Estimating RCARMA models with uncorrelated but non-independent error terms

Yacouba Boubacar Maïnassara and **Landy Rabehasaina**

Université Bourgogne Franche-Comté,
Laboratoire de mathématiques de Besançon,
UMR CNRS 6623,
16 route de Gray,
25030 Besançon, France.

e-mail: yacouba.boubacar_mainassara@univ-fcomte.fr; lrabehas@univ-fcomte.fr

Abstract: In this paper we derive the asymptotic properties of the least squares estimator (LSE) of random coefficient autoregressive moving-average (RCARMA) models under the assumption that the errors are uncorrelated but not necessarily independent. Relaxing the independence assumption considerably extends the range of application of the class of RCARMA models. Conditions are given for the consistency and asymptotic normality of the LSE. A particular attention is given to the estimation of the asymptotic variance matrix, which may be very different from that obtained in the standard framework. A set of Monte Carlo experiments is presented.

AMS 2000 subject classifications: Primary 62M10, 62F03, 62F05; secondary 91B84, 62P05.
Keywords and phrases: Least square estimation, Random coefficients, weak ARMA models.

1. Introduction

Since the works of [26, 27] and [35], the time series models with time-varying coefficients have become increasingly popular. In statistical applications, a large part of the literature is devoted to the non-stationary autoregressive moving-average (ARMA) models with time-varying parameters (see [3, 4, 7, 12]), see also the class of ARMA models with periodic coefficients (for instance [1, 5]). But the most popular class deals with the treatment of regime shifts and non-linear modeling strategies. For instance, a Markov-switching model is a non-linear specification in which different states of the world affect the evolution of a time series (see, for examples, [18, 28, 30]). The asymptotic properties of Markov-switching ARMA models are well known in the literature (see, for instance, [8, 19, 21, 22, 33] or [29]).

The fact that changes in regimes may be very important for the evolution of interest rates has been emphasized in a number of recent studies. Our attention here is focused on the class of random coefficient ARMA (RCARMA) models with recurrent but non necessarily periodic changes in regime. We consider a time series $(X_t)_{t\in\mathbb{Z}}$ exhibiting changes in regime at known dates and we suppose that we have finite regimes. Contrarily to the famous Markov-switching approach, we assume that the realization of the regimes is observed. For such models, [16, 17] gave general conditions ensuring consistency and asymptotic normality of least squares (LS) and quasi-generalized least-squares (QGLS) estimators under the assumption that the innovation processes is independent. This independence assumption is often considered too restrictive by practitioners. We considerably relax this independence assumption on the errors to extend the range of application of the RCARMA models.

This paper is devoted to the problem of estimating RCARMA representations under the assumption that the errors are uncorrelated but not necessarily independent. Thus, the main goal of our paper is to complete the above-mentioned results concerning the statistical analysis of RCARMA models with uncorrelated but not necessarily independent innovation, by considering the estimation problem under general error terms. We establish the asymptotic distribution of the LS estimator of RCARMA models when the linear innovations are not necessarily independent, but are strongly mixing.

The paper is organized as follows. Section 2 presents the RCARMA models that we consider here. In Section 3, we established the strict stationarity condition and it is shown that the LS estimator (LSE) is asymptotically normally distributed when linear innovation process (ϵ_t) satisfies mild mixing assumptions. The asymptotic variance of the LSE may be very different in the weak and strong cases. Particular attention is given to the estimation of this covariance matrix. Modified version of the Wald test is proposed for testing linear restrictions on the parameters. In Section 4, we present two examples of weak RCAR(1) models with iid and correlated realization of the regimes. Numerical experiments are presented in Section 5. The proofs of the main results are collected in the appendix.

2. Model and assumptions

Let $(\Delta_t)_{t\in\mathbb{Z}}$ be a stationary ergodic observed process with values in a finite set $\mathcal{S}$ of size $\mathrm{Card}(\mathcal{S}) = K$. We consider the random coefficient ARMA(p, q) process $(X_t)_{t\in\mathbb{Z}}$ defined by

$$X_t - \sum_{i=1}^{p} a_i^0(\Delta_t)X_{t-i} = \epsilon_t - \sum_{j=1}^{q} b_j^0(\Delta_t)\epsilon_{t-j} \tag{1}$$

where the linear innovation process $\epsilon := (\epsilon_t)_{t\in\mathbb{Z}}$ is assumed to be a stationary sequence satifies $\mathbb{E}(\epsilon_t) = 0, \quad \mathbb{E}(\epsilon_t\epsilon_{t'}) = \sigma^2\mathbb{1}_{[t=t']}$. Under the above assumptions, the process ϵ is called a weak white noise.

This representation is said to be a weak RCARMA(p, q) representation under the assumption that ϵ is a weak white noise. For the statistical inference of ARMA models, the weak white noise assumption is often replaced by the strong white noise assumption, *i.e.* the assumption that ϵ is an independent and identically distributed (iid for short) sequence of random variables with mean 0 and common variance. Obviously the strong white noise assumption is more restrictive than the weak white noise assumption, because independence entails uncorrelatedness. Consequently weak RCARMA representation is more general than the strong one.

The unknown parameter of interest denoted $\theta_0 := (a_i^0(s), b_j^0(s),\ i = 1,\dots,p,\ j = 1,\dots,q,\ s \in \mathcal{S})$ lies in a compact set of the form

$$\Theta \subset \left\{(a_i(s), b_j(s),\ i = 1,\dots,p,\ j = 1,\dots,q,\ s \in \mathcal{S}) \in \mathbb{R}^{(p+q)\times K}\right\},$$

with non empty interior, within which we suppose that θ_0 lies. The parameter σ^2 is considered as a nuisance parameter. In order to estimate θ_0, one thus has at our disposal observations (X_t, Δ_t), $t = 1,\dots,n$, from which one aims at building an strongly consistent and asymptotically normal estimator $\hat{\theta}_n$. The strong mixing coefficients $(\alpha_Z(h))_{h\in\mathbb{Z}}$ of a stationary process

$(Z_t)_{t\in\mathbb{Z}}$ are defined by

$$\alpha_Z(h) := \sup_{A\in\mathcal{F}_{-\infty}^t,\ B\in\mathcal{F}_{t+h}^\infty} \left|\mathbb{P}(A\cap B) - \mathbb{P}(A)\cdot\mathbb{P}(B)\right|, \tag{2}$$

where $\mathcal{F}_{-\infty}^t$, and $\mathcal{F}_{t+h}^\infty$ be the σ-fields generated by $\{Z_u,\ u\le t\}$ and $\{Z_u,\ u\ge t+h\}$, respectively. Let us suppose the following assumptions.

