Asymptotic Properties of QML Estimators for VARMA Models with Time-Dependent Coefficients

Abdelkamer Alj
Université de Moulay Ismail, Meknès

Rajae Azrak
Université Mohammed V, Rabat

Christophe Ley
Ghent University

Guy Mélard
ECARES, SBS-EM, Université libre de Bruxelles

December 2016

ECARES working paper 2016-41
Asymptotic properties of QML estimators for VARMA models with time-dependent coefficients

Abdelkamel ALJ
Faculté des Sciences juridiques, économiques et sociales,
Université Moulay Ismail, Meknès, Morocco,
Rajae AZRAK
Faculté des Sciences juridiques, économiques et sociales,
Université Mohammed V, Rabat, Morocco,
Christophe LEY
Department of Applied Mathematics, Computer Science and Statistics,
Ghent University, Ghent, Belgium
and Guy MÉLARD
Faculty SBS-EM & ECARES,
Université libre de Bruxelles, Brussels, Belgium

November 23, 2016

Abstract

This paper is about vector autoregressive-moving average (VARMA) models with time-dependent coefficients to represent non-stationary time series. Contrary to other papers in the univariate case, the coefficients depend on time but not on the series’ length n. Under appropriate assumptions, it is shown that a Gaussian quasi-maximum likelihood estimator is almost surely consistent and asymptotically normal. The theoretical results are illustrated by means of two examples of bivariate processes. It is shown that the assumptions underlying the theoretical results apply. In the second example the innovations are marginally heteroscedastic with a correlation ranging from −0.8 to 0.8. In the
two examples, the asymptotic information matrix is obtained in the Gaussian case. Finally, the finite-sample behavior is checked via a Monte Carlo simulation study for $n$ from 25 to 400. The results confirm the validity of the asymptotic properties even for short series and the asymptotic information matrix deduced from the theory.

Key words and phrases: non-stationary process; multivariate time series; time-varying models.
Running title: Asymptotics of QMLEs for tdVARMA models
1 Introduction

A large part of the literature on time series models is concerned with stationary models. This is of course due to the ensuing mathematical simplifications of stationarity. Even in that simple context, an asymptotic analysis is not necessarily easy; indeed, the celebrated autoregressive-moving average (ARMA) models popularized by Box and Jenkins (Box et al., 2015) require several pages in Brockwell & Davis (1991, pp. 375-396) for the derivation of their asymptotic properties. However, the assumption of invariance over time (especially for long time intervals) is difficult to justify in most practical situations. Therefore, recent years have seen an increasing interest in models with time-dependent coefficients and non-stationary time series. Initiated by the seminal work of Quenouille (1957), models with time-dependent or time-varying coefficients for univariate time series have been investigated over the years by, inter alia, Whittle (1965), Subba Rao (1970), Tjostheim (1984), Kwoun & Yajima (1986), Priestley (1988), Dahlhaus (2000), Bibi & Francq (2003), and Azrak & Mélard (2006). We refer to the introduction of Azrak & Mélard (2006) or of Van Bellegem & Dahlhaus (2006) for further references. In several of these papers, the coefficients of the ARMA models are not constant but are deterministic functions of time. Also the innovation variance can be a deterministic function of time instead of being constant, such as in Van Bellegem & von Sachs (2004). We can speak of marginal heteroscedasticity as opposed to conditional heteroscedasticity which is encountered in ARCH and GARCH models. All these functions of time are supposed to depend on a small number of parameters.

The present paper inscribes itself in this line of research but for multivariate time series. The generality of the models we consider evidently entails numerous challenges, since the convenient asymptotic theory of stationary ergodic processes does no longer apply. Also, the asymptotic theory of time series models makes a large use of Fourier transforms and, hence, of what is called spectral analysis. As a consequence, deriving conditions for consistency and asymptotic normality of estimators of the coefficients, as well as obtaining the asymptotic covariance matrix, becomes highly complicated.

We consider multidimensional time series models, with particular emphasis on vector ARMA (VARMA) models, the multivariate extension of the ARMA models. The main difference between ARMA and VARMA models lies in the fact that the coefficients change from scalars to square matrices. The main developments in the area of statistical inference of standard stationary VARMA models are due to Kohn (1978), Hannan & Deistler (1988), and Yao & Brockwell (2006), and quite recently, Boubacar
Mainassara & Francq (2011) who study the consistency and asymptotic normality of quasi-maximum likelihood estimators for weak VARMA models. However, the field of time-dependent VARMA (tdVARMA) models with marginally heteroscedastic innovation covariance matrix remains largely unexplored. An exception is Dahlhaus (2000) using an entirely different approach and assuming that the coefficients depend on time $t$ but also on the length of the series $n$ through their ratio $t/n$. Here we assume dependency on $t$ only. Although our theory is illustrated on pure VAR examples, it should be emphasized that it is valid for VMA and VARMA models. Note that Lütkepohl (2005, Chap. 14) treats tdVAR models by Gaussian maximum likelihood but does not discuss asymptotic properties in the general case.

Thus we want to fill in this gap in the literature by extending to the multivariate setting the methodology of Azrak & Mélard (2006) who, to the best of the authors’ knowledge, were the first to obtain asymptotic properties of estimators for the general class of univariate time-dependent ARMA models by having recourse to quasi-maximum likelihood estimation (QMLE). Like other QMLE approaches, the estimation method in Azrak & Mélard (2006) does not use the true, unknown, density of the observations but rather acts as if that density were Gaussian, thus using the Gaussian log-likelihood, which is an extension of the generalized least-squares method since it takes care of possible heteroscedasticity. There is no assumption of stationarity but, although it is not illustrated in our examples, there is an adjustment in the asymptotic theory for allowing non-normal observations. Existence of eight-order moments is assumed. One major advantage of QMLE is that the Gaussian likelihood function can be computed exactly, with an efficient algorithm, see Alj et al. (2016), and this is very important for short time series. The main task in the Azrak-Mélard approach, hence also in our extension, consists in checking conditions from two crucial theorems in Klimko & Nelson (1978), which respectively ensure existence of an almost surely (a.s.) consistent estimator and prove asymptotic normality of that estimator, whilst providing the asymptotic covariance matrix. This is precisely what we are aiming at, but, as we shall see in the rest of this paper, it is all but an easy task.

Another related approach, see for instance Tiao & Grupe (1980) or Basawa & Lund (2001), consists in ARMA models with coefficients that vary as periodic functions of time. If the period $s$ is an integer, $s$ consecutive variables can be stacked as a vector which satisfies a stationary VARMA model. Here we do not assume periodic coefficients although, to simplify the derivations, our two examples will have
periodic coefficients or innovation covariance matrix, but with large or irrational periods. Therefore stacking the variables will not be feasible in practice and standard asymptotic theory for stationary VARMA models will not apply.

Another aspect of the present paper is that it provides an alternative theory for the asymptotics of standard VARMA models that does not rely on stationarity or ergodicity arguments, although in the standard case, the assumptions will imply the usual conditions on the roots of the autoregressive and moving average polynomials in the lag operator. Our alternative theory also does not require spectral analysis.

The paper is organized as follows. In Section 2 we first develop asymptotics for quasi-maximum likelihood estimators in a general multivariate time series model which is not necessarily stationary. Then, in Section 3, we focus our attention on tdVARMA models: after setting the notations, we analyze pure VAR and pure VMA representations, with an illustration, before finally stating the main theorem for the tdVARMA case. We illustrate our theoretical findings by means of two examples in Section 4 and examine the finite-sample behavior of our estimators via a Monte Carlo simulation study in Section 5. Appendix A collects the main proofs. In online Supporting Information, we further add two appendices: Appendix S1 contains a verification of the main assumptions for the two examples studied with a few nice mathematical derivations, and Appendix S2 contains the main theoretical technicalities underpinning the proofs of this paper.

2 QMLE for a general multivariate time series model

2.1 Some preliminaries

Let \( \{x_t : t \in \mathbb{N}\} \) be a stochastic process defined on a probability space \((\Omega, F, P_\theta)\), taking values in \(\mathbb{R}^r\), and whose distribution depends on a vector \(\theta = (\theta_1, \ldots, \theta_m)^T\) of unknown parameters to be estimated, with \(\theta\) lying in some open set \(\Theta\) of a Euclidean space \(\mathbb{R}^m\). Let \(E_\theta(.)\) and \(E_\theta(.|.)\) denote expectation and conditional expectation under \(P_\theta\), respectively. The true value of \(\theta\) is denoted by \(\theta^0 = (\theta_1^0, \ldots, \theta_m^0)^T\), assumed to be an interior point of \(\Theta\). Let \(\{F_t : t \in \mathbb{N}\}\) be an increasing sequence of sub-sigma algebras of \(F\) with \(F_t\) generated by \(\{x_u : u = 1, 2, \ldots, t\}\) with \(F_0 = \{\emptyset, \Omega\}\) so that, for each \(t\), \(x_t\) is measurable with respect to \(F_t\). Given a set of observations \(\{x_t : t = 1, 2, \ldots, n\}\), we want to estimate \(\theta\) by trying to minimize the general real-valued objective function \(Q_n(\theta) = Q_n(\theta; x_1, \ldots, x_n)\) which depends on \(\theta\) and the observations \(\{x_t : t = 1, 2, \ldots, n\}\). Therefore we solve the system of equations \(\partial Q_n(\theta)/\partial \theta_i = 0\) for \(i = 1, \ldots, m\).
We suppose that the objective function $Q_n(\theta)$ is twice continuously differentiable in $\theta$. Let $\hat{\theta}_n = (\hat{\theta}_1, ..., \hat{\theta}_m)^T$ be a sequence of estimators indexed by $n$. Klimko & Nelson (1978) showed conditions for strong consistency and asymptotic normality of $\hat{\theta}_n$, see also Hall & Heyde (1980, pp. 174-176) and Taniguchi & Kakizawa (2000, pp. 97-98).

### 2.2 General theory of quasi-maximum likelihood estimation

Denote

$$e_t(\theta) = x_t - \hat{x}_{t-1}(\theta) \quad \text{with} \quad \hat{x}_{t-1}(\theta) = E_\theta(x_t/F_{t-1}),$$

for which obviously $E_\theta(e_t(\theta)) = 0$. We denote by $\Sigma_t(\theta) = E_\theta [e_t(\theta)e_t^T(\theta)/F_{t-1}]$ the conditional covariance matrix given $F_{t-1}$. The quasi-likelihood function $L_n(\theta; x_1, ..., x_n)$ computed as if the process were Gaussian is given by

$$L_n(\theta; x_1, ..., x_n) = (2\pi)^{-nr/2} \prod_{t=1}^n \det (\Sigma_t(\theta))^{-1/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^n e_t^T(\theta)\Sigma_t^{-1}(\theta)e_t(\theta) \right\}.$$  

We take the objective function $Q_n(\theta) = -\log (L_n(\theta; x_1, ..., x_n)) = (1/2)\sum_{t=1}^n \alpha_t(\theta) + (rn/2) \log(2\pi)$, with $\alpha_t(\theta) = \log (\det (\Sigma_t(\theta))) + e_t^T(\theta)\Sigma_t^{-1}(\theta)e_t(\theta)$. Then the QMLE of $\theta$ is defined as any measurable solution $\hat{\theta}_n$ of arg min$_{\hat{\theta} \in \Theta} Q_n(\theta)$.

