambda' I \lambda \geq (\inf_t \sigma_t^2) \lambda' J \lambda$, under A5 and A6, I is invertible if and only if J is invertible. □

LEMMA 5.2. *Let*

$$Z_t = Z_t(\theta, \theta_0) = \sum_{1 \le i \ne j \le t} \psi_{t,i}\psi_{t,j}\eta_{t-i}\eta_{t-j}.$$

There exists a constant $K = K(\theta, \theta_0) > 0$ *such that, under* A2,

$$E\left\{\frac{1}{n}\sum_{t=1}^{n} Z_t\right\}^2 \le \frac{K}{n} \tag{5.2}$$

and, under A3,

$$E\left\{\frac{1}{n}\sum_{t=1}^{n} Z_t\right\}^4 \le \frac{K}{n^2} \tag{5.3}$$

PROOF. From A2, we have

$$\begin{aligned}|EZ_tZ_{t+|h|}| &= \left|\sum_{1\le i_1\ne j_1, i_2\ne j_2\le t} \psi_{t,i_1}\psi_{t,j_1}\psi_{t+|h|,i_2}\psi_{t+|h|,j_2}E\eta_{t-i_1}\eta_{t-j_1}\eta_{t+|h|-i_2}\eta_{t+|h|-j_2}\right| \\ &= 2\left|\sum_{1\le i_1\ne j_1\le t} \psi_{t,i_1}\psi_{t,j_1}\psi_{t+|h|,|h|+i_1}\psi_{t+|h|,|h|+j_1}\right| \le 2\gamma(|h|).\end{aligned}$$

Therefore, from A2,

$$E\left\{\frac{1}{n}\sum_{t=1}^{n} Z_t\right\}^2 \le \frac{2}{n^2}\sum_{|h|<n}(n-|h|)\gamma(|h|) = O(n^{-1}).$$

Thus (5.2) holds. Similarly, for $t_1 \le t_2 \le t_3 \le t_4$, we have

$$|EZ_{t_1}Z_{t_2}Z_{t_3}Z_{t_4}| = \left|\sum_{(1)}\prod_{k=1}^{4}\psi_{t_k,i_k}\psi_{t_k,j_k}E\prod_{k=1}^{4}\eta_{t_k-i_k}\eta_{t_k-j_k}\right|,$$

where $\sum_{(1)}$ denotes the sum over the indices i_k, j_k satisfying $1 \le i_k \ne j_k \le t_k$, $k = 1,2,3,4$. Most of the expectations in this sum are equal to zero. Indeed, if one index $t_k - i_k$ or $t_k - j_k$ is different from all the others, the summand vanishes. This is also the case when three indices coincide and are different from all the others. The expectation in the summand is equal to m_4^2 when $i_k = t_k - t_1 + i_1$ and $j_k = t_k - t_1 + j_1$, $k = 2,3,4$ (and in 2^3 other situations obtained by permuting some i_k and j_k). Let $\sum_{(2)}$ the sum of the $\prod_{k=1}^{4}\psi_{t_k,i_k}\psi_{t_k,j_k}E\prod_{k=1}^{4}\eta_{t_k-i_k}\eta_{t_k-j_k}$'s over all these indices. From A3, we have

$$\left|\sum_{(2)}\right| \le 8m_4^2\sum_{1\le i_1\ne j_1\le t_1}\left|\prod_{k=1}^{4}\psi_{t_k,t_k-t_1+i_1}\psi_{t_k,t_k-t_1+j_1}\right| \le K\gamma\left(\max_{k=2,3,4}|t_k - t_1|\right).$$

The expectation also equals m_4 when $t_1 - i_1 = t_2 - i_2 = t_3 - i_3 = t_4 - j_4 \ne t_1 - j_1 = t_2 - j_2 \ne t_3 - j_3 = t_4 - i_4$, and in a finite number of similar situations. Let $\sum_{(3)}$ the sum over all these indices. From A3, we have

$$\left|\sum_{(3)}\right| \leq Km_4 \sum_{1\leq i_1,j_1,j_3} |\psi_{t_1,i_1}\psi_{t_1,j_1}\psi_{t_2,t_2-t_1+i_1}\psi_{t_2,t_2-t_1+j_1} \cdot \psi_{t_3,t_3-t_1+i_1}\psi_{t_3,j_3}\psi_{t_4,t_4-t_3+j_3}\psi_{t_4,t_4-t_1+i_1}| \leq K\gamma\left(\max\left\{\max_{k=2,3,4}|t_k-t_1|, |t_4-t_3|\right\}\right).$$

The last situation is when the expectation is equal to 1. This is for instance the case when the four indices $t_1-i_1=t_2-j_2$, $t_1-j_1=t_2-i_2$, $t_3-i_3=t_4-i_4$, $t_3-j_3=t_4-j_4$ are different. Let $\sum_{(4)}$ the sum over all these indices. From A3, we have

$$\left|\sum_{(4)}\right| \leq K \sum_{1\leq i_1,j_1,i_3,j_3} |\psi_{t_1,i_1}\psi_{t_1,j_1}\psi_{t_2,t_2-t_1+j_1}\psi_{t_2,t_2-t_1+i_1} \cdot \psi_{t_3,i_3}\psi_{t_3,j_3}\psi_{t_4,t_4-t_3+i_3}\psi_{t_4,t_4-t_3+j_3}| \leq \gamma(\max\{|t_2-t_1|, |t_4-t_3|\}).$$