(**A1**) The processes $(\epsilon_t)_{t\in\mathbb{Z}}$ and $(\Delta_t)_{t\in\mathbb{Z}}$ are ergodic sequences, strictly stationary, independent from each other.

(**A2**) For some $\nu>0$, the processes $(\epsilon_t)_{t\in\mathbb{Z}}$ and $(\Delta_t)_{t\in\mathbb{Z}}$ satisfy $\sum_{h=0}^\infty \alpha_\epsilon(h)^{\frac{\nu}{\nu+2}}<+\infty$ and $\sum_{h=0}^\infty \alpha_\Delta(h)^{\frac{\nu}{\nu+2}}<+\infty$.

(**A3**) The process $(\epsilon_t)_{t\in\mathbb{Z}}$ also satisfies $\mathbb{E}[|\epsilon_t|^{2\nu+4}]<+\infty$.

(**A4**) We have $\theta_0\in\overset{\circ}{\Theta}$, where $\overset{\circ}{\Theta}$ denotes the interior of Θ.

We introduce the following notation so as to emphasize dependence of unknown parameter θ_0 in (1). For all $\theta=(a_i(s),b_j(s),\ i=1,\dots,p,\ j=1,\dots,q,\ s\in\mathcal{S})\in\Theta$, we let $\underline{a_i}:=(a_i(s),s\in\mathcal{S})$, $i=1,\dots,p$ and $\underline{b_j}:=(b_j(s),s\in\mathcal{S})$, $j=1,\dots,q$. Let $\mathbf{e}(s)$ be the row vector of size $1\times K$ such that the ith component is $\mathbb{1}_{[s=i]}$. Then one notices that $\forall t\in\mathbb{Z}$

$$a_i(\Delta_t) =< \mathbf{e}(\Delta_t),\underline{a_i} >:= g_i^a(\Delta_t,\theta),\quad b_j(\Delta_t) =< \mathbf{e}(\Delta_t),\underline{a_j} >:= g_j^b(\Delta_t,\theta),\quad i=1,\dots,p,\ j=1,\dots,q,$$

where $<\cdot,\cdot>$ denotes the scalar product between vectors of appropriate dimension. Thus (1) reads

$$X_t - \sum_{i=1}^p g_i^a(\Delta_t,\theta_0)X_{t-i} = \epsilon_t - \sum_{j=1}^q g_j^b(\Delta_t,\theta_0)\epsilon_{t-j}. \tag{3}$$

Let us furthermore note that for all i, j and s, $g_i^a(s,\theta)$ and $g_j^b(s,\theta)$ are *linear* in θ. We thus introduce matrices

$$A(s) := \begin{pmatrix} g_1^a(s,\theta_0) & \cdots & \cdots & g_p^a(s,\theta_0)\\ & & & 0\\ & I_{p-1} & & \vdots\\ & & & 0\end{pmatrix},\quad B(s,\theta) := \begin{pmatrix} g_1^b(s,\theta) & \cdots & \cdots & g_q^b(s,\theta)\\ & & & 0\\ & I_{q-1} & & \vdots\\ & & & 0\end{pmatrix}$$

for all $s\in\mathcal{S}$, $\theta\in\Theta$. A remark that will prove useful later on is that $\theta\mapsto B(s,\theta)$ is, for all $s\in\mathcal{S}$, an *affine* function.

We next introduce the residues corresponding to parameter $\theta\in\Theta$ as the stationary process $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ satisfying

$$\epsilon_t(\theta) - \sum_{j=1}^q g_j^b(\Delta_t,\theta)\epsilon_{t-j}(\theta) = X_t - \sum_{i=1}^p g_i^a(\Delta_t,\theta)X_{t-i},\quad \forall t\in\mathbb{Z}. \tag{4}$$

This process is unique in L^2, as explained in Proposition 3.1. In particular, one has $(\epsilon_t(\theta_0))_{t\in\mathbb{Z}} = (\epsilon_t)_{t\in\mathbb{Z}}$, the initial white noise. We next define the approximating residue as the process $(e_t(\theta))_{t\in\mathbb{Z}}$ verifying

$$e_t(\theta) - \sum_{j=1}^q g_j^b(\Delta_t,\theta)e_{t-j}(\theta) = \tilde{X}_t - \sum_{i=1}^p g_i^a(\Delta_t,\theta)\tilde{X}_{t-i},\quad \forall t\in\mathbb{Z}, \tag{5}$$

where values corresponding to negative indices are set to zero, i.e. processes $(e_t(\theta))_{t\in\mathbb{Z}}$ and $(\tilde{X})_{t\in\mathbb{Z}}$ verify

$$\begin{aligned} e_t(\theta) &= 0, \quad t \le 0, \\ \tilde{X}_t &= X_t \mathbb{1}_{[t\ge 1]}, \quad \forall t \in \mathbb{Z}. \end{aligned}$$

The basic idea behind definition of $(e_t(\theta))_{t\in\mathbb{Z}}$ is that, given a realization $X_1, X_2, \ldots, X_n$ of length n, $\epsilon_t(\theta)$ is approximated, for $0 < t \le n$, by $e_t(\theta)$. Next, we define the cost function

$$Q_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} e_t^2(\theta). \tag{6}$$

Finally, we let for all $n \in \mathbb{N}$ the random variable $\hat{\theta}_n$ the least squared estimator that satisfies, almost surely,

$$Q_n(\hat{\theta}_n) = \min_{\theta\in\Theta} Q_n(\theta), \tag{7}$$

We finish this section by giving some notation. In the following, $||.||$ will denote the norm of matrices or vectors of appropriate size, depending on the context, whereas $||.||_p$ will denote the L^p norm defined by $||X||_p = [\mathbb{E}(|X|^p)]^{1/p}$ for all random variable X admitting a $p-$th order moment, $p \ge 1$. For all matrix M, M' will denote its transpose. For all three times differentiable function $f : \Theta \longrightarrow \mathbb{R}$, we will let $\nabla f(\theta) = \left(\frac{\partial}{\partial\theta_k} f(\theta)\right)_{k=1,\ldots,(p+q)K}$, $\nabla^2 f(\theta) = \left(\frac{\partial^2}{\partial\theta_i\partial\theta_j} f(\theta)\right)_{i,j=1,\ldots,(p+q)K}$ and $\nabla^3 f(\theta) = \left(\frac{\partial^3}{\partial\theta_\ell\partial\theta_i\partial\theta_j} f(\theta)\right)_{\ell,i,j=1,\ldots,(p+q)K}$ respectively the first, second and third order derivatives with respect to the variable θ.