In order to check the assumptions of the Klimko & Nelson (1978) theorems, we proceed like Azrak & Mélard (2006) and we make some additional assumptions as follows. Let the $r$-vector stochastic process $\{x_t : t \in \mathbb{N}\}$ be such that $E_\theta (||x_t||^8) < \infty$, where $||.||$ is the Euclidean norm, for all $\theta$ and $e_t(\theta)$ and $\Sigma_t(\theta)$ are almost surely twice continuously differentiable in $\theta$. Henceforth, for simplicity, we denote $[E_{\theta} \{ \cdot (\theta) \}]_{\theta = \theta^0}$ by $E_{\theta^0} \{ \cdot (\theta) \}$. We suppose that there exist two positive constants $C_1$ and $C_2$ such that for all $t \geq 1$:

$$H_{1.1} \quad \mathbb{E}_{\theta^0} \left\{ \left| \frac{\partial \alpha_t(\theta)}{\partial \theta_i} \right|^4 \right\} \leq C_1 \text{ for } i = 1, ..., m;$$

$$H_{1.2} \quad \mathbb{E}_{\theta^0} \left\{ \left| \frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j} - \mathbb{E}_{\theta} \left( \frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j} / F_{t-1} \right) \right|^2 \right\} \leq C_2 \text{ for } i, j = 1, ..., m.$$  

Suppose further that

$$H_{1.3} \quad \lim_{n \to \infty} \frac{1}{2n} \sum_{t=1}^n \mathbb{E}_{\theta^0} \left\{ \left| \frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j} / F_{t-1} \right|^2 \right\} = V_{ij} \quad \text{a.s. for } i, j = 1, ..., m,$$

where $V = (V_{ij})_{1 \leq i, j \leq m}$ is a strictly positive definite matrix of constants;
\[ H_{1.4} \]
\[
\lim_{n \to \infty} \sup_{\Delta \downarrow 0} (n \Delta)^{-1} \left| \sum_{t=1}^{n} \left( \left\{ \frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j} \right\}_{\theta = \theta^*_i j} - \left\{ \frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j} \right\}_{\theta = \theta^0} \right) \right| < \infty \quad \text{a.s.}
\]
for \( i, j = 1, \ldots, m \), where \( \theta^*_i j \) is a point of the straight line joining \( \theta^0 \) to every \( \theta \), such that \( \| \theta - \theta^0 \| < \Delta, 0 < \Delta \);

\[ H_{1.5} \]
\[
\frac{1}{n} \sum_{t=1}^{n} E_{\theta^0} \left( \frac{\partial \alpha_t(\theta)}{\partial \theta} \frac{\partial \alpha_t(\theta)}{\partial \theta^T} / F_{t-1} \right) - \frac{1}{n} \sum_{t=1}^{n} E_{\theta^0} \left( \frac{\partial \alpha_t(\theta)}{\partial \theta} \frac{\partial \alpha_t(\theta)}{\partial \theta^T} \right) \xrightarrow{a.s.} 0 \quad \text{as } n \to \infty;
\]

\[ H_{1.6} \]
\[
\lim_{n \to \infty} \frac{1}{4n} \sum_{t=1}^{n} E_{\theta^0} \left( \frac{\partial \alpha_t(\theta)}{\partial \theta_i} \frac{\partial \alpha_t(\theta)}{\partial \theta_j} \right) = W_{ij}, \quad i, j = 1, \ldots, m, \quad (2.2)
\]
where \( W = (W_{ij})_{1 \leq i, j \leq m} \) is a positive definite matrix of constants.

**Theorem 1.** Suppose that assumptions \( H_{1.1} - H_{1.4} \) hold. Then there exists a sequence of estimators \( \hat{\theta}_n = (\hat{\theta}_1, \ldots, \hat{\theta}_m)^T \) such that \( \hat{\theta}_n \to \theta^0 \) almost surely as \( n \to \infty \). If furthermore assumptions \( H_{1.5} - H_{1.6} \) hold, then \( n^{1/2} (\hat{\theta}_n - \theta^0) \overset{L}{\to} N(0, V^{-1}WV^{-1}) \) as \( n \to \infty \).

Theorem 1 is similar to Theorem 1 in Azrak & Mélard (2006) except that \( \alpha_t(\theta) \) is here a scalar product instead of a square. Therefore, we need Lemmas 4.4 and 4.5 in Appendix S2 in order to use a strong law of large numbers for martingale sequences.

### 3 VARMA models with time-dependent coefficients

#### 3.1 tdVARMA models: definition and notations

The process \( \{x_t: t \in \mathbb{N}\} \) is called a zero mean \( r \)-vector mixed autoregressive-moving average process of order \( (p, q) \) with time-dependent coefficients, and is denoted by tdVARMA \( (p, q) \), if and only if it satisfies the equation

\[
x_t = \sum_{i=1}^{p} A_{ti} x_{t-i} + g_t \epsilon_t + \sum_{j=1}^{q} B_{tj} g_{t-j} \epsilon_{t-j}, \quad (3.1)
\]

where \( p \) and \( q \) are integer constants, \( \{\epsilon_t: t \in \mathbb{N}\} \) is an independent white noise process, consisting of independent random variables, not necessarily identically dis-
tributed, with zero mean, covariance matrix $\Sigma$ which is invertible, and finite eighth-order moments, and where the coefficients $A_{t_1}, ..., A_{t_p}$ and $B_{t_1}, ..., B_{t_q}$, as well as $g_t$, are $r \times r$ matrices and deterministic functions of time $t$. The initial values $x_t, t < 1$, and $\epsilon_t, t < 1$, are supposed to be equal to the zero vector. In the sequel, we will also use $A_{t_0} = B_{t_0} = I_r$, with $I_r$ the $r$-dimensional identity matrix and set to zero the coefficients $A_{tk}$ with $k > p$ and $B_{tk}$ with $k > q$, for all $t$. Writing $\otimes$ for the Kronecker product, we let $\kappa_t = E \left( \text{vec}(\epsilon_t \epsilon_t^T) \text{vec}(\epsilon_t \epsilon_t^T)^T \right) = E \left( (\epsilon_t \epsilon_t^T) \otimes (\epsilon_t \epsilon_t^T) \right)$, which depends on $t$, in general. For $k, l \in \mathbb{N}$ and $k \neq l$, we consider the matrix $E \left( \text{vec}(\epsilon_{t-k} \epsilon_{t-l}) \text{vec}(\epsilon_{t-t} \epsilon_{t-k})^T \right) = K_{r,r}(\Sigma \otimes \Sigma)$, which does not depend on $t, k$ or $l$ and where the $r^2 \times r^2$ matrix $K_{r,r}$ is a commutation matrix, see Kollo & von Rosen (2005, p. 79) or Lemma 4.2 of Appendix S2.

Let us now consider the parametric model corresponding to (3.1), namely

$$x_t = \sum_{i=1}^{p} A_{ti}(\theta)x_{t-i} + e_t(\theta) + \sum_{j=1}^{q} B_{tj}(\theta)e_{t-j}(\theta), \quad (3.2)$$

where the $e_t(\theta)$ can be considered as the residuals of the model and are defined as in (2.1) and where $A_{ti}(\theta)$, for $i = 1, ..., p$, and $B_{tj}(\theta)$, for $j = 1, ..., q$, are the parametric coefficients. Furthermore the covariance matrix of $e_t(\theta)$ is parametrized as $\Sigma_t(\theta) = \text{def} E_\theta \left( e_t(\theta) e_t^T(\theta) \right) = g_t(\theta) \Sigma g_t^T(\theta)$. For $\theta = \theta^0$, we have $A_{ti}(\theta^0) = A_{ti}$, $B_{tj}(\theta^0) = B_{tj}$, $g_t(\theta^0) = g_t$, $e_t(\theta^0) = e_t$, and $\Sigma_t = \text{def} \Sigma_t(\theta^0) = E \left( e_t(\theta^0) e_t^T(\theta^0) \right) = g_t \Sigma g_t^T$. We assume that the $m$-dimensional vector $\theta$ contains all the parameters of interest to be estimated, those in the coefficients $A_{t_1}(\theta), ..., A_{tp}(\theta), B_{t_1}(\theta), ..., B_{tq}(\theta)$ and $g_t(\theta)$ but not the nuisance parameters in the scale factor matrix $\Sigma$ which are estimated separately. In usual VARMA($p, q$) models, the coefficients $A_1(\theta), ..., A_p(\theta), B_1(\theta), ..., B_q(\theta)$ do not depend on $t$, and the parameters are the coefficients themselves and $g_t(\theta)$ is unnecessary. Note that for a given $\theta$ we have

$$\hat{x}_{t,t-1}(\theta) = E_\theta(x_t/F_{t-1}) = \sum_{i=1}^{p} A_{ti}(\theta)x_{t-i} + \sum_{j=1}^{q} B_{tj}(\theta)e_{t-j}(\theta).$$

According to the assumptions made about initial values, it is possible to write out properly the pure autoregressive and the pure moving average representation of the model (3.2), as we shall see in the next section.
3.2 The pure autoregressive and the pure moving average representations

By using the assumption about initial values and using (3.2) recurrently, Mélard (1985) has established expressions for the pure autoregressive representation and the pure moving average representation of tdVARMA processes. In the univariate case these representations can be found in Azrak & Mélard (2006). In our setting, for any \( \theta \), the pure autoregressive representation corresponds to

\[
x_t = \sum_{k=1}^{t-1} \pi_{tk}(\theta) x_{t-k} + e_t(\theta),
\]

(3.3)

where the coefficients \( \pi_{tk}(\theta) \) can be obtained from the autoregressive and moving average coefficients by using the following recurrences (see Mélard, 1985, pp. 43-45):

\[
\pi_{t0}(\theta) = I_r, \quad \pi_{tj}^{(0)}(\theta) = A_{tj}(\theta), \quad \tilde{\pi}_{tj}^{(0)}(\theta) = B_{tj}(\theta), \quad \text{for} \quad j = 1, \ldots, t - 1,
\]

\[
\pi_{tk}^{(k)}(\theta) = \pi_{tj}^{(k-1)}(\theta) - \tilde{\pi}_{tk}^{(k-1)}(\theta) A_{t-k,j-k}(\theta), \quad \text{for} \quad j = k, \ldots, t - 1,
\]

\[
\tilde{\pi}_{tj}^{(k)}(\theta) = \tilde{\pi}_{tj}^{(k-1)}(\theta) - \tilde{\pi}_{tk}^{(k-1)}(\theta) B_{t-k,j-k}(\theta), \quad \text{for} \quad j = k + 1, \ldots, t - 1,
\]

and \( \pi_{tk}(\theta) = \pi_{tk}^{(k)}(\theta) \) for \( k = 1, \ldots, t - 1 \). By (3.3) we have \( e_t(\theta) = x_t - \sum_{k=1}^{t-1} \pi_{tk}(\theta) x_{t-k} \), and its first derivative with respect to \( \theta_i, \ i = 1, \ldots, m \), is given by

\[
\frac{\partial e_t(\theta)}{\partial \theta_i} = -\sum_{k=1}^{t-1} \frac{\partial \pi_{tk}(\theta)}{\partial \theta_i} x_{t-k}.
\]

(3.4)

On the other hand, for the pure moving average representation we have

\[
x_t = e_t(\theta) + \sum_{k=1}^{t-1} \psi_{tk}(\theta) e_{t-k}(\theta),
\]