Since $\gamma(\cdot)$ decreases, we have shown that, for $t_1\leq t_2\leq t_3\leq t_4$,

$$|EZ_{t_1}Z_{t_2}Z_{t_3}Z_{t_4}| \leq K\gamma(\max\{|t_2-t_1|, |t_4-t_3|\}). \tag{5.4}$$

Now we have

$$E\left|\frac{1}{n}\sum_{t=1}^{n} Z_t\right|^4 \leq \frac{4!}{n^4}\sum_{1\leq t_1\leq t_2\leq t_3\leq t_4\leq n} |EZ_{t_1}Z_{t_2}Z_{t_3}Z_{t_4}| \leq \frac{4!}{n^4}\sum_{j=2}^{4}\sum\nolimits^{(j),n}|EZ_{t_1}Z_{t_2}Z_{t_3}Z_{t_4}|, \tag{5.5}$$

where $\sum^{(j),n}$ denotes the sum over the subscripts t_1,t_2,t_3,t_4 satisfying $1\leq t_1\leq t_2\leq t_3\leq t_4\leq n$ and $\max_{2\leq k\leq 4}\{t_k-t_{k-1}\}=t_j-t_{j-1}$. For fixed j and k, the number of indices $1\leq t_1\leq t_2\leq t_3\leq t_4\leq n$ satisfying $\max_{2\leq \ell\leq 4}\{t_\ell-t_{\ell-1}\}=t_j-t_{j-1}=k$ is less than nk^2. Therefore, from (5.4),

$$\sum_{j=2,4}\sum\nolimits^{(j),n}|EZ_{t_1}Z_{t_2}Z_{t_3}Z_{t_4}| \leq K\sum_{k=0}^{n-1} nk^2\gamma(k) \leq Kn^2\sum_{k=0}^{\infty}k\gamma(k),$$

and

$$\sum\nolimits^{(3),n}|EZ_{t_1}Z_{t_2}Z_{t_3}Z_{t_4}| \leq Kn\sum_{k=0}^{n-1}\sum_{i=0}^{k}\sum_{j=0}^{i}\gamma(i) \leq Kn^2\sum_{k=0}^{\infty}k\gamma(k),$$

which, in view of A3 and (5.5), entails (5.3). □

LEMMA 5.3. *Let*

$$Z_t^* = \sum_{1\leq i\leq t} (\pi_{t,i}(\theta)-\pi_{t,i}(\theta_0))x_{t-i}\eta_t.$$

There exists a constant $K > 0$ *such that, under* A2,

$$E\left\{\frac{1}{n}\sum_{t=1}^{n} Z_t^*\right\}^2 \leq \frac{K}{n} \tag{5.6}$$

and, under A3,

$$E\left\{\frac{1}{n}\sum_{t=1}^{n} Z_t^*\right\}^4 \leq \frac{K}{n^2} \tag{5.7}$$

PROOF. Let $\mathcal{F}_n$ the σ-field generated by $\{\eta_0, \dots, \eta_n\}$. From (2.1) and (2.4), this is also the σ-field generated by $\{x_0, \dots, x_n\}$. Therefore $EZ_t^* Z_{t+|h|}^* = E\eta_{t+|h|}E\sum_{1\leq i\leq t+|h|}(\pi_{t+|h|,i}(\theta) - \pi_{t+|h|,i}(\theta_0))x_{t+|h|-i}Z_t^* = 0$ when $h \neq 0$. Since $\sum_{i=1}^{t}(\pi_{t,i}(\theta) - \pi_{t,i}(\theta_0))x_{t-i} = \sum_{i=1}^{t}\psi_{t,i}\eta_{t-i}$, using A2 with $h = 0$, we obtain $E(Z_t^{*2}) = \sum_{i=1}^{t}\psi_{t,i}^2 \leq \sqrt{\gamma(0)}$. Arguing as in Lemma 5.2, we deduce (5.6). Similarly, we obtain (5.7) by the arguments used to show (5.3). □

LEMMA 5.4. *Assume* A2. *For any positive integer* q, *let*

$$W_t^{(q)} = \sum_{1\leq i\leq q} \psi_{t,i}^2(\eta_{t-i}^2 - 1).$$

Then almost surely,

$$\frac{1}{n}\sum_{t=1}^{n} W_t^{(q)} \to 0, \qquad as \quad n \to \infty.$$

PROOF. Assumption A2 implies that, for any fixed i, $\sup_t \psi_{t,i}^4 < K$. Thus

$$\sum_{t\geq 1} t^{-2}\,\mathrm{Var}\{\psi_{t,i}^2(\eta_{t-i}^2 - 1)\} < \infty.$$

Therefore the strong law of large numbers for independent random variables entails,

$$\frac{1}{n}\sum_{t=1}^{n}\psi_{t,i}^2(\eta_{t-i}^2 - 1) \to 0 \qquad \text{a.s. as} \quad n \to \infty.$$

The conclusion follows. □

LEMMA 5.5. *Under Assumptions* A1 *and* A3, *for all* $\theta \in \Theta^*$, *almost surely,*

$$\liminf_n Q_n(\theta) \geq \liminf_n Q_n(\theta_0) \qquad and \qquad \limsup_n Q_n(\theta) \geq \limsup_n Q_n(\theta_0),$$

with equalities if and only if $\theta = \theta_0$.