3. Case of general correlated process $(\boldsymbol{\Delta_t})_{t\in\mathbb{Z}}$

In this section, we display our main results.

3.1. Stationarity

A first step consists in giving sufficient conditions such that processes $(X_t)_{t\in\mathbb{Z}}$ and $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ defined in (1) and (4) are stationary and admits moments of order 2. Let $||.||$ be any norm on the set of matrices, and let us introduce the following notation

$$\begin{aligned}
e_1 &:= (1,0,\dots,0) \in \mathbb{R}^{p+q}, \\
e_{p+1} &:= (e_{p+1,i})_{i=1,\dots,p+q}, \quad e_{p+1,1}=1, \quad e_{p+1,i} = \mathbb{1}_{[i=p+1]},\ i=2,\dots,p+q, \\
M &:= (m_{ij})_{i,j=1,\dots,p+q}, \quad m_{i,j} = \mathbb{1}_{[i=q+1,j=1 \text{ or } i=1,j=1]}, \\
\Phi(s,\theta) &:= \left(\begin{array}{c|ccc} B(s,\theta) & \begin{array}{ccc} g_1^a(s,\theta) & \cdots & g_p^a(s,\theta) \\ & 0 & \end{array} \\ \hline 0 & \begin{array}{ccc} 0 & \cdots & 0 \\ & I_{p-1} & \vdots \\ & & 0 \end{array} \end{array}\right), \quad s\in\mathcal{S},\ \theta\in\Theta, \\
\Psi(s) &:= \left(\begin{array}{c|ccc} A(s) & \begin{array}{ccc} g_1^b(s,\theta_0) & \cdots & g_q^b(s,\theta_0) \\ & 0 & \end{array} \\ \hline 0 & \begin{array}{ccc} 0 & \cdots & 0 \\ & I_{q-1} & \vdots \\ & & 0 \end{array} \end{array}\right), \quad s\in\mathcal{S}.
\end{aligned}$$

As for $B(s,\theta)$, one also notices in particular that $\theta \mapsto \Phi(s,\theta)$ is an affine function for all $s\in\mathcal{S}$. One has the following result.

Proposition 3.1. *Let us suppose that*

$$\limsup_{t\to\infty} \frac{1}{t} \ln \mathbb{E}\left(\sup_{\theta\in\Theta}\left\|\prod_{i=1}^{t}\Phi(\Delta_i,\theta)\right\|^8\right) < 0, \quad \limsup_{t\to\infty} \frac{1}{t} \ln \mathbb{E}\left(\left\|\prod_{i=1}^{t}\Psi(\Delta_i)\right\|^8\right) < 0, \tag{8}$$

then for all $t\in\mathbb{Z}$ and $\theta\in\Theta$, the unique stationary solution to (4) is given by

$$\epsilon_t(\theta) = \sum_{i=0}^{\infty} c_i(\theta,\Delta_t,\dots,\Delta_{t-i+1})\epsilon_{t-i}, \quad \textit{where} \tag{9}$$

$$c_i(\theta,\Delta_t,\dots,\Delta_{t-i+1}) = \sum_{k=0}^{i} e_1 \prod_{j=0}^{k-1}\Phi(\Delta_{t-j},\theta) M \prod_{j'=k}^{i-1}\Psi(\Delta_{t-j'})e'_{p+1}, \tag{10}$$

with the usual convention $\prod_i^j = 1$ if $i>j$. Furthermore, for each $t\in\mathbb{Z}$, sequence $(c_i(\theta,\Delta_t,\dots,\Delta_{t-i+1}))_{i\in\mathbb{N}}$ is unique in the set of sequences of random variables

$$\mathcal{H} := \left\{(d_i)_{i\in\mathbb{N}} \textit{ independent from } (\epsilon_t)_{t\in\mathbb{Z}} \textit{ s.t. } \mathbb{E}\left(\sum_{i=0}^{\infty} d_i^2\right) < +\infty\right\}.$$

Let us note that decomposition (9) is a slight generalization of the Wold decomposition of stationary processes which are squared integrable, see Theorem 5.7.1 p.187 of [11]. Also note that stability condition (8) is reminiscent of the one in [21, Theorem 1] (see also [10]); it is however stronger as we need integrability conditions for process $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ (as well as on its derivatives), uniformly on $\theta \in \Theta$.

Corollary 3.2. *The process $(e_t(\theta))_{t\in\mathbb{Z}}$ defined by (5) has the following decomposition*

$$e_t(\theta) = \sum_{i=0}^{\infty} c_i^e(t,\theta,\Delta_t,\ldots,\Delta_{t-i+1})\epsilon_{t-i}, \quad t \geq p+1, \text{ where} \tag{11}$$

$$c_i^e(t,\theta,\Delta_t,\ldots,\Delta_{t-i+1}) = \sum_{k=0}^{\min(t-1,i)} e_1 \prod_{j=0}^{k-1} \Phi(\Delta_{t-j},\theta) M \prod_{j'=k}^{i-1} \Psi(\Delta_{t-j'}) e'_{p+1}, \tag{12}$$

where matrix M and vectors e_1, e_{p+1}, are defined at the beginning of the section.

Lemma 3.3. *Random coefficients $c_i(\theta,\Delta_t,\ldots,\Delta_{t-i+1})$, $i \in \mathbb{Z}$, $t \in \mathbb{Z}$, verify the following properties:*

- *$\theta \mapsto c_i(\theta,\Delta_t,\ldots,\Delta_{t-i+1})$, $\theta \mapsto \nabla[c_i(\theta,\Delta_t,\ldots,\Delta_{t-i+1})]^2$ and $\theta \mapsto \nabla^2[c_i(\theta,\Delta_t,\ldots,\Delta_{t-i+1})]^2$ are a.s. polynomial functions,*
- *Let us assume, instead of (8), the stronger assumption*

$$(\mathbf{A5}) \limsup_{t\to\infty} \frac{1}{t} \ln \mathbb{E}\left(\sup_{\theta\in\Theta}\left|\left|\prod_{i=1}^{t}\Phi(\Delta_i,\theta)\right|\right|^{4\nu+8}\right) < 0, \quad \limsup_{t\to\infty} \frac{1}{t} \ln \mathbb{E}\left(\left|\left|\prod_{i=1}^{t}\Psi(\Delta_i)\right|\right|^{4\nu+8}\right) < 0$$

holds. Then one has

$$\begin{array}{rcl} \limsup_{i\to\infty} \frac{1}{i} \ln \mathbb{E}\left(\sup_{\theta\in\Theta}[c_i(\theta,\Delta_i,\ldots,\Delta_1)]^{2\nu+4}\right) & < & 0, \\ \limsup_{i\to\infty} \frac{1}{i} \ln \mathbb{E}\left(\sup_{\theta\in\Theta}\left|\left|\nabla^j[c_i(\theta,\Delta_i,\ldots,\Delta_1)]\right|\right|^{2\nu+4}\right) & < & 0, \quad j=2,3. \end{array} \tag{13}$$