(3.5)

where the coefficients \( \psi_{tk}(\theta) = \psi_{tk}^{(k)}(\theta), \ k = 0, 1, \ldots, t - 1 \), can be obtained from the autoregressive and moving average coefficients by using the following recurrences (see Mélard, 1985, pp. 36-38):

\[
\psi_{t0}^{(0)}(\theta) = I_r, \quad \psi_{tj}^{(0)}(\theta) = B_{tj}(\theta), \quad \tilde{\psi}_{tj}^{(0)}(\theta) = A_{tj}(\theta), \quad \text{for} \quad j = 1, \ldots, t - 1,
\]

\[
\psi_{tk}^{(k)}(\theta) = \psi_{tj}^{(k-1)}(\theta) + \tilde{\psi}_{tk}^{(k-1)}(\theta) B_{t-k,j-k}(\theta), \quad \text{for} \quad j = k, \ldots, t - 1,
\]

\[
\tilde{\psi}_{tj}^{(k)}(\theta) = \tilde{\psi}_{tj}^{(k-1)}(\theta) + \tilde{\psi}_{tk}^{(k-1)}(\theta) A_{t-k,j-k}(\theta), \quad \text{for} \quad j = k + 1, \ldots, t - 1,
\]
for each \( k = 1, ..., t - 1 \), and \( \psi_{tk}(\theta) = 0 \), \( k \geq t \). Hence, denoting \( \psi_{tk} = \psi_{tk}(\theta^0) \) and \( \psi_{t0} = I_r \), we have

\[
x_t = \sum_{k=0}^{t-1} \psi_{tk} g_{t-k} \epsilon_{t-k}, \quad E(x_t x_t^T) = \sum_{k=0}^{t-1} \psi_{tk} \Sigma_{t-k} \psi_{tk}^T.
\] (3.6)

By using (3.5), \( e_t(\theta) \) and its first derivative in (3.4) can be written as a pure moving average in terms of the innovation process:

\[
e_t(\theta) = g_t \epsilon_t + \sum_{k=1}^{t-1} \psi_{t0k}(\theta, \theta^0) g_{t-k} \epsilon_{t-k},
\]

\[\frac{\partial e_t(\theta)}{\partial \theta_i} = \sum_{k=1}^{t-1} \psi_{tik}(\theta, \theta^0) g_{t-k} \epsilon_{t-k}.\] (3.7)

for \( i = 1, ..., m \), where the coefficients \( \psi_{t0k}(\theta, \theta^0) \) and \( \psi_{tik}(\theta, \theta^0) \) are obtained from the autoregressive and moving average coefficients by the following relations:

\[
\psi_{t0k}(\theta, \theta^0) = \psi_{tk}(\theta^0) - \sum_{u=1}^{k} \pi_{tu}(\theta) \psi_{t-u,k-u},
\]

\[
\psi_{tik}(\theta, \theta^0) = - \sum_{u=1}^{k} \frac{\partial \pi_{tu}(\theta)}{\partial \theta_i} \psi_{t-u,k-u}.
\] (3.8)

Similarly, for the second and third derivatives, for \( i, j, l = 1, ..., m \), we have

\[
\psi_{tijk}(\theta, \theta^0) = - \sum_{u=1}^{k} \frac{\partial^2 \pi_{tu}(\theta)}{\partial \theta_i \partial \theta_j} \psi_{t-u,k-u},
\]

\[
\psi_{tijlk}(\theta, \theta^0) = - \sum_{u=1}^{k} \frac{\partial^3 \pi_{tu}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l} \psi_{t-u,k-u}.
\]

We denote \( \psi_{t0k} = \psi_{t0k}(\theta^0, \theta^0) \), \( \psi_{tik} = \psi_{tik}(\theta^0, \theta^0) \), \( \psi_{tijk} = \psi_{tijk}(\theta^0, \theta^0) \) and \( \psi_{tijlk} = \psi_{tijlk}(\theta^0, \theta^0) \). Note that \( \psi_{t0k} = 0 \), for \( k \geq 1 \), and 1, for \( k = 0 \).

### 3.3 An illustration: tdVAR(1)

Let us consider the tdVAR(1) model defined by \( x_t = A_t(\theta) x_{t-1} + e_t(\theta) \). It is a special case of the model defined in (3.2) with \( p = 1 \) and \( q = 0 \). Now, in this case, the coefficients of the pure moving average representation are given by \( \psi_{tk}(\theta) = \)
\[ \prod_{l=0}^{k-1} A_{l-1}(\theta), \text{ for } k = 1, 2, \ldots, t - 1, \text{ where a product for } l = 0 \text{ to } -1 \text{ is set to } I_r. \] The coefficients of the pure autoregressive form are \( \pi_{1i}(\theta) = A_i(\theta) \), and \( \pi_{tk}(\theta) = 0 \), for \( k = 2, 3, \ldots, t - 1 \), so, for \( i = 1, \ldots, m \), their first derivatives are given by \( \partial \pi_{1i}(\theta)/\partial \theta_i = \partial A_i(\theta)/\partial \theta_i \). Then

\[
\psi_{tik}(\theta, \theta^0) = \frac{\partial \pi_{1i}(\theta)}{\partial \theta_i} \prod_{l=1}^{k-1} A_{l-1}(\theta^0), \quad \frac{\partial e_t(\theta)}{\partial \theta_i} = -\frac{\partial A_i(\theta)}{\partial \theta_i} x_{t-1}, \quad (3.9)
\]

### 3.4 QMLE of tdVARMA(\( p, q \)) models: asymptotic results

Let \( \{x_t : t = 1, 2, \ldots, n\} \) be a partial realization of length \( n \) of the process \( \{x_t : t \in \mathbb{N}\} \) defined in (3.1). In the present section, we shall apply the general results of Section 2.2 to the tdVARMA(\( p, q \)) setting after formulating the minimal requirements for satisfying assumptions \( H_{1.1} - H_{1.4} \) (resp., \( H_{1.1} - H_{1.6} \)). The notations \( Q_n(\theta), \alpha_t(\theta) \) as well as the QMLE solution remain of course the same here.

Theorem 2 below establishes the strong consistency of the QMLE and further the asymptotic normality of this estimator. For convenience, we suppose that the parameters in \( A_{ti}(\theta) \) for \( i = 1, \ldots, p \), in \( B_{tj}(\theta) \) for \( j = 1, \ldots, q \), and in \( g_t(\theta) \) are functionally independent. Without loss of generality we suppose that the vector \( \theta \) is composed of three sub-vectors \( A, B \) and \( g \), more concretely \( \theta = (A^T, B^T, g^T)^T \), \( A \) being the sub-vector of the parameters included in \( A_{ti}(\theta) \) for \( i = 1, \ldots, p \), with dimension \( s_1 \), \( B \) the sub-vector of the parameters included in \( B_{tj}(\theta) \) for \( j = 1, \ldots, q \), with dimension \( s_2 \) and \( g \) the sub-vector of the parameters included in \( g_t(\theta) \) with dimension \( m - s_1 - s_2 \).

We now introduce a set of assumptions that will allow us, as mentioned above, to use results from Section 2.2. Using the Frobenius norm \( \|A\|_F = \sqrt{\text{tr}(A^T A)} \) for a \( m \times n \) matrix \( A \), we assume for all \( t \in \mathbb{N} \):

**H\(_{2.1}\)** : The matrices \( A_{ti}(\theta), B_{tj}(\theta) \) and \( g_t(\theta) \) are three times continuously differentiable with respect to \( \theta \), in an open set \( \Theta \) which contains the true value \( \theta^0 \) of \( \theta \);

**H\(_{2.2}\)** : There exist positive constants \( N_1, N_2, N_3, N_4, N_5 \), a polynomial \( P(\nu) \) such that \( P(1) = 1 \), and \( 0 < \Phi < 1 \) such that, for \( \nu = 1, \ldots, t - 1 \),

\[
\sum_{k=\nu}^{t-1} \|\psi_{tik}\|_F^2 < N_1 P(\nu) \Phi^{\nu-1}, \quad \sum_{k=\nu}^{t-1} \|\psi_{ljk}\|_F^2 < N_2 P(\nu) \Phi^{\nu-1},
\]

\[
\sum_{k=\nu}^{t-1} \|\psi_{ljk}\|_F^2 < N_3 P(\nu) \Phi^{\nu-1}, \quad \sum_{k=\nu}^{t-1} \|\psi_{ljk}\|_F^4 < N_4 P(\nu) \Phi^{\nu-1},
\]
\[
\sum_{k=1}^{t-1} \| \psi_{ijk} \|^2_F < N_5, \quad i, j, l = 1, \ldots, m;
\]

**H\textsubscript{2.3}**: There exist positive constants \( K_1, K_2, K_3, K_4, K_5 \) such that

\[
\left\| \frac{\partial \Sigma_t}{\partial \theta_i} \right\|_{F, \theta=\theta^0}^2 \leq K_1, \quad \left\| \frac{\partial^2 \Sigma_t}{\partial \theta_j \partial \theta_j} \right\|_{F, \theta=\theta^0}^2 \leq K_2, \quad \left\| \frac{\partial^3 \Sigma_t^{-1}}{\partial \theta_j \partial \theta_j \partial \theta_i} \right\|_{F, \theta=\theta^0}^2 \leq K_3,
\]

\[
\left\| \frac{\partial \Sigma_t^{-1}}{\partial \theta_i} \right\|_{F, \theta=\theta^0}^2 \leq K_4, \quad \left\| \frac{\partial^2 \Sigma_t^{-1}}{\partial \theta_j \partial \theta_j} \right\|_{F, \theta=\theta^0}^2 \leq K_5, \quad i, j, l = 1, \ldots, m;
\]

**H\textsubscript{2.4}**: There exist positive constants \( M_1, M_2, \) and \( M_3 \) such that

\[
(a) \quad E \left[ (\epsilon_t^T \epsilon_t)^4 \right] \leq M_1, \quad (b) \quad \| E (\epsilon_t \epsilon_t^T \otimes \epsilon_t^T) \|_F \leq M_2,
\]

\[
(c) \quad \| \kappa_t \|_F + \| \text{vec}(\Sigma). \text{vec}(\Sigma)^T \|_F + \| \Sigma \otimes \Sigma \|_F + \| K_{r,r}(\Sigma \otimes \Sigma) \|_F \leq M_3;
\]

**H\textsubscript{2.5}**: There exist positive constants \( m_1 \) and \( m_2 \) such that \( \| g_t \|^2_F \leq m_1, \quad \| \Sigma_t^{-1} \|^2_F \leq m_2. \)

Furthermore, we suppose that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^n \left\{ E_{\theta^0} \left( \frac{\partial e_t^T}{\partial \theta_i} \Sigma_t^{-1} \frac{\partial e_t}{\partial \theta_j} \right) + \text{tr} \frac{1}{2} \left[ \Sigma_t^{-1} \frac{\partial \Sigma_t}{\partial \theta_i} \Sigma_t^{-1} \frac{\partial \Sigma_t}{\partial \theta_j} \right] \right\} = V_{ij},
\]

for \( i, j = 1, \ldots, m, \) where the matrix \( V = (V_{ij})_{1 \leq i,j \leq m} \) is strictly positive definite;

**H\textsubscript{2.7}**: 

(A) \quad \frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k=1}^{t-1} \| g_{t-k} \|^2_F \| \psi_{t+k} \|_F \| \psi_{t+d,i,k+d} \|_F = O \left( \frac{1}{n} \right),

(B) \quad \frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k=1}^{t-1} M_{t0kk}^{jiT} \Xi_{t-k} M_{tikk}^{ij} + \sum_{k_1=1}^{t-1} \sum_{k_2=1}^{t-1} M_{t0k2k1}^{jiT} K_{r,r}(\Sigma \otimes \Sigma) M_{t2k1k2}^{ij} \nonumber

+ \sum_{k_1=1}^{t-1} \sum_{k_2=1}^{t-1} M_{t0k2k1}^{jiT} (\Sigma \otimes \Sigma) M_{t2k1k2}^{ij} = O \left( \frac{1}{n} \right),

with \( \Xi_t(\Sigma) = \kappa_t(\Sigma) - \text{vec}(\Sigma). \text{vec}(\Sigma)^T - (\Sigma \otimes \Sigma) - K_{r,r}(\Sigma \otimes \Sigma), \) and, for \( k, k' = k, k_1, k_2, \) \( M_{t, k,k'}^{ij} = \text{vec}(g_{t-k}^T \psi_{t+k}^T \psi_{t+f,j,k'} + \Sigma_t^{-1} \psi_{t+f,j,k'} + f g_{t-k}^T), \) \( f = 0, d; \)

**H\textsubscript{2.8}**: There exists a positive definite matrix of constants \( W \) defined by (22).
See Remark A.1 for comments on the assumptions and their use. With these assumptions in hand, we are able to show that the conditions for Theorem 1 hold (see Appendix A.2 for a sketch of the proof) and hence obtain the following result.