PROOF. We have

$$e_t(\theta) = \eta_t\sigma_t + \sum_{i=1}^{t}(\pi_{t,i}(\theta) - \pi_{t,i}(\theta_0))x_{t-i} = \eta_t\sigma_t + \sum_{i=1}^{t}\psi_{t,i}\eta_{t-i}.$$

Thus

$$Q_n(\theta) = Q_n(\theta_0) + \frac{1}{n}\sum_{t=1}^{n}\left\{\sum_{i=1}^{t}\psi_{t,i}\eta_{t-i}\right\}^2 + \frac{2}{n}\sum_{t=1}^{n}\left\{\sum_{i=1}^{t}(\pi_{t,i}(\theta)-\pi_{t,i}(\theta_0))x_{t-i}\right\}\eta_t\sigma_t. \tag{5.8}$$

Using Lemma 5.3, the Markov inequality and the Borel-Cantelli lemma,

$$\frac{1}{n}\sum_{t=1}^{n}\left\{\sum_{i=1}^{t}(\pi_{t,i}(\theta)-\pi_{t,i}(\theta_0))x_{t-i}\right\}\eta_t \to 0 \text{ almost surely as } n\to\infty. \tag{5.9}$$

Note also that A3 entails that $\sup_t \sigma_t = \sup_t \psi_{t,0} < \infty$. Therefore, from (5.8) and (5.9), for all $\theta \in \Theta^*$,

$$\liminf_n Q_n(\theta) \geq \liminf_n Q_n(\theta_0) \quad \text{and} \quad \limsup_n Q_n(\theta) \geq \limsup_n Q_n(\theta_0).$$

We have

$$\frac{1}{n}\sum_{t=1}^{n}\left\{\sum_{i=1}^{t}\psi_{t,i}\eta_{t-i}\right\}^2 = \frac{1}{n}\sum_{t=1}^{n}\sum_{1\leq i\neq j\leq t}\psi_{t,i}\psi_{t,j}\eta_{t-i}\eta_{t-j} + \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{t}\psi_{t,i}^2\eta_{t-i}^2. \tag{5.10}$$

Using (5.3) in Lemma 5.2, the Markov inequality and the Borel-Cantelli lemma,

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{1\leq i\neq j\leq t}\psi_{t,i}\psi_{t,j}\eta_{t-i}\eta_{t-j} \to 0 \text{ almost surely as } n\to\infty. \tag{5.11}$$

By A1 and Lemma 5.4, almost surely, for n large enough,

$$\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{t}\psi_{t,i}^2\eta_{t-i}^2 \geq \frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{q_0}\psi_{t,i}^2\eta_{t-i}^2 \geq \frac{1}{2}\frac{1}{n}\sum_{t=1}^{n}\sum_{i=1}^{q_0}\psi_{t,i}^2 \geq K > 0 \quad \text{when} \quad \theta\neq\theta_0. \tag{5.12}$$

The conclusion follows from (5.8), (5.9), (5.10), (5.11) and (5.12). □

LEMMA 5.6. *Under* A4 *and* A5,

$$\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta^*}|e_t(\theta)|\left\|\frac{\partial}{\partial\theta}e_t(\theta)\right\| < \infty \qquad a.s.$$

PROOF. From A5, the Cauchy-Schwarz inequality and the strong law of large numbers, almost surely,

$$\begin{aligned}
&\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sup_{\theta\in\Theta^*}|e_t(\theta)|\left\|\frac{\partial}{\partial\theta}e_t(\theta)\right\| \\
&\qquad\le K\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n}\sum_{i=0}^{t}\sum_{j=0}^{t}\bar{\gamma}_i\bar{\gamma}_j|\eta_{t-i}\eta_{t-j}| \\
&\qquad\le K\limsup_{n\to\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\bar{\gamma}_i\bar{\gamma}_j n^{-1}\sum_{t=0}^{n}\eta_t^2<\infty. \qquad\square
\end{aligned}$$

PROOF OF THEOREM 2.1. Because the $\hat{\theta}_n$'s belong to the compact set Θ^*, almost surely, there exists a subsequence $(\hat{\theta}_{n_k})$ of $(\hat{\theta}_n)$ which converges to some limit θ_∞. Since $Q_{n_k}(\cdot)$ is minimum at $(\hat{\theta}_{n_k})$, almost surely,

$$Q_{n_k}(\hat{\theta}_{n_k})\le Q_{n_k}(\theta_0),\qquad \forall k. \tag{5.13}$$