Furthemore, coefficients $c_i^e(t,\theta,\Delta_{t-1},\ldots,\Delta_{t-i})$, $i\in\mathbb{Z}$, $t\geq 0$, satisfy

$$\begin{array}{rcl} \limsup_{i\to\infty} \frac{1}{i} \ln \sup_{t\geq 0} \mathbb{E}\left(\sup_{\theta\in\Theta}[c_i^e(t,\theta,\Delta_t,\ldots,\Delta_{t-i+1})]^{2\nu+4}\right) & < & 0, \\ \limsup_{i\to\infty} \frac{1}{i} \ln \sup_{t\geq 0} \mathbb{E}\left(\sup_{\theta\in\Theta}\left|\left|\nabla^j[c_i^e(t,\theta,\Delta_t,\ldots,\Delta_{t-i+1})]\right|\right|^{2\nu+4}\right) & < & 0, \quad j=2,3. \end{array} \tag{14}$$

3.2. Preliminary results

We define the cost function

$$O_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} \epsilon_t^2(\theta). \tag{15}$$

Similarly to $\hat{\theta}_n$, let us introduce estimator $\check{\theta}_n$ the least squared estimators corresponding to the cost function $O_n(\theta)$:

$$O_n(\check{\theta}_n) = \min_{\theta\in\Theta} O_n(\theta). \tag{16}$$

The following set of lemmas and propositions are necessary in order to prove the main consistency and normal asymptotics results for estimators $\hat{\theta}_n$ and $\check{\theta}_n$ defined in (16) and (7). They also justify that $e_t(\theta)$ asymptotically behaves as $\epsilon_t(\theta)$ as $t\to\infty$ for all θ.

Lemma 3.4. *Let us suppose that* **(A1)** *and that Stationarity condition* (8) *hold. Sequences* $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ *and* $(e_t(\theta))_{t\in\mathbb{Z}}$ *satisfy*

1. $||\sup_{\theta\in\Theta}|\epsilon_0(\theta)|||_4 < +\infty$ *and* $\sup_{t\geq 0}||\sup_{\theta\in\Theta}|e_t(\theta)|||_4 < +\infty$,
2. $||\sup_{\theta\in\Theta}|\epsilon_t(\theta) - e_t(\theta)|||_2$ *tends to* 0 *exponentially fast as* $t\to\infty$,
3. *For all* $\alpha > 0$, $t^\alpha \sup_{\theta\in\Theta}|\epsilon_t(\theta) - e_t(\theta)| \longrightarrow 0$ *a.s. as* $t\to\infty$,
4. *For all* $j = 1,2,3$, $\left|\left|\sup_{\theta\in\Theta}||\nabla^j\epsilon_0(\theta)||\right|\right|_4 < +\infty$, $\sup_{t\geq 0}\left|\left|\sup_{\theta\in\Theta}||\nabla^j e_t(\theta)||\right|\right|_4 < +\infty$ *and one has* $t^\alpha ||\sup_{\theta\in\Theta}||\nabla(e_t-\epsilon_t)(\theta)||||_{8/5} \longrightarrow 0$, $t^\alpha\left|\left|\sup_{\theta\in\Theta}||\nabla^2(e_t-\epsilon_t)(\theta)||\right|\right|_{4/3} \longrightarrow 0$ *and* $t^\alpha\left|\left|\sup_{\theta\in\Theta}||\nabla^3(e_t-\epsilon_t)(\theta)||\right|\right|_1 \longrightarrow 0$ *as* $t\to\infty$ *for all* $\alpha > 0$.

Proposition 3.5. *Under the same assumptions in Lemma 3.4, we have that, for all* $\alpha\in(0,1)$,

1. $\sup_{\theta\in\Theta}|Q_n(\theta) - O_n(\theta)|$ *converges a.s. to* 0, *and* $n^\alpha ||\sup_{\theta\in\Theta}|Q_n(\theta) - O_n(\theta)|||_1$ *tends to* 0 *as* $n\to\infty$,
2. $\sup_{\theta\in\Theta}||\nabla(Q_n(\theta) - O_n(\theta))||$ *and* $\sup_{\theta\in\Theta}||\nabla^j(Q_n(\theta) - O_n(\theta))||$, *for* $j = 2,3$ *converge a.s. to* 0,
3. $n^\alpha ||\sup_{\theta\in\Theta}|\nabla(Q_n - O_n)(\theta)|||_1 \longrightarrow 0$ *as* $n\to\infty$.

3.3. Asymptotic properties

Proposition 3.6. *Let* **(A1)**, **(A4)** *as well as stationarity condition* (8) *hold. Estimator* $\breve\theta_n$ *defined by (16) converges a.s. towards* θ_0.

Theorem 3.7 (Consistency of estimator)**.** *Let* **(A1)**, **(A4)** *as well as stationarity condition* (8) *hold. Estimator* $\hat\theta_n$ *defined by (7) converges a.s. towards* θ_0.

Theorem 3.8 (Asymptotic normality for the estimator)**.** *Let us suppose that assumptions* **(A1)** *to* **(A5)** *hold, and let* $\hat\theta_n$ *defined in (7). We have the following Central Limit Theorem*

$$\sqrt{n}\left(\hat\theta_n - \theta_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Omega := J^{-1}IJ^{-1}\right), \quad n\to+\infty, \tag{17}$$

matrices I *and* J *being defined as*

$$\begin{aligned} J &:= J(\theta_0) = 2\mathbb{E}\left(\nabla\epsilon_t(\theta_0)[\nabla\epsilon_t(\theta_0)]'\right), & (18)\\ I &:= I(\theta_0) = 4\sum_{k=-\infty}^{\infty}\mathbb{E}\left(\epsilon_t(\theta_0)\epsilon_{t-k}(\theta_0)\nabla\epsilon_t(\theta_0)[\nabla\epsilon_{t-k}(\theta_0)]'\right) & \\ &= \sum_{k=-\infty}^{+\infty}\mathrm{Cov}(\Upsilon_t,\Upsilon_{t-k}), \quad \textit{where} & (19)\\ \Upsilon_t &:= \Upsilon_t(\theta_0) = 2\epsilon_t(\theta_0)\nabla\epsilon_t(\theta_0). & (20)\end{aligned}$$

Remark 3.9. In the strong RCARMA case, *i.e.* when **(A1)** is replaced by the assumption that (ϵ_t) is iid, we have $I = 2J$, so that $\Omega = 2J^{-1}$. In the general case we have $I \neq 2J$. As a consequence the ready-made software used to fit RCARMA do not provide a correct estimation for weak RCARMA processes.