**Theorem 2.** Suppose that assumptions $H_{2.1}-H_{2.8}$ hold. Then there exists a sequence of estimators $\hat{\theta}_n = (\hat{\theta}_1, ..., \hat{\theta}_m)^T$ such that, as $n \to \infty$, we have

- $\hat{\theta}_n \to \theta^0$ a.s.;
- $n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{L} \mathcal{N}(0, V^{-1}WV^{-1})$.

**Remark 1.** Note that for a Gaussian process, we have $W = V$. For the sake of simplicity in the proof of Theorem 2 (see Appendix A.2), we have made assumptions on $\Sigma_t$ in addition to those on $g_t$. The proof is somewhat similar to that of Azrak & Mélard (2006) but is extended to multivariate processes. Note however several corrections with respect to that paper (e.g. the treatment of the third term of (A.3); $\theta^0$ and $\theta$ were sometimes not distinguished where they should, especially in Section 3.2) and improvements (the treatment of the last three terms of (A.3) is more detailed; also the role of the assumptions is better enlightened).

### 4 Some examples

The two examples will show that the theory can be applied and that the assumptions can be verified.

**4.1 Example 1: tdVAR(1) a generalization of Kwoun & Yajima (1986)**

In the following example we discuss a generalization of Kwoun & Yajima (1986). To achieve this we consider the bivariate tdVAR(1) model

$$
\begin{pmatrix}
    x_{t1} \\
    x_{t2}
\end{pmatrix}
= 
\begin{pmatrix}
    A_{11}^{11} & A_{12}^{12} \\
    A_{21}^{21} & A_{22}^{22}
\end{pmatrix}
\begin{pmatrix}
    x_{t-1,1} \\
    x_{t-1,2}
\end{pmatrix}
+ 
\begin{pmatrix}
    \epsilon_{t1} \\
    \epsilon_{t2}
\end{pmatrix},
$$

where the coefficients $A_{ij}^{ab}(\theta)$ for $i, j = 1, 2$ are defined as

$$A_{ij}^{ab}(\theta) = A_{ij}^a \sin(\alpha_{ij} t + A_{ij}^b).$$
The unknown parameters $A'_{ij}$ and $A''_{ij}$ are such that $A'_{ij} \in [\delta, 1-\delta]$ and $A''_{ij} \in [0, 2\pi-\delta]$ for some fixed $1/2 > \delta > 0$ and $\alpha_{ij}$ are known constants. Then

$$\theta = (A'_{11}, A'_{21}, A'_{12}, A'_{22}, A''_{11}, A''_{21}, A''_{12}, A''_{22})^T.$$ 

The numerical example proposed by Kwoun & Yajima (1986) for $r = 1$ contains a process with periodic coefficients of period 4 because $\alpha_{ij} = \pi/2$. However, it is well known, see e.g. Tiao & Grupe (1980) and Azrak & Méard (2006), that an $r$-dimensional autoregressive process with periodic coefficients of period $s \in \mathbb{N}$ can be embedded into an $s$-dimensional stationary autoregressive process. To avoid this simplification we consider coefficients $A'_{ij}(\theta)$ either with distinct irrational periods or at least with large relatively prime periods. We check the assumptions of Theorem 2 in Appendix S1.1 in the simplified case where we have

$$A_t(\theta) = \begin{pmatrix} A'_{11} \sin(at) & \frac{1}{2} \\ 0 & A''_{22} \sin(bt) \end{pmatrix}, \quad (4.3)$$

with $\theta = (A'_{11}, A''_{22})^T$. For simplicity, we take here $\Sigma = I_2$. Some parts of the verification assume that the two sines have an integer period. For the simulation study, we shall rather have recourse to the model

$$A_t(\theta) = \begin{pmatrix} A'_{11} \sin(at) & A'_{12} \\ 0 & A''_{22} \sin(bt) \end{pmatrix}, \quad (4.4)$$

with

$$\theta = (A'_{11}, A'_{12}, A''_{22})^T, \quad a = \frac{2\pi}{\sqrt{2499}} \quad \text{and} \quad b = \frac{2\pi}{\sqrt{2399}}. \quad (4.5)$$

We impose that $A'_{11}, A''_{22} \in [0, 1[$, partly similarly to Kwoun & Yajima (1986), and write $\theta^0 = (A'_{11}^0, 0.5, A''_{22}^0)^T$ for the true value of $\theta$.

### 4.2 Example 2: tdVAR(1) with heteroscedasticity

Let us re-consider the model defined in (4.1)-(4.3), except that $A'_{12}$ is no longer a parameter, with the added difficulty that the innovations are now $g_t\epsilon_t$ instead of $\epsilon_t$. Therefore we introduce a matrix $g_t(\theta)$ and we have a bounded time-dependent covariance matrix $\Sigma_t(\theta) = g_t(\theta)\Sigma g_t^T(\theta)$. We have extended the Kwoun & Yajima (1986)
parametrization by taking
\[ g_t(\theta) := \begin{pmatrix} \exp(-\eta_{11}\sin(ct)) & 1 \\ -1 & \exp(-\eta_{22}\sin(ct)) \end{pmatrix} \] (4.6)
with \( c \in \mathbb{R} \). Also we use a matrix \( \Sigma \) which is no longer the identity matrix:
\[ \Sigma := \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}. \]
A close examination of \( \Sigma_t(\theta) \) shows that if \( g_t(\theta) \) were diagonal, then the correlation between the residuals would be constant, which is not very realistic for a time-dependent process. This is why we have put off-diagonal elements different from 0 in (4.6). Here, the vector \( \theta \) reduces to \( \theta = (A_1', A_2', \eta_{11}, \eta_{22})^T \). The assumptions of Theorem 2 are checked in Appendix S1.2, and simulation results are shown in Section 5.2.

**Remark 2.** Other functions of time can be considered. To simplify, we illustrate univariate processes. Let us first consider a tdAR(1) process like in Example 1 of Section 4.1. Let \( A_t(\theta) = \tilde{A} \) for \( t \leq t_0 \), and \( A'sin(at) \) for \( t > t_0 \), where \( \tilde{A} \), \( a \) and \( t_0 > 1 \) are given constants. Let \( g_t = 1 \) and \( \theta = (A') \). Even if \( |\tilde{A}| > 1 \), all the assumptions are fulfilled. Any bounded function can be used for \( g_t(\theta) \) provided it does not involve parameters. Other functions can be considered for \( A_t(\theta) \) as well as for \( g_t(\theta) \). In particular, let us take \( A_t(\theta) = \phi \exp(-b/t) \) for \( \theta = (\phi) \), \( 0 < \phi < 1 \), \( b \) is a given constant and \( g_t = \Sigma = 1 \), to simplify. Then the variance of \( x_t \) given by \( \sum_{k=0}^{t-1} \phi^{2k} \exp(-2b \sum_{i=0}^{k-1} (t - i)^{-1} \) does exist for all \( t \). With the notations in Appendix S1, \( V_{11}(t) = \exp(-2b/t) \to 1 \) when \( t \to \infty \), so that the limit \( V > 0 \) does exist. Note that treating \( b \) as a parameter \( \theta_2 \) will not work since then \( V_{22}(t) = (\phi^2/t^2) \exp(-2b/t) \) behaves like \( \phi^2/t^2 \) when \( t \to \infty \), but \( \sum_{i=1}^{\infty} t^{-2} = \pi^2/6 \) hence \( V_{22} \to 0 \) when \( n \to \infty \). tdMA(1) processes can also be considered although the details are more complex. Finally, a tdAR(p) process defined by \( x_t = \sum_{i=1}^{p} A_i' \sin(a_i t) x_{t-i} + g_t \epsilon_t \) can be written as a p-variate tdVAR(1) process by using \( (x_t, x_{t-1}, \ldots, x_{t-p+1})^T \) as the variable and a companion matrix for the coefficient.
5 Simulation results

5.1 Example 1: tdVAR(1) a generalization of Kwoun & Ya-jima (1986)

The simulation experiment is performed in Matlab by using the program, which we call AJM, described in Alj et al. (2016) and based on a special case of tdVAR(1) process defined in (4.4)-(4.5), with \( A'_{11} = 0.8 \) and \( A'_{22} = -0.9 \) and \((\epsilon_t, \epsilon_t)^T\) has a bivariate normal distribution with covariance \( \Sigma = I_2 \). A simulated series using these specifications is shown in Fig. 1. The true value of \( \theta \) is \( \theta^0 = (A'^0_{11}, A'^0_{12}, A'^0_{22})^T = (0.8, 0.5, -0.9)^T \). We take \( \theta^i = (0.1, 0.1, 0.1)^T \) as initial value of \( \theta \).

The experiment was replicated 1000 times. The results are summarized in Table 1. As the sample size becomes larger, we can see that

- the averages of the estimates become closer to their true value in accordance with the theory,
- the sample standard deviation on line (b) becomes also closer to the averages across simulations of estimated standard errors in line (c) showing that the standard errors are well estimated, and
- the percentage of rejecting the hypothesis \( H_0 \) is close to 5%.

We compare, for the sample size \( n = 400 \), a histogram of the 1000 replications of \( \hat{\theta}_i \) to the normal probability curve with mean equal to \( \theta^0 \) and standard deviation given in line (b) of Table 1. As we can see from the corresponding Figure 2, this histogram shows empirically consistency and normality of the estimates.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1}
\caption{Figure 1 should be here}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2}
\caption{Figure 2 should be here}
\end{figure}

5.2 Example 2: tdVAR(1) with heteroscedasticity

We keep the bivariate model defined by (4.1)-(4.3), with the same numerical values for \( a \) and \( b \) but without \( A'_{12}(\theta) \), with \( g_t \epsilon_t \) instead of \( \epsilon_t \) and a covariance matrix \( \Sigma_t(\theta) = g_t(\theta) \Sigma g_t^T(\theta) \) bounded but time-dependent, where \( g_t(\theta) \) is defined by (4.6), with \( \Sigma_{11} = \Sigma_{22} = 1, \Sigma_{12} = 0.5, c = 2\pi/25 \) and \( \eta_{11} = 1, \eta_{22} = -1 \) so that the innovation correlation coefficient varies between \(-0.8\) and \(0.8\).
Figure 1: Time plots of the coefficients and the simulated tdVAR(1) generated by the process defined in Section 5.1 of length $n = 400$.