Using a Taylor expansion and Lemma 5.6, we obtain

$$\begin{aligned}
|Q_{n_k}(\hat{\theta}_{n_k})-Q_{n_k}(\theta_\infty)| &\le 2\|\hat{\theta}_{n_k}-\theta_\infty\|n_k^{-1}\sum_{t=1}^{n_k}\sup_{\theta\in\Theta^*}|e_t(\theta)|\left\|\frac{\partial}{\partial\theta}e_t(\theta)\right\| \\
&\to 0 \quad \text{a.s.}
\end{aligned} \tag{5.14}$$

as $k\to\infty$. From (5.13) and (5.14), we obtain

$$\liminf_{k\to\infty} Q_{n_k}(\theta_\infty)\le\liminf_{k\to\infty} Q_{n_k}(\theta_0) \quad\text{and}\quad \limsup_{k\to\infty} Q_{n_k}(\theta_\infty)\le\limsup_{k\to\infty} Q_{n_k}(\theta_0).$$

Lemma 5.5 still holds when $(n)_{n\ge1}$ is replaced by the increasing subsequence $(n_k)_{k\ge1}$. This entails that $\theta_\infty=\theta_0$. Since all the subsequences which converge have the same limit θ_0, the sequence $(\hat{\theta}_n)$ converges almost surely to θ_0. $\square$

LEMMA 5.7. *Under the assumptions of Theorem* 2.1 *and Assumptions* A6–A7, *the random vector* $n^{1/2}(\partial/\partial\theta)Q_n(\theta_0)$ *has a limiting normal distribution with mean* 0 *and covariance matrix* $4I$.

PROOF. The random vectors

$$n^{1/2}\frac{\partial}{\partial\theta}Q_n(\theta_0)=2n^{-1/2}\sum_{t=1}^{n}\sigma_t\eta_t\sum_{i\ge1}\psi_{t,i}^{(\cdot)}\eta_{t-i},$$

are centered. For any positive integer $r\ge r_0$, where r_0 is defined in A7, define the following truncated variables

$${}_rU_t=2\sigma_t\eta_t\sum_{i=1}^{r}\psi_{t,i}^{(\cdot)}\eta_{t-i},\qquad {}^rU_t=\frac{\partial}{\partial\theta}e_t^2(\theta_0)-{}_rU_t.$$

We have $n^{1/2}(\partial/\partial\theta)Q_n(\theta_0)=n^{-1/2}\sum_{t=1}^{n}{}_rU_t+n^{-1/2}\sum_{t=1}^{n}{}^rU_t$. The sequence $({}_rU_t)_t$ is r-dependent, but not identically distributed. Therefore, the standard central limit theorems do not directly apply. We will combine the arguments used to prove the central limit theorem for stationary r-dependent sequences with those used to prove the central limit theorem for sequences of independent but not necessarily identically distributed random variables. For $n > 2r$, each $\lambda = (\lambda_1, \ldots, \lambda_d)' \neq 0 \in \mathbb{R}^d$ and each integer k_n in $]2r, n[$, we have $n^{-1/2}\sum_{t=1}^{n} \lambda'{}_rU_t = S_n + R_n$, where

$$\begin{aligned} S_n &= n^{-1/2}\lambda'[({}_rU_1 + \cdots + {}_rU_{k_n-r}) + ({}_rU_{k_n+1} + \cdots + {}_rU_{2k_n-r}) \\ &\qquad + \cdots + ({}_rU_{([n/k_n]-1)k_n+1} + \cdots + {}_rU_{[n/k_n]k_n-r})], \\ R_n &= n^{-1/2}\lambda'[({}_rU_{k_n-r+1} + \cdots + {}_rU_{k_n}) + ({}_rU_{2k_n-r+1} + \cdots + {}_rU_{2k_n}) \\ &\qquad + \cdots + ({}_rU_{[n/k_n]k_n-r+1} + \cdots + {}_rU_{[n/k_n]k_n}) + ({}_rU_{[n/k_n]k_n} + \cdots + {}_rU_n)]. \end{aligned}$$

From A5,

$$\|\lambda'{}_rU_t\|_2 := \sqrt{E(\lambda'{}_rU_t)^2} \leq 2\sup_t \sigma_t \sum_{i=1}^{r}\sum_{j=1}^{d} |\lambda_j|\bar{\gamma}_i \leq K < \infty,$$

where K does not depend on t. Using the independence of the first $[n/k_n]$ summands into brackets in the expression of R_n, we have