3.4. Estimating the asymptotic variance matrix

Theorem 3.8 can be used to obtain confidence intervals and significance tests for the parameters. The asymptotic variance Ω must however be estimated. The matrix J can easily be estimated by its empirical counterpart

$$\hat{J}_n = \frac{2}{n}\sum_{t=1}^{n}\nabla e_t(\hat{\theta}_n)[\nabla e_t(\hat{\theta}_n)]'.$$

In the standard strong RCARMA case $\hat{\Omega} = 2\hat{J}_n^{-1}$ is a strongly consistent estimator of Ω. In the general weak RCARMA case this estimator is not consistent when $I \neq 2J$ (see Remark 3.9). So we need a consistent estimator of I, defined by (19). The estimation of this long-run variance I is more complicated. In the literature, two types of estimators are generally employed: the nonparametric kernel estimator, also called Heteroskedasticity and Autocorrelation Consistent (HAC) estimators (see [2] and [34] for general references, and [23] for an application to testing strong linearity in weak ARMA models) and spectral density estimators (see e.g. [6] and [15] for a general references and [9] for estimating I when θ is not necessarily equal to θ_0).

In the present paper, we focus on an estimator based on a spectral density form for I.

Interpreting $(2\pi)^{-1}I$ as the spectral density of the stationary process (Υ_t) evaluated at frequency 0 (see [11], p. 459) of the process (20). This approach, which has been studied by [6] (see also [15]), rests on the expression

$$I = \mathbf{\Phi}^{-1}(1)\Sigma_u\mathbf{\Phi}^{-1}(1) \tag{21}$$

when (Υ_t) satisfies an AR(∞) representation of the form

$$\mathbf{\Phi}(L)\Upsilon_t := \Upsilon_t + \sum_{i=1}^{\infty}\Phi_i\Upsilon_{t-i} = u_t, \tag{22}$$

where u_t is a $(p+q)K$-variate weak white noise with variance matrix Σ_u. Let $\hat{\Upsilon}_t$ be the vector obtained by replacing θ_0 by $\hat{\theta}_n$ in Υ_t and $\hat{\mathbf{\Phi}}_r(z) = \mathrm{I}_{(p+q)K} + \sum_{i=1}^{r}\hat{\Phi}_{r,i}z^i$, where $\hat{\Phi}_{r,1},\dots,\hat{\Phi}_{r,r}$ denote the coefficients of the least squares regression of $\hat{\Upsilon}_t$ on $\hat{\Upsilon}_{t-1},\dots,\hat{\Upsilon}_{t-r}$. Let $\hat{u}_{r,t}$ be the residuals of this regression, and let $\hat{\Sigma}_{\hat{u}_r}$ be the empirical variance of $\hat{u}_{r,1},\dots,\hat{u}_{r,n}$.

In the framework of linear processes with independent innovations, Berk [6] showed that the spectral density can be consistently estimated by fitting autoregressive models of order $r = r(n)$, whenever $r \to \infty$ and $r^3/n \to 0$ as $n \to \infty$. It can be shown that this result remains valid for the linear process (Υ_t), though its innovation (u_t) is not an independent process. Another difference with [6], is that (Υ_t) is not directly observed and is replaced by $(\hat{\Upsilon}_t)$.

We are now able to state the following theorem.

Theorem 3.10. *In addition to the assumptions of Theorem 3.8, assume that the process (Υ_t) defined in (20) admits an AR(∞) representation (22) in which the roots of $\det\mathbf{\Phi}(z) = 0$ are outside the unit disk, $\|\Phi_i\| = o(i^{-2})$, and $\Sigma_u = Var(u_t)$ is non-singular. Moreover we assume that $\mathbb{E}|\epsilon_t|^{8+4\nu} < \infty$ and $\sum_{k=0}^{\infty}\{\alpha_\epsilon(k)\}^{\nu/(2+\nu)} < \infty$ and $\sum_{k=0}^{\infty}\{\alpha_\Delta(k)\}^{\nu/(2+\nu)} < \infty$ for some $\nu > 0$. Then the spectral estimator of I*

$$\hat{I}^{\mathrm{SP}} := \hat{\mathbf{\Phi}}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\mathbf{\Phi}}_r'^{-1}(1) \to I$$

in probability when $r = r(n) \to \infty$ and $r^3/n \to 0$ as $n \to \infty$.

The matrix Ω is then estimated by a "sandwich" estimator of the form

$$\hat{\Omega}^{\mathrm{SP}} = \hat{J}_n^{-1}\hat{I}^{\mathrm{SP}}\hat{J}_n^{-1}, \quad \hat{I}^{\mathrm{SP}} = \hat{\mathbf{\Phi}}_r^{-1}(1)\hat{\Sigma}_{\hat{u}_r}\hat{\mathbf{\Phi}}_r'^{-1}(1).$$

3.5. Testing linear restrictions on the parameter

It may be of interest to test s_0 linear constraints on the elements of θ_0. Let R be a given matrix of size $s_0 \times (p+q)K$ and rank s_0, and let r_0 and r_1 be given vectors of size s_0 such that $r_1 \neq r_0$. Consider the testing problem

$$H_0 : R\theta_0 = r_0 \qquad \text{against} \qquad H_1 : R\theta_0 = r_1. \tag{23}$$

The Wald principle is employed frequently for testing (23). We now examine if this principle remains valid in the non standard framework of weak RCARMA models.

Let $\hat{\Omega} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}$, where $\hat{J}$ and $\hat{I}$ are consistent estimator of J and I, as defined in Section 3.4. Under the assumptions of Theorems 3.8 and 3.10, and the assumption that I is invertible, the modified Wald statistic

$$\mathbf{W}_M := n(R_0\hat{\theta}_n - r_0)'(R_0\hat{\Omega}R_0')^{-1}(R_0\hat{\theta}_n - r_0)$$

asymptotically follows a $\chi^2_{s_0}$ distribution under H_0. Therefore, the standard formulation of the Wald test remains valid. More precisely, at the asymptotic level α, the modified Wald test consists in rejecting H_0 when $\mathbf{W}_M > \chi^2_{s_0}(1-\alpha)$. It is however important to note that a consistent estimator of the form $\hat{\Omega} = \hat{J}^{-1}\hat{I}\hat{J}^{-1}$ is required. The estimator $\hat{\Omega}_S := 2\hat{J}^{-1}$, which is routinely used in the time series softwares, is only valid in the strong RCARMA case. Thus standard Wald statistic takes the following form

$$\mathbf{W}_S := n(R_0\hat{\theta}_n - r_0)'(R_0\hat{\Omega}_S R_0')^{-1}(R_0\hat{\theta}_n - r_0),$$

which asymptotically follows a $\chi^2_{s_0}$ distribution under H_0.