Figure 2: Histograms of parameter estimates with the normal density function of 1000 replications of the process of Section 5.1 with $n = 400$. 
Table 1: Estimation results for the model \( (4.1) \) under \( (4.4)-(4.5) \) via the program AJM, where lines (a) give the averages of the parameter estimates, lines (b) give the averages across simulations of estimated standard errors of the corresponding estimates for the 1000 replications, lines (c) the sample standard deviations of the corresponding estimates for the 1000 replications and lines (d) give percentages of simulations where we reject the hypothesis \( H_0(\theta_i = \theta^0_i) \) at significance level 5%.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>( \hat{A}_{11} )</th>
<th>( \hat{A}_{12} )</th>
<th>( \hat{A}_{22} )</th>
<th>( \hat{\Sigma}_{11} )</th>
<th>( \hat{\Sigma}_{12} )</th>
<th>( \hat{\Sigma}_{22} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>0.7441</td>
<td>0.5017</td>
<td>-0.8011</td>
<td>1.0038</td>
<td>0.0005</td>
<td>0.9962</td>
</tr>
<tr>
<td>(a)</td>
<td>0.1997</td>
<td>0.1555</td>
<td>0.1991</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(b)</td>
<td>0.2035</td>
<td>0.1663</td>
<td>0.2193</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(c)</td>
<td>5.1</td>
<td>7.6</td>
<td>5.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(d)</td>
<td>0.7734</td>
<td>0.5031</td>
<td>-0.8506</td>
<td>1.0018</td>
<td>0.0003</td>
<td>0.9982</td>
</tr>
<tr>
<td>50</td>
<td>0.1378</td>
<td>0.1053</td>
<td>0.1319</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(a)</td>
<td>0.1352</td>
<td>0.1149</td>
<td>0.1423</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(b)</td>
<td>5.2</td>
<td>6.3</td>
<td>5.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>100</td>
<td>0.7834</td>
<td>0.5008</td>
<td>-0.8722</td>
<td>1.0008</td>
<td>0.0002</td>
<td>0.9992</td>
</tr>
<tr>
<td>(a)</td>
<td>0.0966</td>
<td>0.0735</td>
<td>0.0905</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(b)</td>
<td>0.1002</td>
<td>0.0777</td>
<td>0.0974</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(c)</td>
<td>5.5</td>
<td>6.7</td>
<td>5.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>200</td>
<td>0.7916</td>
<td>0.4986</td>
<td>-0.8861</td>
<td>1.0004</td>
<td>0.0001</td>
<td>0.9996</td>
</tr>
<tr>
<td>(a)</td>
<td>0.0677</td>
<td>0.0512</td>
<td>0.0625</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(b)</td>
<td>0.0677</td>
<td>0.0531</td>
<td>0.0665</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(c)</td>
<td>4.7</td>
<td>5.8</td>
<td>6.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>400</td>
<td>0.7964</td>
<td>0.4997</td>
<td>-0.8925</td>
<td>1.0002</td>
<td>0.0000</td>
<td>0.9998</td>
</tr>
<tr>
<td>(a)</td>
<td>0.0474</td>
<td>0.0359</td>
<td>0.0438</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(b)</td>
<td>0.0490</td>
<td>0.0367</td>
<td>0.0455</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>(c)</td>
<td>5.6</td>
<td>5.9</td>
<td>5.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Again, the number of replications is 1000. The initial values were \( \theta^i = (0.1, 0.1, 0.01, 0.01)^T \). The results are presented in Table 2. Moreover a program for computing the asymptotic information matrix on the basis of the formulae in Appendix S1 gives, for \( n = 50 \) for example, the standard errors 0.0905, 0.0908, 0.1995, 0.1995, whereas the averages of the standard errors estimated by the QMLE program were 0.0906, 0.1154, 0.1936, 0.1462, respectively, and the standard deviations of the 1000 estimates were 0.0976, 0.1168, 0.2327, 0.1779.
Table 2: Estimation results for the model (4.1) under (4.3), (4.5) and (4.6), via the program AJM, where columns (a) give the averages of the parameter estimates and columns (d) give percentages of 1000 simulations where we reject the hypothesis $H_0(\theta_i = \theta_{i0})$ at significance level 5%.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\hat{A}<em>{11}' (A</em>{11}^0 = 0.8)$</th>
<th>$\hat{A}<em>{22}' (A</em>{22}^0 = -0.9)$</th>
<th>$\hat{\eta}<em>{11} (\eta</em>{11}^0 = 1.0)$</th>
<th>$\hat{\eta}<em>{22} (\eta</em>{22}^0 = -1.0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a) 0.7796 (8.1)</td>
<td>(d) -0.8511 (7.3)</td>
<td>(a) 0.9972 (14.9)</td>
<td>(d) -1.0299 (13.8)</td>
</tr>
<tr>
<td></td>
<td>(a) 0.7818 (6.9)</td>
<td>(d) -0.8780 (5.7)</td>
<td>(a) 0.9875 (10.1)</td>
<td>(d) -1.0131 (11.0)</td>
</tr>
<tr>
<td></td>
<td>(a) 0.7928 (5.6)</td>
<td>(d) -0.8835 (5.6)</td>
<td>(a) 0.9965 (10.9)</td>
<td>(d) -1.0072 (7.5)</td>
</tr>
<tr>
<td></td>
<td>(a) 0.7959 (6.3)</td>
<td>(d) -0.8910 (4.9)</td>
<td>(a) 1.0009 (8.6)</td>
<td>(d) -1.0033 (8.3)</td>
</tr>
<tr>
<td></td>
<td>(a) 0.7973 (4.7)</td>
<td>(d) -0.8952 (5.6)</td>
<td>(a) 1.0032 (7.6)</td>
<td>(d) -1.0017 (7.2)</td>
</tr>
</tbody>
</table>

**Acknowledgments**

This work was started while Abdelkamel Alj and Christophe Ley were both at Université libre de Bruxelles, Faculty of Sciences, Dept. of Mathematics. Christophe Ley thanks the Fonds National de la Recherche Scientifique, Communauté Française de Belgique, for financial support via a Mandat de Chargé de Recherche FNRS. Guy Mélard has benefited from a Belgian research grant F.R.S.-FNRS 1.5.261.09. He thanks support from Faculty of Sciences, Dept. of Mathematics. We are grateful to two anonymous referees for their suggestions and encouragement, and to the editor and the associate editor in charge of the paper. They have all contributed to a significant improvement of the paper.

**Supporting Information**

Additional information for this article is available online.

Appendix S1. Checking the assumptions for Example 1 of Section 4.1 and for Example 2 of Section 4.2

Appendix S2. Technical Appendix to “Asymptotic properties of QML estimators for VARMA models with time-dependent coefficients”
References


Corresponding author: Christophe Ley, Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Krijgslaan 281, S9, Campus Sterre, B-9000 Gent, Belgium (e-mail: christophe.ley@ugent.be).

# Appendices

## A Proof of theorems

### A.1 Further preliminaries

In order to prove Theorem 2, we need the following Lemma, due to Hamdoune (1995) (see also Azrak & Mélard 2006), and a strong law of large numbers for mixingale sequences.

**Lemma A.1.** Let \( \{w_t, t = 1, \ldots, n\} \) be, for each \( n \in \mathbb{N} \), a scalar process with finite second-order moments, i.e.

**Lemma A.1-i:** \( E(w_t^2) < \infty \), for all \( t = 1, \ldots, n \);

**Lemma A.1-ii:** \( E(n^{-1} \sum_{t=1}^{n} w_t^2) = O(n^{-\delta}) \) with \( \delta > 0 \).

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

We also need a strong law of large numbers for mixingale sequences, e.g. Hall & Heyde (1980, Theorem 2.21) in the special case where their sequence \( b_n = n \). Let us recall the definition from Hall & Heyde (1980, Section 2.3).

**Definition A.1.** Let \( \{w_t, t \geq 1\} \) be square-integrable random variables on a probability space \((\Omega, F, P)\) and \( \{F_t, -\infty < t < \infty\} \) be an increasing sequence of \( \sigma \)-fields of \( F \). Then \( \{w_t, F_t\} \) is a \( L_2 \)-mixingale sequence if for sequences of non-negative constants \( \psi_n \) and \( c_t \) where \( \psi_n \to 0 \) as \( n \to \infty \), we have

(a) \( E \{E(w_t | F_{t-\nu})^2\}^{1/2} \leq \psi_{n-1} c_t \), and (b) \( E \{(w_t - E(w_t | F_{t+\nu}))^2\}^{1/2} \leq \psi_{n+1} c_t \).

**Lemma A.2.** If \( \{w_t, F_t\} \) is a \( L_2 \)-mixingale sequence, and if \( n^{-2} \sum_{t=1}^{n} c_t^2 < \infty \), as \( n \to \infty \), and \( \psi_n = O(\nu^{-1/2}(\log \nu)^{-2}) \) as \( \nu \to \infty \), then \( n^{-1} \sum_{t=1}^{n} w_t \overset{a.s.}{\to} 0 \) as \( n \to \infty \).
A.2 Proof of Theorem 2

Since we proceed as in Azrak & Mélard (2006), we just sketch the proof. Lemmas proceeded by TA refer to the Technical Appendix in Appendix S2.

First of all, as shown in Section 3.2, assumption $H^2_{2.1}$ is used to define the $\psi$’s used in $H^2_{2.2}$ and in the derivatives in the other assumptions. The idea is to check the first five assumptions of Theorem 1, since $H^2_{2.8}$ coincides with $H^1_{1.6}$. Except $H^1_{1.3}$ and $H^1_{1.6}$, all the other conditions involve bounds, even in order to check the conditions of Lemma A.2. These bounds are much more complex to obtain than in Azrak & Mélard (2006) because the matrix product is non commutative, and also because more terms are involved in the expressions. For that reason, they will be given in TA Lemma’s, see however Remark A.2. First $\partial\alpha_t(\theta)/\partial \theta_i$ is given by

$$\text{tr} \left( \Sigma^{-1}_t(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_i} \right) + e^T_t(\theta) \{ \Sigma^{-1}_t(\theta) \} \frac{\partial e_t(\theta)}{\partial \theta_i} + e^T_t(\theta) \frac{\partial \Sigma^{-1}_t(\theta)}{\partial \theta_i} e_t(\theta),$$

with a more complex expression for the second derivatives. Then $H^1_{1.1}$ and $H^1_{1.2}$ are direct consequences of TA Lemma 2.1 (using TA Lemma 4.11) and TA Lemma 2.2 (using TA Lemma 4.12), respectively. This makes use of assumptions $H^2_{2.2}$, $H^2_{2.3}$, $H^2_{2.4}(a)$ and $H^2_{2.5}$. Proving the other assumptions $H^1_{1.3}$, $H^1_{1.4}$, and $H^1_{1.5}$ requires Section A.1 as will be seen.

A.2.1 Proof of $H^1_{1.3}$

Let us consider the process $Z_{tij}$ defined by

$$Z_{tij} = \left\{ \frac{\partial e_t^T(\theta)}{\partial \theta_i} \Sigma^{-1}_t(\theta) \frac{\partial e_t(\theta)}{\partial \theta_j} \right\}_{\theta = \theta_0} - E_{\theta_0} \left( \frac{\partial e_t^T(\theta)}{\partial \theta_i} \Sigma^{-1}_t(\theta) \frac{\partial e_t(\theta)}{\partial \theta_j} \right).$$

By using TA Lemma 4.13 and assumptions $H^2_{2.2}$, $H^2_{2.3}$, $H^2_{2.4}(c)$, $H^2_{2.5}$ and $H^2_{2.7}(B)$, the two assumptions A.1-i and A.1-ii of Lemma A.1 are fulfilled for $Z_{tij}$, hence

$$\frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\partial e_t^T(\theta)}{\partial \theta_i} \Sigma^{-1}_t(\theta) \frac{\partial e_t(\theta)}{\partial \theta_j} \right\}_{\theta = \theta_0} - E_{\theta_0} \left( \frac{\partial e_t^T(\theta)}{\partial \theta_i} \Sigma^{-1}_t(\theta) \frac{\partial e_t(\theta)}{\partial \theta_j} \right) \overset{a.s.}{\rightarrow} 0 \quad (A.1)$$

when $n \to \infty$. However, from TA Lemma 4.5, we have for $i, j = 1, ..., m$:

$$\frac{1}{2n} \sum_{t=1}^{n} E_{\theta_0} \left( \frac{\partial^2 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j} / F_{t-1} \right) = \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{\partial e_t^T(\theta)}{\partial \theta_i} \Sigma^{-1}_t(\theta) \frac{\partial e_t(\theta)}{\partial \theta_j} \right\}_{\theta = \theta_0}.$$
\[ + \frac{1}{2n} \sum_{t=1}^{n} \text{tr} \left\{ \Sigma_t^{-1}(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_i} \Sigma_t^{-1}(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_j} \right\} \theta = \theta^0. \]  

(A.2)

Then (A.1) implies that the a.s. limit of (A.2) for \( n \to \infty \) will be equal to

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \text{tr} \left\{ \Sigma_t^{-1}(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_i} \Sigma_t^{-1}(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_j} \right\} \theta = \theta^0, \]

and this is \( V_{ij} \) as defined in \( H_{2.6} \).