$$\|R_n\|_2 \leq n^{-1/2}\sqrt{\left[\frac{n}{k_n}\right]}rK + n^{-1/2}k_nK.$$

Choose k_n such that $k_n \to \infty$ and $n^{-1/2}k_n \to 0$ as $n \to \infty$. Then R_n tends to zero in mean square, which implies that $n^{-1/2}\sum_{t=1}^{n} \lambda'{}_rU_t$ and S_n have the same asymptotic distribution (when existing). By Assumption A7, this also entails that $s_n^2 := \operatorname{Var} S_n \to 4\lambda'{}_rI\lambda > 0$. We have $S_n = \sum_{t=1}^{[n/k_n]} X_{n,t}$, where $X_{n,t} = n^{-1/2}\lambda'({}_rU_{(t-1)k_n+1} + \cdots + {}_rU_{tk_n-r})$. By previous arguments, $\|\lambda'{}_rU_t\|_4 \leq K < \infty$, where K does not depend on t. Therefore $E(X_{n,t}^4) \leq n^{-2}\{(k_n - r)K\}^4$. For each n, the random variables $X_{n,1}, X_{n,2}, \ldots$ are independent. Moreover,

$$\sum_{t=1}^{[n/k_n]} \frac{1}{s_n^4}E(X_{n,t}^4) \leq \frac{[n/k_n]}{s_n^4}n^{-2}(k_n - r)^4K^4 \to 0$$

as $n \to \infty$, provided k_n is chosen such that $n^{-1/3}k_n \to 0$. By the Lyapounov central limit theorem (see e.g. Billingsley (1995), p. 362) and the Cramer-Wold device (see e.g. Brockwell and Davis (1991), p. 204), $n^{-1/2}\sum_{t=1}^{n} {}_rU_t$ converges in law to the centered normal distribution with variance $4{}_rI$.

From A5,

$$\begin{aligned} \lim_{r\to\infty} \sup_n \operatorname{Var}\left\{n^{-1/2}\sum_{t=1}^{n} {}^rU_t\right\} &= \lim_{r\to\infty} \sup_n \left\{n^{-1}\sum_{t=1}^{n}\sigma_t^2 \sum_{i=r+1}^{t} \psi_{t,i}^{(\cdot)}\psi_{t,i}^{(\cdot)'}\right\} \\ &\leq \sup_t \sigma_t^2 \lim_{r\to\infty} \sum_{i=r+1}^{\infty} \bar{\gamma}_i^2 = 0. \end{aligned}$$

The conclusion comes from Lemma 5.1 and a standard argument (see e.g. Brockwell and Davis ((1991), Proposition 6.3.9)). □

LEMMA 5.8. *Under Assumptions* A4, A5 *and* A8,

$$\frac{\partial^2}{\partial\theta\partial\theta'}Q_n(\theta_0) \to 2J$$

in probability as $n \to \infty$.

PROOF. For $k, k' \in \{1, \dots, d\}$, we have

$$\frac{\partial^2}{\partial\theta(k)\partial\theta(k')}Q_n(\theta_0) = \frac{2}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta(k)}e_t(\theta_0)\frac{\partial}{\partial\theta(k')}e_t(\theta_0) + \frac{2}{n}\sum_{t=1}^{n}e_t(\theta_0)\frac{\partial^2}{\partial\theta(k)\partial\theta(k')}e_t(\theta_0).$$

The random variables $e_t(\theta_0)\{\partial^2 e_t(\theta_0)/\partial\theta(k)\partial\theta(k')\}$, $t = 1, 2, \dots$ are centred and uncorrelated. Moreover, in view of A5,

$$\sup_n \operatorname{Var}\left\{e_t(\theta_0)\frac{\partial^2}{\partial\theta(k)\partial\theta(k')}e_t(\theta_0)\right\} < \infty.$$

Therefore $n^{-1}\sum_{t=1}^{n} e_t(\theta_0)\{\partial^2 e_t(\theta_0)/\partial\theta(k)\partial\theta(k')\}$ converges in mean square to zero. In order to lighten the notations, write $\psi_{t,i}^{(k)}$ instead of $\psi_{t,i}^{(k)}(\theta_0, \theta_0)$. We have

$$E\frac{\partial}{\partial\theta(k)}e_t(\theta_0)\frac{\partial}{\partial\theta(k')}e_t(\theta_0) = \sum_{i\geq 1}\psi_{t,i}^{(k)}\psi_{t,i}^{(k')}$$

and

$$\begin{aligned}
&\operatorname{Cov}\left\{\frac{\partial}{\partial\theta(k)}e_t(\theta_0)\frac{\partial}{\partial\theta(k')}e_t(\theta_0), \frac{\partial}{\partial\theta(k)}e_{t+|h|}(\theta_0)\frac{\partial}{\partial\theta(k')}e_{t+|h|}(\theta_0)\right\}\\
&= \sum_{i_1,j_1,i_2,j_2\geq 1}\psi_{t,i_1}^{(k)}\psi_{t,j_1}^{(k')}\psi_{t+|h|,i_2}^{(k)}\psi_{t+|h|,j_2}^{(k')}\operatorname{Cov}(\eta_{t-i_1}\eta_{t-j_1}, \eta_{t+|h|-i_2}\eta_{t+|h|-j_2})\\
&= \sum_{i_1\neq j_1\geq 1}\psi_{t,i_1}^{(k)}\psi_{t,j_1}^{(k')}\psi_{t+|h|,i_1+|h|}^{(k)}\psi_{t+|h|,j_1+|h|}^{(k')}\\
&\quad + \sum_{i_1\neq j_1\geq 1}\psi_{t,i_1}^{(k)}\psi_{t,j_1}^{(k')}\psi_{t+|h|,j_1+|h|}^{(k)}\psi_{t+|h|,i_1+|h|}^{(k')}\\
&\quad + \sum_{i_1\geq 1}\psi_{t,i_1}^{(k)}\psi_{t,i_1}^{(k')}\psi_{t+|h|,i_1+|h|}^{(k)}\psi_{t+|h|,i_1+|h|}^{(k')}(m_4 - 1).
\end{aligned}$$