4. Examples

In this section, we give examples of weak RCAR(1) model with iid and correlated process $(\Delta_t)_{t\in\mathbb{Z}}$.

4.1. Independent and identically distributed process $(\boldsymbol{\Delta}_t)_{t\in\mathbb{Z}}$: the $AR(1)$ model

We provide here some results that show that one obtains very neat results in the particular case where $(\Delta_t)_{t\in\mathbb{Z}}$ is i.i.d. and centered. We consider a particular $AR(1)$ model where (1) reads

$$X_t - a^0\Delta_t X_{t-1} = \epsilon_t, \tag{24}$$

i.e. $a^0(s) = a^0 s$ for all $s \in \mathcal{S}$, where $a^0 = \theta_0$ is here the unknown (scalar) parameter and belongs to some compact set $\Theta \subset \mathbb{R}$, and the state space $\mathcal{S}$ is a finite subset of $\mathbb{R}$. It is easy to check that, with notation defined in Section 3.1, one has that $B(s,\theta)$ is not defined (as here $q=0$), $A(s) = g_1^a(s,\theta_0) = a^0 s$ and $\Psi(s) = A(s) = a^0 s$. Stationarity condition (8) in Proposition 3.1 is translated as

$$|a^0| \left[\mathbb{E}(|\Delta_0|^8)\right]^{1/8} < 1 \quad \Longleftrightarrow \quad a^0 \in \left(-\frac{1}{\left[\mathbb{E}(|\Delta_0|^8)\right]^{1/8}}, \frac{1}{\left[\mathbb{E}(|\Delta_0|^8)\right]^{1/8}} \right). \tag{25}$$

Let us note that (25) allows some interesting cases where one has $|a^0\Delta_t| \geq 1$, which is a non stable state case and is somewhat a paradox to the usual stability condition in the classical $AR(1)$ model where it is standard that process $(X_t)_{t\in\mathbb{Z}}$ defined by $X_t = aX_{t-1} + \epsilon_t$ is stable iff $|a| < 1$. One simple example is when $(\Delta_t)_{t\in\mathbb{Z}}$ is i.i.d. with distribution $\Delta_t \sim \frac{1}{4}\delta_{-1} + \frac{1}{2} + \frac{1}{4}\delta_{+1}$, in which case (25) reads $|a^0| < 2$, so that $|a^0\Delta_t| = \frac{3}{2} > 1$ if we pick for example $a^0 = \frac{3}{2}$, when $\Delta_t = 1$.
Furthermore, one computes easily that, for all $a = \theta \in \Theta$, $\epsilon_t(a) = X_t - a\Delta_t X_{t-1}$, where X_t has the classical decomposition obtained from (24):

$$X_t = \sum_{i=0}^{\infty}\prod_{j=0}^{i-1}(a^0\Delta_{t-j})\epsilon_{t-i}. \tag{26}$$

Since Assumption **(A2)** is trivially satisfied here, we only need suppose that **(A1)**, **(A3)** and **(A4)** hold for some $\nu > 0$. In that case, Theorems 3.7 and 3.8 translate as

Theorem 4.1. *$\hat{\theta}_n$ defined as (7) converges a.s. towards $\theta_0 = a^0$. Besides, one has the asymptotic normality*

$$\sqrt{n}\left(\hat{\theta}_n - a^0\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega), \quad n \to +\infty, \tag{27}$$

where

$$\Omega = \frac{\left[1 - (a^0)^2\mathbb{E}(\Delta_0^2)\right]^2}{\mathbb{E}(\Delta_0^2)}\sum_{i=0}^{\infty}\left[(a^0)^2\mathbb{E}(\Delta_0^2)\right]^i\mathbb{E}(\epsilon_t^2\epsilon_{t-i}^2). \tag{28}$$

Proof. Strong consistency and asymptotic normality are straigtforward consequences of Theorems 3.7 and 3.8. In order to compute Ω, we need to compute $J = J(a^0)$ and $I = I(a^0)$ in (17). Since $\frac{\partial}{\partial a}\epsilon_t(a) = -\Delta_t X_{t-1}$, and since $\mathbb{E}(X_t^2)$ is equal to $\frac{1}{1-(a^0)^2\mathbb{E}(\Delta_0^2)}$ thanks to (26) and the fact that $(\epsilon_t)_{t\in\mathbb{Z}}$ is a weak noise, independent from $(\Delta_t)_{t\in\mathbb{Z}}$. Hence we have, by independence of Δ_t from X_{t-1},

$$J(a^0) = \mathbb{E}\left(\left[\frac{\partial}{\partial a}\epsilon_t(a^0)\right]^2\right) = \mathbb{E}\left(\Delta_t^2 X_{t-1}^2\right) = \frac{\mathbb{E}(\Delta_0^2)}{1 - (a^0)^2\mathbb{E}(\Delta_0^2)}.$$

There then remains to get $I = I(a^0)$. From Theorem 3.8 we need to compute the expectation of

$$\begin{aligned}
&\epsilon_t(a^0)\epsilon_{t-k}(a^0)\frac{\partial\epsilon_t(a^0)}{\partial a}\frac{\partial\epsilon_{t-k}(a^0)}{\partial a} = \epsilon_t\epsilon_{t-k}\Delta_t X_{t-1}\Delta_{t-k}X_{t-k-1}\\
= \quad &\epsilon_t\epsilon_{t-k}\Delta_t\left[\sum_{i=0}^{\infty}\prod_{j=0}^{i-1}(a^0\Delta_{t-1-j})\epsilon_{t-1-i}\right]\Delta_{t-k}\left[\sum_{i'=0}^{\infty}\prod_{j'=0}^{i'-1}(a^0\Delta_{t-k-1-j'})\epsilon_{t-k-1-i'}\right]
\end{aligned}$$

for all $k \in \mathbb{N}$. Using independence of processes $(\epsilon_t)_{t\in\mathbb{Z}}$ and $(\Delta_t)_{t\in\mathbb{Z}}$, we have

$$\mathbb{E}\left(\epsilon_t(a^0)\epsilon_{t-k}(a^0)\frac{\partial\epsilon_t(a^0)}{\partial a}\frac{\partial\epsilon_{t-k}(a^0)}{\partial a}\right) = \sum_{i,i'=0}^{\infty}V^{i,i',k}d(k, 1+i, k+1+i') \tag{29}$$

where $d(n, m, r) := \mathbb{E}(\epsilon_0\epsilon_{-n}\epsilon_{-m}\epsilon_{-r})$ for all n, m, r in $\mathbb{N}$, and

$$V^{i,i',k} := (a^0)^{i+i'+2}\mathbb{E}\left(\prod_{j=-1}^{i-1}\Delta_{t-1-j}\cdot\prod_{j'=-1}^{i'-1}\Delta_{t-k-1-j'}\right).$$