**A.2.2 Proof of \( H_{1.4} \)**

By using the mean value theorem on TA Lemma 2.3, it suffices to show that

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\partial^3 \alpha_t(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right) \theta = \theta^0 < \infty \quad \text{a.s. for} \quad i, j, l = 1, \ldots, m. \]

As a consequence of TA Lemma 2.4, the expression under the limit is bounded by

\[ \tilde{\Phi}_1 + \frac{1}{n} \sum_{t=1}^{n} \tilde{\Phi}_2t + \frac{1}{n} \sum_{t=1}^{n} \tilde{\Psi}_4t + \frac{1}{n} \sum_{t=1}^{n} \tilde{\Psi}_2t + \frac{1}{n} \sum_{t=1}^{n} \tilde{\Psi}_3t, \]  

(A.3)

where \( \tilde{\Phi}_1 \) is shown in TA Lemma 4.14 (using assumptions \( H_{2.3} \) and \( H_{2.5} \)) to be bounded, and the last four terms \( \tilde{\Phi}_2t, \tilde{\Psi}_4t, \tilde{\Psi}_2t, \) and \( \tilde{\Psi}_3t \) are briefly described now:

- \( \tilde{\Phi}_2t \) contains terms like \( e_t^T(\theta) \left( \frac{\partial^2 \Sigma_t^{-1}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right) e_t(\theta), i, j, l = 1, \ldots, m; \)

- \( \tilde{\Psi}_4t \) contains terms like \( e_t^T(\theta) \left( \frac{\partial^2 \Sigma_t^{-1}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right) \frac{\partial e_t(\theta)}{\partial \theta_i} \), \( e_t^T(\theta) \left( \frac{\partial \Sigma_t^{-1}(\theta)}{\partial \theta_i} \right) \frac{\partial e_t(\theta)}{\partial \theta_j} \), and \( e_t^T(\theta) \left( \frac{\partial^2 \Sigma_t^{-1}(\theta)}{\partial \theta_i \partial \theta_j \partial \theta_l} \right) \), where \( i, j, l = 1, \ldots, m; \)

- the last two terms \( \tilde{\Psi}_2t \) and \( \tilde{\Psi}_3t \) contain respectively terms like \( \left( \frac{\partial e_t^T(\theta)}{\partial \theta_i} \right) \left( \frac{\partial^2 \Sigma_t^{-1}(\theta)}{\partial \theta_i \partial \theta_j} \right) \left( \frac{\partial e_t(\theta)}{\partial \theta_j} \right) \) and \( \left( \frac{\partial^2 e_t^T(\theta)}{\partial \theta_i \partial \theta_j} \right) \Sigma_t^{-1}(\theta) \left( \frac{\partial e_t(\theta)}{\partial \theta_l} \right), \) where \( i, j, l = 1, \ldots, m. \)
For $i,j,l = 1,\ldots, m$, let us define the following six sequences of random variables

\[ X_{t}^{ij} = \left( \frac{\partial e^T_t(\theta)}{\partial \theta^i} \frac{\partial \Sigma^{-1}_t(\theta)}{\partial \theta^j} \right) - E_{\theta^0} \left[ \frac{\partial e^T_t(\theta)}{\partial \theta^i} \frac{\partial \Sigma^{-1}_t(\theta)}{\partial \theta^j} \right], \quad (A.4) \]

\[ Y_{t}^{ijl} = \left( \frac{\partial^2 e^T_t(\theta)}{\partial \theta^j \partial \theta^l} \Sigma^{-1}_t(\theta) \frac{\partial e_t(\theta)}{\partial \theta^i} \right) - E_{\theta^0} \left[ \frac{\partial^2 e^T_t(\theta)}{\partial \theta^j \partial \theta^l} \Sigma^{-1}_t(\theta) \frac{\partial e_t(\theta)}{\partial \theta^i} \right], \quad (A.5) \]

\[ Z_{t} = \left( e^T_t(\theta) \frac{\partial^3 \Sigma^{-1}_t(\theta)}{\partial \theta^i \partial \theta^j \partial \theta^l} \right) - E_{\theta^0} \left[ e^T_t(\theta) \frac{\partial^3 \Sigma^{-1}_t(\theta)}{\partial \theta^i \partial \theta^j \partial \theta^l} \right], \quad (A.6) \]

\[ W_{t,1}^{ij} = \left( e^T_t(\theta) \frac{\partial^2 \Sigma^{-1}_t(\theta)}{\partial \theta^i \partial \theta^j} \right) - E_{\theta^0} \left[ e^T_t(\theta) \frac{\partial^2 \Sigma^{-1}_t(\theta)}{\partial \theta^i \partial \theta^j} \right], \quad (A.7) \]

and $W_{t,2}^{ij}$ and $W_{t,3}^{ij}$ defined similarly to (A.7) by replacing $(\partial^2 \Sigma^{-1}_t(\theta)/\partial \theta^i \partial \theta^j)(\partial e_t(\theta)/\partial \theta^l)$ with, respectively, $(\partial \Sigma^{-1}_t(\theta)/\partial \theta^i)(\partial^2 e_t(\theta)/\partial \theta^j \partial \theta^l)$ and $\Sigma^{-1}_t(\theta)(\partial^3 e_t(\theta)/\partial \theta^i \partial \theta^j \partial \theta^l)$.

First, we have that $E(Z_t|F_{t-1}) = 0$ and, according to TA Lemma 4.15 (using assumptions $H_{2.3}, H_{2.4}(c)$ and $H_{2.5}$), that $E(Z_t^2)$ is uniformly bounded by a constant. Then, the strong law of large numbers for martingale sequences (Stout, 1974, p. 154) implies that $n^{-1} \sum_{t=1}^{n} Z_t \overset{a.s.}{\to} 0$ when $n \to \infty$. The arguments for the other sequences are more involved. From TA Lemmas 4.16, 4.18 and 4.20 we have that $\{W_{t,q}^{ijl}, F_t\}$, $q = 1, 2, 3$, $\{X_t^{ij}, F_t\}$ and $\{Y_t^{ijl}, F_t\}$ are $L_2$-mixingale sequences. From TA Lemmas 4.17, 4.19 and 4.21 we have that these $L_2$-mixingale sequences $\{W_{t,q}^{ijl}, F_t\}$, $q = 1, 2, 3$, $\{X_t^{ij}, F_t\}$ and $\{Y_t^{ijl}, F_t\}$ fulfill the conditions in Lemma A.2, the strong law of large numbers for a mixingale sequence (Hall & Heyde, 1980, p. 41, Theorem 2.21). This makes use of assumptions $H_{2.2}$ (but not only for $\nu = 1$, like before), $H_{2.3}, H_{2.4}(c)$ and $H_{2.5}$. Hence

\[ \frac{1}{n} \sum_{t=1}^{n} W_{t,q}^{ijl} \overset{a.s.}{\to} 0, \quad q = 1, 2, 3, \quad \frac{1}{n} \sum_{t=1}^{n} X_t^{ij} \overset{a.s.}{\to} 0, \quad \text{and} \quad \frac{1}{n} \sum_{t=1}^{n} Y_t^{ijl} \overset{a.s.}{\to} 0, \]

and this for $i, j, l = 1, \ldots, m$.

The proof is completed so that $H_{1.4}$ is checked. Consequently, there exists an estimator $\hat{\theta}_n$ such that $\hat{\theta}_n \overset{a.s.}{\to} \theta^0$ as $n \to \infty$. 

25
A.2.3 Proof of $H_{1.5}$

From TA Lemma 4.10 we can determine the explicit form of the left-hand side of $H_{1.5}$ for all $1 \leq i, j \leq m$:

\[
\frac{4}{n} \sum_{t=1}^{n} \left[ \left\{ \frac{\partial e_{t}^{T}(\theta)}{\partial \theta_{j}} \Sigma_{t}^{-1}(\theta) \frac{\partial c_{t}(\theta)}{\partial \theta_{i}} \right\}_{\theta=\theta^{0}} - E_{\theta^{0}} \left( \frac{\partial e_{t}^{T}(\theta)}{\partial \theta_{j}} \Sigma_{t}^{-1}(\theta) \frac{\partial c_{t}(\theta)}{\partial \theta_{i}} \right) \right] \\
+ \frac{2}{n} \sum_{t=1}^{n} \left[ \left\{ \frac{\partial e_{t}^{T}(\theta)}{\partial \theta_{j}} \right\}_{\theta=\theta^{0}} - E_{\theta^{0}} \left( \frac{\partial e_{t}^{T}(\theta)}{\partial \theta_{j}} \right) \right] K_{tj} \\
+ \frac{2}{n} \sum_{t=1}^{n} \left[ \left\{ \frac{\partial e_{t}^{T}(\theta)}{\partial \theta_{j}} \right\}_{\theta=\theta^{0}} - E_{\theta^{0}} \left( \frac{\partial e_{t}^{T}(\theta)}{\partial \theta_{j}} \right) \right] K_{ti},
\]

where $K_{ti}$ is defined by

\[
K_{ti} = \Sigma_{t}^{-1} g_{t} E \left( \epsilon_{t}^{\otimes 3} \right) (g_{t}^{T} \otimes g_{t}^{T}) \operatorname{vec} \left( \frac{\partial \Sigma_{t}^{-1}(\theta)}{\partial \theta_{i}} \right)_{\theta=\theta^{0}},
\]

with $E \left( \epsilon_{t}^{\otimes 3} \right) = E \left( \epsilon_{t} e_{t}^{T} \otimes e_{t}^{T} \right)$. While checking $H_{1.3}$, we have shown that

\[
\frac{4}{n} \sum_{t=1}^{n} \left[ \left\{ \frac{\partial e_{t}^{T}(\theta)}{\partial \theta_{j}} \Sigma_{t}^{-1}(\theta) \frac{\partial c_{t}(\theta)}{\partial \theta_{i}} \right\}_{\theta=\theta^{0}} - E_{\theta^{0}} \left( \frac{\partial e_{t}^{T}(\theta)}{\partial \theta_{j}} \Sigma_{t}^{-1}(\theta) \frac{\partial c_{t}(\theta)}{\partial \theta_{i}} \right) \right] \overset{a.s.}{\longrightarrow} 0.
\]

There remains to prove that the second and third terms of (A.8) also tend a.s. to zero. To achieve that, let us consider

\[
\tilde{Z}_{t,ij}(\theta) = \left( \left\{ \frac{\partial e_{t}^{T}(\theta)}{\partial \theta_{j}} \right\}_{\theta=\theta^{0}} - E_{\theta} \left( \frac{\partial e_{t}^{T}(\theta)}{\partial \theta_{j}} \right) \right) K_{ti},
\]

for $i, j = 1, ..., m$. Then, by TA Lemma 4.22 (involving assumptions $H_{2.3}$, $H_{2.4}(b)$, $H_{2.5}$ and $H_{2.7}(A)$), the two assumptions A.1-i and A.1-ii of Lemma A.1 are verified, entailing that the last two terms of (A.8) also tend to zero almost surely.