Therefore, in view of A5,

$$\begin{aligned}
&\operatorname{Var}\left(\frac{1}{n}\sum_{t=1}^{n}\frac{\partial}{\partial\theta(k)}e_t(\theta_0)\frac{\partial}{\partial\theta(k')}e_t(\theta_0)\right)\\
&\leq \frac{2}{n^2}\sum_{h=-n+1}^{n-1}(n-|h|)\sum_{i_1\geq 1}\bar{\gamma}_{i_1}\sum_{j_1\geq 1}\bar{\gamma}_{j_1}\bar{\gamma}_{i_1+|h|}\bar{\gamma}_{j_1+|h|}\\
&\quad + \frac{1}{n^2}\sum_{h=-n+1}^{n-1}(n-|h|)\sum_{i_1\geq 1}\bar{\gamma}_{i_1}^2\bar{\gamma}_{i_1+|h|}^2|m_4 - 1|
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{4}{n}\left(\sum_{i_1\ge 1}\bar{\gamma}_{i_1}\right)^2 \sum_{h\ge 1}\bar{\gamma}_h^2 + \frac{2}{n}\left(\sum_{i_1\ge 1}\bar{\gamma}_{i_1}^2\right)^2 |m_4 - 1| \\
&= O\left(\frac{1}{n}\right).
\end{aligned}
$$

The conclusion follows. □

LEMMA 5.9. *Under Assumptions* A4 *and* A5, *almost surely, for each* $i_1, i_2, i_3 \in \{1, \ldots, d\}$,

$$
\limsup_{n\to\infty} \sup_{\theta\in\Theta^*} \left| \frac{\partial^3}{\partial\theta(i_1)\partial\theta(i_2)\partial\theta(i_3)} Q_n(\theta)\right| < \infty.
$$

PROOF. The proof follows from the arguments given in the proof of Lemma 5.6, since

$$
\begin{aligned}
&\sup_{\theta\in\Theta^*}\left|\frac{\partial^3}{\partial\theta(i_1)\partial\theta(i_2)\partial\theta(i_3)}Q_n(\theta)\right| \\
&\quad\le 2n^{-1}\sum_{t=1}^{n}\left\{\sup_{\theta\in\Theta^*}\left|e_t(\theta)\frac{\partial^3}{\partial\theta(i_1)\partial\theta(i_2)\partial\theta(i_3)}e_t(\theta)\right|\right. \\
&\qquad\qquad + \sup_{\theta\in\Theta^*}\left|\frac{\partial}{\partial\theta(i_1)}e_t(\theta)\frac{\partial^2}{\partial\theta(i_2)\partial\theta(i_3)}e_t(\theta)\right| \\
&\qquad\qquad + \sup_{\theta\in\Theta^*}\left|\frac{\partial^2}{\partial\theta(i_1)\partial\theta(i_2)}e_t(\theta)\frac{\partial}{\partial\theta(i_3)}e_t(\theta)\right| \\
&\qquad\qquad \left. + \sup_{\theta\in\Theta^*}\left|\frac{\partial}{\partial\theta(i_2)}e_t(\theta)\frac{\partial^2}{\partial\theta(i_3)\partial\theta(i_1)}e_t(\theta)\right|\right\} \\
&\quad\le 8\sum_{i,j\ge 0}\bar{\gamma}_i\bar{\gamma}_j n^{-1}\sum_{t=0}^{n}\eta_t^2.
\end{aligned}
$$

□

PROOF OF THEOREM 2.2. A Taylor expansion of the criterion around θ_0 yields

$$
0 = \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\hat{\theta}_n) = \sqrt{n}\frac{\partial}{\partial\theta}Q_n(\theta_0) + \left[\frac{\partial^2}{\partial\theta(i)\partial\theta(j)}Q_n(\theta_{i,j}^*)\right]\sqrt{n}(\hat{\theta}_n - \theta_0),
$$

where the $\theta_{i,j}^*$'s are between $\hat{\theta}_n$ and θ_0. Theorem 2.1 and Lemma 5.9 entail that

$$
\begin{aligned}
&\left|\frac{\partial^2}{\partial\theta(i)\partial\theta(j)}Q_n(\theta_{i,j}^*) - \frac{\partial^2}{\partial\theta(i)\partial\theta(j)}Q_n(\theta_0)\right| \\
&\quad\le \sup_{\theta\in\Theta^*}\left\|\frac{\partial}{\partial\theta}\left\{\frac{\partial^2}{\partial\theta(i)\partial\theta(j)}Q_n(\theta_{i,j}^*)\right\}\right\| \|\hat{\theta}_n - \theta_0\| \to 0
\end{aligned}
$$

almost surely. The conclusion follows from Lemmas 5.7 and 5.8. □

LEMMA 5.10. *Suppose that the assumptions of Theorem* 2.3 *hold. Suppose also that these assumptions hold when* τ *is replaced by* τ_0. *Then we have*

$$
\Sigma^{(\tau)} \succeq \Sigma^G \quad (\textit{i.e. } \Sigma^{(\tau)} - \Sigma^G \textit{ is a nonnegative definite matrix}).
$$