Since Δ_t is centered, one checks immediately that $V^{i,i',k}$ is non zero if and and only if $k=0$ and $i=i'$, in which case one has $V^{i,i,0}=\left[(a^0)^2\mathbb{E}(\Delta_0^2)\right]^{i+1}$. Hence (29) is in that case equal to $(a^0)^2\mathbb{E}(\Delta_0^2)\sum_{i=0}^\infty\left[(a^0)^2\mathbb{E}(\Delta_0^2)\right]^i\mathbb{E}(\epsilon_t^2\epsilon_{t-1-i}^2)$, which is also the expression for $I(a^0)$, yielding (28). □

4.2. Modulating Markov chain

We now give an example of process $(\Delta_t)_{t\in\mathbb{Z}}$ with correlated trajectories by considering a discrete time stationary ergodic Markov chain with state space $\mathcal{S}=\{1,2\}$ and transition probabilities matrix

$$P=(p(i,j))_{i,j=1,2}=\left(\begin{array}{cc}0&1\\ p&1-p\end{array}\right),$$

where p lies in $(0,1)$, and with stationary distribution

$$(\mathbb{P}(\Delta_t=1),\ \mathbb{P}(\Delta_t=2))=(\pi_1,\pi_2)=\left(\frac{p}{p+1},\frac{1}{p+1}\right). \tag{30}$$

We also consider, as in the previous section, an $AR(1)$ model of the form

$$X_t-a^0(\Delta_t)X_{t-1}=\epsilon_t, \tag{31}$$

where parameter $\theta_0=(a^0(1),a^0(2))$ verifies $a^0(1)=0$, in order to have nice expressions later for asymptotic normality. In order to establish the stationarity condition (8) we need to compute $\mathbb{E}\left[||\prod_{k=1}^t a^0(\Delta_k)||^8\right]$ which, because of $a^0(1)=0$, simplifies to

$$\mathbb{E}\left[||\prod_{k=1}^t a^0(\Delta_k)||^8\right]=|a^0(2)|^{8t}\mathbb{P}(\Delta_1=...=\Delta_t=2)=|a^0(2)|^{8t}\pi_2(1-p)^{t-1},$$

so that stationarity condition (8) here reads

$$a^0(2)\in\left(-\frac{1}{(1-p)^{1/8}},\frac{1}{(1-p)^{1/8}}\right). \tag{32}$$

Here again, as in the i.i.d. case for $(\Delta_t)_{t\in\mathbb{Z}}$, and since $\frac{1}{(1-p)^{1/8}}>1$, one can allow $|a^0(2)|$ to be larger than 1 so that state $2\in\mathcal{S}$ is non stable, although the process is stationary. Let us furthermore note that the Markov chain $(\Delta_t)_{t\in\mathbb{Z}}$ verifies the Doeblin condition so is geometrically ergodic, hence has exponentially fast strong mixing property (see [32]), so that **(A2)** is satisfied. We furthermore suppose that **(A1)**, **(A3)** and **(A4)** hold for some $\nu>0$. As in (26), one has

$$X_t=\sum_{i=0}^\infty\prod_{j=0}^{i-1}a^0(\Delta_{t-j})\epsilon_{t-i}, \tag{33}$$

and $\epsilon_t(a)=X_t-a(\Delta_t)X_{t-1}$ for all $\theta=(a(1),a(2))\in\Theta$. We introduce matrices $Q(l)$, $l\in\mathcal{S}=\{1,2\}$ as well as vector π_V defined by

$$Q(1)=\left(\begin{array}{cc}0&p\\ 0&0\end{array}\right),\quad Q(2)=\left(\begin{array}{cc}0&0\\ 1&1-p\end{array}\right),\quad \pi_V=(0,\pi_2)'. \tag{34}$$

Theorems 3.7 and 3.8 read

Theorem 4.2. *$\hat{\theta}_n$ defined as (7) converges a.s. towards $\theta_0 = (a^0(1), a^0(2))$. Besides, one has the asymptotic normality*

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Omega\right), \quad n \to +\infty, \tag{35}$$

where $\Omega = J^{-1} I J^{-1}$, matrices $J = (J(l,l'))_{l,l' \in \mathcal{S}^2}$ and $I = (I(l,l'))_{l,l' \in \mathcal{S}^2}$ being defined by

$$\begin{array}{rcl} J(1,1) & = & \sigma^2 \frac{p}{p+1} \frac{1+a^0(2)^2 p}{1-a^0(2)^2(1-p)}, \\ J(1,2) & = & J(2,1) = 0, \\ J(2,2) & = & \sigma^2 \frac{1}{p+1} \frac{1}{1-a^0(2)^2(1-p)} \end{array} \tag{36}$$

and $I(l,l') = I(l,l',0) + 2\sum_{k=1}^{\infty} I(l,l',k)$, where

$$I(l,l',k) = \left\{ \begin{array}{ll} \sum_{i'=0}^{\infty} \mathbb{1}' \left[\sum_{i<k} a^0(2)^{i+i'} Q(l) Q(2)^i P'^{k-i-1} Q(l') Q(2)^{i'} d(k, i+1, k+i'+1) \right. & \\ \quad \sum_{k \le i \le k+i'} a^0(2)^{i+i'} Q(l) Q(2)^{k+i'} d(k, i+1, k+i'+1) & \\ \quad \left. + \sum_{i>k+i'} a^0(2)^{i+i'} Q(l) Q(2)^i d(k, i+1, k+i'+1) \right] \pi_V, & l' = 2, \\ \sum_{i'=0}^{\infty} \mathbb{1}' \left[\sum_{i<k} a^0(2)^{i+i'} Q(l) Q(2)^i P'^{k-i-1} Q(l') Q(2)^{i'} d(k, i+1, k+i'+1) \right] \pi_V, & l' = 1, \end{array} \right. \tag{37}$$

where $d(i,i',i'') := \mathbb{E}(\epsilon_t \epsilon_{t-i} \epsilon_{t-i'} \epsilon_{t-i''})$, i, i', i'' in $\mathbb{N}$.