As a conclusion, the asymptotic convergence of the estimator $\hat{\theta}_{n}$ towards the normal distribution is ensured and the proof of Theorem 2 is achieved. \hfill \Box

Remark A.1. The assumptions $H_{2.1}$-$H_{2.8}$ are a generalization of those in Azrak & Mélard (2006), where for instance $P(\nu)$ is simply equal to 1.

Remark A.2. To show an example on how bounds are derived, let us consider an
upper bound of the absolute value of the expectation at $\theta = \theta^0$ of the scalar

$$\frac{\partial e_t^T(\theta)}{\partial \theta_i} I_m \frac{\partial e_t(\theta)}{\partial \theta_j}. \quad (A.10)$$

This is done in two stages.

First, using (3.7), the expression (A.10) is equal to

$$\sum_{k_1=1}^{t-1} \sum_{k_2=1}^{t-1} \text{tr}\{\epsilon_{t-k_1}^T g_{t-k_1}^T \psi_{tik_1}(\theta, \theta^0) \psi_{tjk_2}(\theta, \theta^0) g_{t-k_2} \epsilon_{t-k_2} \} = \sum_{k_1=1}^{t-1} \sum_{k_2=1}^{t-1} \text{tr}\{g_{t-k_1}^T \psi_{tik_1}(\theta, \theta^0) \psi_{tjk_2}(\theta, \theta^0) g_{t-k_2} \epsilon_{t-k_2}^T \}. \quad (A.11)$$

Taking the expectation at $\theta = \theta^0$ leads to a simple sum

$$\sum_{k=1}^{t-1} \text{tr}\{g_{t-k}^T \psi_{tik} \psi_{tjk} g_{t-k} \Sigma\}. \quad (A.11)$$

For other cases, we need to deal with other matrices than $I_m$, possibly $\Sigma^{-1}(\theta)$ or its derivatives with respect to elements of $\theta$, or other derivatives of $e_t(\theta)$, or expectations of products of two expressions like in (A.10), sometimes centered, and possibly at different times $t$ and $t + d$. In the case of a product, it is necessary to move circularly factors within traces of products in order to make sure that random vectors are contiguous. Finally, to check the bounds in the conditions of Lemma A.2, we need to treat products of conditional expectations, given $F_{t-\nu}$, of expressions like (A.10). It is here that sums from $t = \nu$ are needed in assumption H$_{2.2}$. All these cases are treated in a single TA Lemma 4.3. It is used in the proofs of TA Lemma’s 4.7, 4.9, 4.11, 4.13, 4.15, 4.16, 4.18, 4.20, 4.22.

Second, to obtain an upper bound of the absolute value of (A.11), we use standard inequalities (recalled in the TA). On the example of (A.11), we obtain successively

$$\sum_{k=1}^{t-1} \|g_{t-k}\|_F^2 \|\psi_{tik}\|_F \|\psi_{tjk}\|_F \|\Sigma\|_F \leq m_1 \|\Sigma\|_F \left( \sum_{k=1}^{t-1} \|\psi_{tik}\|_F^2 \times \sum_{k=1}^{t-1} \|\psi_{tjk}\|_F^2 \right)^{1/2} \leq m_1 N_1 \|\Sigma\|_F,$$
using the first part of assumption $H_{2,2}$ for $\nu = 1$, given $P(1) = 1$, and assumption $H_{2,5}$. This bound is used in the proof of TA Lemma 4.8, more precisely for proving (4.35) there. In order to obtain some of the requested bounds in the other cases, we have to use other of our assumptions, e.g. other parts of $H_{2,2}$, with derivatives of order 2 and 3 of $e_1(\theta)$, or $H_{2,3}$ when derivatives of $\Sigma_\theta(\theta)$ or its inverse are involved. All the cases are treated in a single TA Lemma 4.7. It is used in the proofs of TA Lemmas 4.8, 4.11, 4.13, 4.15, 4.17, 4.19, 4.21.
S1. Checking the assumptions for the examples

We check the assumptions of Theorem 2 for the two examples of Section 4, hereby providing a theoretical foundation for the simulation results of Section 5. Since several of the assumptions are somewhat similar, we have only covered once each argument. Also, since Example 2 is a generalization of Example 1, we have avoided to repeat some of the verifications when they are too similar. Assumptions that are trivially verified are omitted.

S1.1 Example 1

S1.1.1 Assumption H_{2.2}

In order check this hypothesis, we shall have recourse to the results of Section 3.2. In this example the coefficients of the pure moving average representation of (4.1) are given by direct application of Section 3.3. Then, by using (3.9), we can calculate \( \psi_{tk}(\theta, \theta^0) \). For example, denoting \( A_{t-1} = \prod_{l=1}^{k-1} A_{t-l}(\theta) \), and its \((i,j)\) element \( A_{t,i,j}^{(k-1)}, i, j = 1, 2 \), we have \( \psi_{tk} = A_{t+1}^{(k-1)} \) and

\[
\psi_{tk}(\theta, \theta^0) = \frac{\partial \pi_{11}(\theta)}{\partial \theta_i} \prod_{l=1}^{k-1} A_{t-l}(\theta^0) = \frac{\partial A_{t}(\theta)}{\partial \theta_i} A_{t}^{(k-1)},
\]

for \( k = 1, 2, \ldots, t - 1 \), where \( A_{t}^{(0)} = I_2 \). Obviously, checking the assumptions in the general setup happens to be a complicated and tedious task, hence, for the sake of simplification, we shall consider the model defined in (4.3). It can be shown by induction that

\[
A_{t}^{(k-1)} = \begin{pmatrix} A_{t,1,1}^{(k-1)} & A_{t,1,2}^{(k-1)} \\ 0 & A_{t,2,2}^{(k-1)} \end{pmatrix},
\]

(A.12)

for \( k \geq 2 \), where

\[
A_{t,1,1}^{(k-1)} = (A_{t,11}^{(0)})^{k-1} \prod_{l=1}^{k-1} \sin(a(t-l)), \quad A_{t,2,2}^{(k-1)} = (A_{t,22}^{(0)})^{k-1} \prod_{l=1}^{k-1} \sin(b(t-l)),
\]

(A.13)

\[
A_{t,1,2}^{(k-1)} = \frac{1}{2} \sum_{l=1}^{k-1} (A_{t,11}^{(0)})^{k-1-l}(A_{t,22}^{(0)})^{l-1} \prod_{f=1}^{k-2} \sin(c_{lf}(t-f-\delta_{lf})),
\]

(A.14)
and $c_{lf} = a$ and $\delta_{lf} = 0$, for $l + f \leq k - 1$, and $c_{lf} = b$ and $\delta_{lf} = 1$, for $l + f > k - 1$.

For example, for $k = 4$

$$A_{t,1,2}^{(3)} = \frac{1}{2} \left( (A_{11}^0)^2 \sin(a(t - 1)) \sin(a(t - 2)) + A_{11}^0 A_{22}^0 \sin(a(t - 1)) \sin(b(t - 3)) + (A_{22}^0)^2 \sin(b(t - 2)) \sin(b(t - 3)) \right).$$

For $\theta_1 = A_{11}'$ and $\theta_2 = A_{22}'$, and using (A.12) we have

$$\frac{\partial A_t(\theta)}{\partial \theta_1} = \begin{pmatrix} \sin(at) & 0 \\ 0 & 0 \end{pmatrix}, \quad \psi_{1lk} = \sin(at) \begin{pmatrix} A_{t,1,1}^{(k-1)} & A_{t,1,2}^{(k-1)} \\ 0 & 0 \end{pmatrix}, \quad (A.15)$$

$$\frac{\partial A_t(\theta)}{\partial \theta_2} = \begin{pmatrix} 0 & 0 \\ 0 & \sin(bt) \end{pmatrix}, \quad \psi_{2lk} = \sin(bt) \begin{pmatrix} 0 & 0 \\ 0 & A_{t,2,2}^{(k-1)} \end{pmatrix}. \quad (A.16)$$

In this case $\psi_{1ijk} = 0$ and $\psi_{2ijk} = 0$, for all $i, j, l = 1, 2$ and all $t$ and $k$. We define

$$\Phi^{1/2} = \max \left\{ \left| A_{11}^0 \right|, \left| A_{22}^0 \right| \right\}, \quad (A.17)$$

a constant which is such that $0 < \Phi < 1$. Now by using (A.13)-(A.17) and bounding sines by 1, we can show that

(a) $A_{t,1,1}^{(k-1)} \leq \Phi^{\frac{k-1}{2}}$,  \hspace{0.5cm} (b) $A_{t,1,2}^{(k-1)} \leq (k - 1)\Phi^{\frac{k-1}{2}}$  \hspace{0.5cm} (c) $A_{t,2,2}^{(k-1)} \leq \Phi^{\frac{k-1}{2}}. \quad (A.18)$

Consequently, from (A.16) and (A.18(c)), \sum_{k=\nu}^{t-1} \|\psi_{2lk}\|_F^2 = \sum_{k=\nu}^{t-1} \left[ \sin(bt)A_{t,2,2}^{(k-1)} \right]^2 \leq \sum_{k=\nu}^{t-1} \Phi^{k-1} \leq N_1 \Phi^{\nu-1}$, where $N_1 = 1/(1 - \Phi)$. By the same way, from (A.15) and (A.18(a)-(A.18(b)),

$$\sum_{k=\nu}^{t-1} \|\psi_{1lk}\|_F^2 = \sum_{k=\nu}^{t-1} \sin^2(at) \left[ (A_{t,1,1}^{(k-1)})^2 + (A_{t,1,2}^{(k-1)})^2 \right] \leq \Phi^{k-1} + \sum_{k=\nu}^{t-1} (k - 1)^2 \Phi^{k-2}, \quad (A.19)$$

but this cannot be bounded by $N_1 \Phi^{\nu-1}$ for some constant $N_1$, independently of $\nu$, but well by $N_1 P(\nu) \Phi^{\nu-1}$, where $P(\nu)$ is a polynomial of degree 2. For the purpose of the verification of the subsequent assumptions, a more subtle upper bound can be found.
Then, for any given \( t \)

\[
\text{Frobenius norm can be bounded by some expression }
\]

\[
g
\]

is a strictly positive definite matrix. From (3.7) we have for \( i, j \)

\[
I
\]

taken \( \Sigma = \) the term does exist and is denoted \( V \)

vanishes. Hence, taking \( \nu > \tilde{k} \), the second term of (A.20) vanishes for all \( t \) and the Frobenius norm can be bounded by some expression \( N_1 \Phi^{\nu-1} \) for some \( N_1 \). Note also that, under the conditions of integral periods, we have \( \| \psi_{t1k} \|^2_F = \| \psi_{2k} \|^2_F = 0, k > \tilde{k} \).