PROOF. Let the random vectors

$$V_n^{(\tau)} = (J^{(\tau)})^{-1} \frac{\partial}{\partial \theta} Q_n^{(\tau)}(\theta_0).$$

From a direct extension of Lemma 5.7, the asymptotic variance of $V_n^{(\tau)}$ is $\mathrm{Var}_{as} V_n^{(\tau)} = 4\Sigma^{(\tau)}$. Moreover

$$\begin{aligned}
&\mathrm{Cov}_{as}(V_n^{(\tau)}, V_n^{(\tau_0)}) \\
&= (J^{(\tau)})^{-1} \lim_n \mathrm{Cov}\left(\frac{2}{n}\sum_{t=1}^{n} \tau_t \sigma_t \eta_t \sum_{i=1}^{t} \psi_{t,i}^{(\cdot)} \eta_{t-i}, \frac{2}{n}\sum_{s=1}^{n} \sigma_s^{-1} \eta_s \sum_{j=1}^{s} \psi_{s,j}^{(\cdot)\prime} \eta_{s-j}\right) \\
&\quad \cdot (J^{(\tau_0)})^{-1} \\
&= 4(J^{(\tau)})^{-1}(J^{(\tau)})(J^{(\tau_0)})^{-1} = 4\Sigma^{(\tau_0)}.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathrm{Var}_{as}(V_n^{(\tau)} - V_n^{(\tau_0)}) \\
&= \mathrm{Var}_{as} V_n^{(\tau)} + \mathrm{Var}_{as} V_n^{(\tau_0)} - \mathrm{Cov}_{as}(V_n^{(\tau)}, V_n^{(\tau_0)}) - \mathrm{Cov}_{as}(V_n^{(\tau_0)}, V_n^{(\tau)}) \\
&= 4(\Sigma^{(\tau)} - \Sigma^{(\tau_0)})
\end{aligned}$$

is nonnegative definite. □

PROOF OF THEOREM 2.4. The consistency is obtained as in the proof of Theorem 2.1. Since the objective function depends on the estimate of the nuisance parameter, write $Q_n(\theta, \hat{\beta}_n)$ instead of $Q_n^{(\hat{\tau}_0)}(\theta)$. In view of A5 and A6, let $m = \inf_t \sigma_t(\beta_0) > 0$ and $M = \sup_t \sigma_t(\beta_0) < \infty$. Almost surely, for all sufficiently small $\varepsilon > 0$ and sufficiently large n, we have

$$\sup_t |\sigma_t(\hat{\beta}_n) - \sigma_t(\beta_0)| < \varepsilon \quad \text{and} \quad m - \varepsilon > 0.$$

From A5, we have

$$\begin{aligned}
&n^{1/2} \left\| \frac{\partial}{\partial \theta} Q_n(\theta_0, \hat{\beta}_n) - \frac{\partial}{\partial \theta} Q_n^{(\tau_0)}(\theta_0) \right\| \\
&= 2n^{-1/2} \left\| \sum_{t=1}^{n} \left(\frac{1}{\sigma_t^2(\hat{\beta}_n)} - \frac{1}{\sigma_t^2(\beta_0)} \right) \eta_t \sum_{i=1}^{t} \psi_{t,i}^{(\cdot)} \eta_{t-i} \right\| \\
&\leq 2 \frac{\varepsilon(\varepsilon + 2M)}{m^2(m-\varepsilon)^2} \sum_{i=1}^{\infty} \bar{\gamma}_i n^{-1/2} \sum_{t=1}^{n} \eta_t^2 := K(\varepsilon) n^{-1/2} \sum_{t=1}^{n} \eta_t^2.
\end{aligned}$$

The Markov inequality entails that, for all $\varepsilon_0 > 0$,

$$P\left(K(\varepsilon) n^{-1/2} \sum_{t=1}^{n} \eta_t^2 > \varepsilon_0\right) \leq \frac{K(\varepsilon)^2 m_4}{\varepsilon_0^2}.$$

The right hand side of the previous inequality can be made arbitrarily small by choosing ε sufficiently small. Thus $n^{1/2} \|(\partial/\partial\theta) Q_n(\theta_0, \hat{\beta}_n) - (\partial/\partial\theta) Q_n^{(\tau_0)}(\theta_0)\| \to 0$ in probability as

$n \to \infty$. Arguing as in the proof of Lemma 5.7, it can be shown that $n^{1/2}(\partial/\partial\theta)Q_n^{(\tau_0)}(\theta_0)$ has a limiting normal distribution with mean 0 and covariance matrix $4I^{(\tau_0)}$. We deduce that $n^{1/2}(\partial/\partial\theta)Q_n(\theta_0, \hat{\beta}_n)$ has the same asymptotic distribution. By similar arguments, it can be shown that

$$|(\partial^2/\partial\theta(i)\partial\theta(j))Q_n(\theta_0, \hat{\beta}_n) - (\partial^2/\partial\theta(i)\partial\theta(j))Q_n^{(\tau_0)}(\theta_0)| \to 0$$

in probability as $n \to \infty$. The proof is completed as in the proof of Theorem 2.2. □

Acknowledgements

We would like to thank Guy Mélard for giving us a copy of his paper jointly written with Rajae Azrak. We wish to thank the referees for helpful suggestions and comments. We are also very grateful to the referees for indicating us several major references.

References

Adams, G. J. and Goodwin, G. C. (1995). Parameter estimation for periodic ARMA models, *J. Time Ser. Anal.*, **16**, 127–146.

Alpay, D., Freydin, B. and Loubaton, Ph. (2000). An extension problem for discrete-time almost periodically correlated stochastic processes, *Linear Algebra Appl.*, **308**, 163–181.

Alpay, D., Chevreuil, A. and Loubaton, Ph. (2001). An extension problem for discrete-time periodically correlated stochastic processes, *J. Time Ser. Anal.*, **22**, 1–11.

Anderson, P. L. and Vecchia, A. V. (1983). Asymptotic results for periodic autoregressive moving-average processes, *J. Time Ser. Anal.*, **14**, 1–18.

Azrak, R. and Mélard, G. (2000). Asymptotic properties of quasi-likelihood estimators for ARMA models with time-dependent coefficients, unpublished working paper.

Basawa, I. V. and Lund, R. B. (2001). Large sample properties of parameter estimates for periodic ARMA models, *J. Time Ser. Anal.*, **22**, 551–663.

Bentarzi, M. and Hallin, M. (1993). On the invertibility of periodic moving-average models, *J. Time Ser. Anal.*, **15**, 262–268.

Billingsley, P. (1995). *Probability and Measure*, Wiley, New York.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, Springer, New York.

Cochrane, D. and Orcutt, D. H. (1949). Applications of least squares regression to relationships containing autocorrelated error terms, *J. Amer. Statist. Assoc.*, **44**, 32–61.

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

de Gooijer, J. G. (1998). On threshold moving-average models, *J. Time Ser. Anal.*, **19**, 1–18.

Dehay, D. and Monsan, V. (1996). Random sampling estimation for almost periodically correlated processes, *J. Time Ser. Anal.*, **17**, 425–445.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root, *J. Amer. Statist. Assoc.*, **74**, 427–431.

Gladyshev, E. G. (1963). Periodically and almost-periodically correlated random processes with continuous time parameter, *Theory Probab. Appl.*, **8**, 173–177.

Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation, *International Statistical Review*, **55**, 231–244.

Gouriéroux, C. and Monfort, A. (1995). *Statistics and Econometric Models*, Cambridge University Press, Cambridge.

Hallin, M. (1986). Non-stationary q-dependent processes and time-varying moving-average models: Invertibility properties and the forecasting problem, *Adv. in Appl. Probab.*, **18**, 170–210.

Hamilton, J. D. (1994). *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Kowalski, 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**, 297–309.

Lund, R. B. and Basawa, I. V. (2000). Recursive prediction and likelihood evaluation for periodic ARMA models, *J. Time Ser. Anal.*, **21**, 75–93.

Lund, R. B., Basawa, I. V. and Shao, Q. (2001). Parsimonious periodic time series modeling, Tech. Report 2001-12, Department of Statistics, University of Georgia.

Makagon, A. and Miamee, A. G. (1996). Weak law of large numbers for almost periodically correlated processes, *Proc. Amer. Math. Soc.*, **124**, 1899–1902.

Priestley, M. B. (1965). Evolutionary spectra and nonstationary processes, *J. Roy. Statist. Soc. Ser. B*, **27**, 204–237.

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

Tiao, G. C. and Grupe, M. R. (1980). Hidden periodic autoregressive-moving average models in time series data, *Biometrica*, **67**, 365–373.

Tjøstheim, D. and Paulsen, J. (1985). Least squares estimates and order determination procedures for autoregressive processes with a time dependent variance, *J. Time Ser. Anal.*, **6**, 117–133.

Tyssedal, J. S. and Tjøstheim, D. (1982). Autoregressive processes with a time dependent variance, *J. Time Ser. Anal.*, **3**, 209–217.

Vecchia, A. V. (1985). Periodic autoregressive-moving average (PARMA) modeling with applications to water resources, *Water Resources Bulletin*, **21**, 721–730.

White, S. J. (1958). The limiting distribution of serial correlation coefficient in the explosive case, *Ann. Math. Statist.*, **29**, 1188–1197.

Whittle, P. (1965). Recursive relations for predictors of nonstationary processes, *J. Roy. Statist. Soc. Ser. B*, **27**, 523–532.