**S1.1.2 Assumption H_{2.6}**

The second term in \( H_{2.6} \) is equal to 0, so it remains to show that the limit of the first term does exist and is denoted \( V_{ij} \), for \( i, j = 1, 2 \), where the matrix \( V = (V_{ij})_{1 \leq i, j \leq 2} \) is a strictly positive definite matrix. From (3.7) we have for \( i, j = 2 \), since we have taken \( \Sigma = I_2 \), using (3.4) and (A.16)

\[
E_{\theta_0}\left( \frac{\partial e_i^T(\theta)}{\partial \theta_2} \Sigma^{-1}_t(\theta) \frac{\partial e_i(\theta)}{\partial \theta_2} \right) = \text{tr} \left[ \left\{ \frac{\partial A_i^T(\theta)}{\partial \theta_2} \frac{\partial A_i(\theta)}{\partial \theta_2} \right\}_{\theta=\theta_0} E(x_{t-1} x_{t-1}^T) \right]
\]

\[
= \sin^2(bt) \sum_{k=0}^{t-2} \left( A_{1,2,2}^{(k)} \right)^2.
\]

Denote this by \( V_{22}(t) \), a strictly positive periodic function of \( t \) with period \( P_2 \). Suppose that \( P_2 \) is an integer. Letting \( S > 0 \) denote the sum of \( P_2 \) consecutive terms of \( V_{22}(t) \), and \( s_n \) the integer quotient of \( n \) by \( P_2 \), \( n^{-1} \sum_{t=1}^n V_{22}(t) \) is equal to \( (s_n S + \ldots) \).
\[ \sum_{t=s_n, P_2+1}^{n} V_{22}(t)/n. \] When \( n \to \infty \), the first term tends to \( S/P_2 \) and the second term tends to 0. Thus the limit \( V_{22} > 0 \) does at least exist when \( P_2 \) is an integer. By the same method for \( i, j = 1 \), using (A.15), \( V_{11}(t) \), the term \( t \) of \( V_{11} \) is equal to

\[ \text{tr} \left[ \left\{ \frac{\partial A_t^T(\theta) \partial A_t(\theta)}{\partial \theta_1 - \partial \theta_1} \right\}_{\theta_0} E(x_{t-1} x_{t-1}^T) \right] = \sin^2(at) \sum_{k=0}^{t-2} \left[ \left( A_{t,1,1}^{(k)} \right)^2 + \left( A_{t,1,2}^{(k)} \right)^2 \right]. \]

(A.22)

The sum is strictly positive, at least for some \( t \)'s. As a function of \( t \), \( V_{11}(t) \) is a periodic function of \( t \) with period \( P_1 P_2 \) at most. If this is an integer, the same arguments as for \( V_{22} \) in (A.21) can be used. Therefore, taking the average for \( t = 1 \) to \( n \) gives a finite strictly positive limit \( V_{11} \) when \( n \to \infty \). Furthermore, it can easily be seen that \( V_{12} = V_{21} = 0 \). Consequently, the matrix \( V = (V_{ij})_{1 \leq i, j \leq 2} \) is strictly positive definite.

If the process is Gaussian, then \( W = V \). Otherwise, the existence of the limit \( W \) has to be proven and this is surely more complex.

**Remark A.3.** If we handle the model described in Section 5, but without the parameter \( A_1^0 \), for the case where the number of observations is 25, we obtain the following standard errors from that theoretical formula for the two remaining parameters: 0.2175 and 0.2303, respectively, which largely agree with the averages drawn from the simulation results: 0.2161 and 0.2226, as shown in Table 1.

**S1.1.3 Assumption H_{2.7}**

For the first part of this assumption, since \( g_t \) is bounded, there remains to show that

\[ \frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k=1}^{t-1} \| \psi_{tk} \|_F \| \psi_{t+d,i,k+d} \|_F = O \left( \frac{1}{n} \right). \]

For \( i = 1 \) we have

\[ \sum_{k=1}^{t-1} \| \psi_{tk} \|_F \| \psi_{t+d,1,k+d} \|_F = | \sin(at) \sin(a(t + d)) | \]

\[ \times \sum_{k=1}^{t-1} \left[ \left( A_{t,1,1}^{(k-1)} \right)^2 + \left( A_{t,1,2}^{(k-1)} \right)^2 \right]^{1/2} \left[ \left( A_{t+d,1,1}^{(k-d-1)} \right)^2 + \left( A_{t+d,1,2}^{(k-d-1)} \right)^2 \right]^{1/2}. \] (A.23)

To simplify the proof we assume again that \( a = 2\pi/P_1 \) and \( b = 2\pi/P_2 \), for some integers \( P_1 \) and \( P_2 \) and use \( \tilde{k} \) as defined above. Then (A.23) becomes a sum for \( k = 1 \)
Given that $\Phi < 1$, the general term can be bounded by $[\Phi^{k-2}(1+(k-1)^2)\Phi^{k+d-2}(1+(k + d - 1)^2)]^{1/2}$ which is of order $d\Phi^{d/2}$. Hence

$$\sum_{d=1}^{n-1} \sum_{t=1}^{n-d} (A.23) = \sum_{t=1}^{n-1} \sum_{d=1}^{n-t} (A.23)$$

$$\leq k_1(\Phi) \sum_{t=1}^{n-1} \sum_{d=1}^{n-t} \Phi^{d/2}d$$

$$\leq k_1(\Phi) \frac{\Phi^{1/2}}{(1 - \Phi^{1/2})^2}(n - 1),$$

which is bounded by $k_2(\Phi)n$, where $k_1(\Phi)$ and $k_2(\Phi)$ are constants and where we have used the formula $\sum_{j=1}^{\infty} jx^j = x/(1 - x)^2$, provided $|x| < 1$. Dividing by $n^2$ thus gives $O(1/n)$, as requested.

Applying the same method for $i = 2$, with again a sum for $k = 1$ to $\tilde{k}$ but of $A_{i+1,2}^{(k-1)} A_{i+d+2,2}^{(k+d-1)}$ instead of the product of square roots of sums of squares in (A.23), with a general term bounded this time by $\Phi^{(k-1)/2}\Phi^{(k+d-1)/2}$, we can show that

$$\sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k=1}^{t-1} \frac{\psi_{t+2,k}}{F} \frac{\psi_{t+4,k+d+2}}{F} \leq \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} k_3(\Phi)\Phi^{d/2},$$

which is bounded by $k_4(\Phi)n$, with other constants $k_3(\Phi)$ and $k_4(\Phi)$ and reach the same final conclusion as for $i = 1$.

For the second part of this assumption, since $\Xi_{i}(\Sigma)$, $\Sigma \otimes \Sigma$ and $\text{vec}(\Sigma)\text{vec}(\Sigma)^T$ are finite constants, similarly to the first part we can show that the second part of assumption $H_{2.7}$ is fulfilled.

**S1.2 Example 2**

The verification of assumption $H_{2.2}$ remains unchanged but requires the same conditions on integers $P_1, P_2$ as for the previous example. Note that from (4.4) the vector of the parameters is $\theta = (\Lambda_{11}^0, \Lambda_{22}^0, \eta_{11}^0, \eta_{22}^0)\text{T}$ and $\theta^0 = (\Lambda_{11}^0, \Lambda_{22}^0, \eta_{11}^0, \eta_{22}^0)\text{T}$ is its true value. Note that we assume again that $\Phi$ is of the form (A.17).

**S1.2.1 Assumption H_{2.6}**

We start by making an important observation, namely that the matrix $V = (V_{ij})_{1 \leq i, j \leq 4}$ is block-diagonal, with blocks for $(i, j) \in \{1, 2\}$ and $(i, j) \in \{3, 4\}$, respectively.
Regarding the second block, tedious calculations for \( \theta_3 = \eta_{11} \) yield

\[
\text{tr} \left\{ \Sigma_t^{-1}(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_3} \Sigma_t^{-1}(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_3} \right\}_{\theta = \theta^0} = 2 \sin^2(ct) \frac{\left( e^{\eta_{12} \sin(ct)} s_{11} - s_{12} \right)^2 + 2(s_{11}s_{22} - s_{12}^2)}{(1 + e^{(\eta_{11}+\eta_{22}) \sin(ct)})^2 (s_{11}s_{22} - s_{12}^2)} = V_{33}(t),
\]

where \( V_{ij}(t) \) is the term \( t \) in the sum defining \( V_{ij} \), \( i, j = 1, 2, 3, 4 \). Since \( \Sigma \) is an invertible matrix, \( s_{11}s_{22} - s_{12}^2 = \det(\Sigma) > 0 \), hence this expression is clearly positive. Bounding the numerator is straightforward (since \( |\sin(ct)| < 1 \)), and \( 1/(1 + e^{(\eta_{11}+\eta_{22}) \sin(ct)}) \) can simply be bounded by \( 1 \). Hence, there exists a constant \( \alpha \) not depending on \( t \) such that \( V_{33}(t) \leq \alpha \). This is a strictly positive periodic function of \( t \) with period \( P_3 = 2\pi/c \). If \( P_3 \) is an integer, the same arguments as in Section S1.1.2 can be used to prove existence of the limit \( V_{33} \). The same conclusion obviously holds for \( V_{44} \). Turning our attention towards \( V_{34} \), we obtain the \( t \)-term \( V_{34}(t) = \)

\[
\text{tr} \left\{ \Sigma_t^{-1}(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_3} \Sigma_t^{-1}(\theta) \frac{\partial \Sigma_t(\theta)}{\partial \theta_4} \right\}_{\theta = \theta^0} = 2 \sin^2(ct) \frac{s_{12} \left( e^{\eta_{12} \sin(ct)} s_{11} - s_{12} - e^{\eta_{11} \sin(ct)} s_{22} \right) - e^{(\eta_{11}+\eta_{22}) \sin(ct)} (s_{11}s_{22} - 2s_{12}^2)}{(1 + e^{(\eta_{11}+\eta_{22}) \sin(ct)})^2 (s_{11}s_{22} - s_{12}^2)}.
\]

Again, it is a simple exercise to show that this term is bounded independently of \( t \) and to establish the existence of a limit, at least when \( P_3 \) is an integer.

Let us turn our attention towards the block \((1, 2)\) now. Following Section S1.1.2 for \( i, j = 1 \) with \( \theta_1 = A'_{11} \), given (3.6), we obtain a generalized version of the left-hand side of (A.22) as

\[
\text{tr} \left\{ \left( \frac{\partial A_t^T(\theta)}{\partial \theta_1} \Sigma_t^{-1}(\theta) \frac{\partial A_t(\theta)}{\partial \theta_1} \right)_{\theta = \theta^0} E(x_{t-1}x_{t-1}^T) \right\} = \sin^2(at) \sum_{k=1}^{t-1} \text{tr} \left[ \begin{pmatrix} A_{1,1,1}^{(k-1)} & A_{1,1,2}^{(k-1)} \\ A_{1,1,1} & 0 \end{pmatrix} \Sigma_{t-k} \begin{pmatrix} A_{1,1,1}^{(k-1)} & A_{1,1,2}^{(k-1)} \\ 0 & 0 \end{pmatrix} \Sigma_{t-1}^{-1} \right],
\]

an expression clearly bounded (under the same conditions as those in Section S1.1.2). Its positiveness follows from the fact that we consider a Mahalanobis distance in the metric \( \Sigma^{-1} \). We have again a periodic period of \( t \), but this time the period is at most \( P_1P_2P_3 \). When the three periods are integers, convergence is proved like before.

Finally, checking that the blocks \((1, 2)\) and \((3, 4)\) are positive definite has been
done numerically in the numerical example of Section 5.2.

**S1.2.2 Assumption H_{2,7}**

In the first part of this assumption we have to show that

\[
\frac{1}{n^2} \sum_{d=1}^{n-1} \sum_{t=1}^{n-d} \sum_{k=1}^{t-1} \|g_{t-k}\|_F^2 \|\psi^{(n)}_{tik}\|_F \|\psi^{(n)}_{t+d,i,k+d}\|_F = O \left( \frac{1}{n} \right).
\]

From (4.6) we have \(\|g_{t-k}\|_F^2 = 2 + e^{-2\eta_1 \sin(ct)} + e^{-2\eta_2 \sin(ct)}\) which is an easily bounded function. Consequently, we are left with exactly the same expression as for Example 1 of Section 4.1, which solves the question.

**References**