Proof. It is not hard to check that, for all $i \in \mathcal{S} = \{1,2\}$ and $a = (a(1), a(2))$, $\frac{\partial}{\partial a(i)} \epsilon_t(a) = -\mathbb{1}_{[\Delta_t = i]} X_{t-1}$. One computes easily

$$\nabla \epsilon_t(\theta_0) [\nabla \epsilon_t(\theta_0)]' = \begin{pmatrix} \mathbb{1}_{[\Delta_t=1]} X_{t-1}^2 & 0 \\ 0 & \mathbb{1}_{[\Delta_t=2]} X_{t-1}^2 \end{pmatrix},$$

so that it suffices to compute $\mathbb{E}(\mathbb{1}_{[\Delta_t=l]} X_{t-1}^2)$ for all $l = 1,2$, in order to compute J. By the usual argument of independence of the Markov chain from the weak white noise, and since $a^0(1) = 0$, we get, for $l = 1,2$,

$$\begin{array}{rcl} \mathbb{E}(\mathbb{1}_{[\Delta_t=l]} X_{t-1}^2) & = & \sigma^2 \sum_{i=0}^{\infty} \mathbb{E}\left(\mathbb{1}_{[\Delta_t=l]} \prod_{j=0}^{i-1} (a^0(\Delta_{t-1-j}))^2 \right) \\ & = & \sigma^2 \pi_l + \sigma^2 \sum_{i=1}^{\infty} a^0(2)^{2i} \pi_2 (1-p)^{i-1} p(2,l) = \sigma^2 \pi_l + \sigma^2 \dfrac{a^0(2)^2 \pi_2}{1 - a^0(2)^2(1-p)} p(2,l), \end{array}$$

so that those quantities along with (30) yield the expression for the for matrix J in (36).
In order to compute I, we need to take the expectation of $\epsilon_t(\theta_0)\epsilon_{t-k}(\theta_0) \frac{\partial}{\partial a(l)} \epsilon_t(\theta_0) \frac{\partial}{\partial a(l')} \epsilon_{t-k}(\theta_0) = \epsilon_t \epsilon_{t-k} \mathbb{1}_{[\Delta_t=l]} X_{t-1} \mathbb{1}_{[\Delta_{t-k}=l']} X_{t-1-k}$ for all l, l' in $\mathcal{S}$ and $k \in \mathbb{N}$. As in (29) in the proof of Theorem 4.1, this expectation is equal to $\sum_{i,i'=0}^{\infty} V^{i,i',k}(l,l') d(k, 1+i, k+1+i')$ where

$$V^{i,i',k}(l,l') := \mathbb{E}\left(\mathbb{1}_{[\Delta_t=l]} \prod_{j=0}^{i-1} a^0(\Delta_{t-1-j}) . \mathbb{1}_{[\Delta_{t-k}=l']} \prod_{j'=0}^{i'-1} a^0(\Delta_{t-k-1-j'}) \right).$$

This quantity can be obtained straightforwardly using e.g. Lemma 1 of [16]. Remembering that $Q(1)$, $Q(2)$ and π_V are defined by (34), we then have the following expression for $V^{i,i',k}(l,l')$, according to whether $t-i>t-k \iff i<k$, $t-k\geq t-i\geq t-k-i' \iff k\leq i\leq k+i'$ or $t-k-i'\geq t-i \iff k+i'<i$:

$$V^{i,i',k}(l,l')=\begin{cases} a^0(2)^{i+i'}\mathbb{1}'Q(l)Q(2)^iP'^{k-i-1}Q(l')Q(2)^{i'}\pi_V, & i<k,\\ a^0(2)^{i+i'}\mathbb{1}'Q(l)Q(2)^{k+i'}\pi_V, & k\leq i\leq k+i',\ l'=2,\\ 0, & k\leq i\leq k+i',\ l'=1,\\ a^0(2)^{i+i'}\mathbb{1}'Q(l)Q(2)^{i}\pi_V, & k+i'<i,\ l'=2,\\ 0, & k+i'<i,\ l'=1,\end{cases}$$

yielding (37). □

5. Numerical illustrations

We study numerically the behaviour of our estimator for strong and weak RCARMA models. We consider the following RCARMA$(1,1)$ model

$$X_t=a_1^0(\Delta_t)X_{t-1}+\epsilon_t+b_1^0(\Delta_t)\epsilon_{t-1}, \tag{38}$$

where the innovation process (ϵ_t) follows a strong or weak white noise. The process (Δ_t) is simulated (independently of (ϵ_t)) according to the law of a stationary Markov chain with state-space $\mathcal{S}=\{1,2\}$ and transition probabilities matrix

$$\begin{pmatrix} p(1,1) & 1-p(1,1)\\ 1-p(2,2) & p(2,2)\end{pmatrix}=\begin{pmatrix} 0.95 & 0.05\\ 0.05 & 0.95\end{pmatrix}.$$

This Markov chain is geometrically ergodic, so that Condition **(A2)** is satisfied. We first consider the strong RCARMA case. To generate this model, we assume the innovation process (ϵ_t) in (38) is defined by an iid sequence such that

$$\epsilon_t \stackrel{\mathcal{D}}{=} \mathcal{N}(0,1). \tag{39}$$

Following [36], we propose a set of experiment for weak RCARMA with innovation process ϵ_t in (38) defined by

$$\epsilon_t=\eta_t(|\eta_{t-1}|+1)^{-1}, \tag{40}$$

where $(\eta_t)_{t\geq 1}$ is a sequence of iid standard Gaussian random variable.

The numerical illustrations of this section are made with the free statistical software R (see http://cran.r-project.org/). We simulated $N=1,000$ independent trajectories of size $n=2,000$ of Model (38), first with the strong Gaussian noise (39), second with the weak noise (40). Recall that the regimes (Δ_t) are supposed to be known. For each of these N replications, we estimate the coefficient $\theta_0=(a_1^0(1),a_1^0(2),b_1^0(1),b_1^0(2))'=(0.90,-0.45,0.10,0.85)'$ and

Figures 1 and 2 display the realization of length 400 of Model (38) in the strong (39) and weak (40) noise cases. Note that here stationarity condition (8) in Proposition 3.1 is trivially satisfied as all coefficients $a_1^0(1)$, $a_1^0(2)$, $b_1^0(1)$, $b_1^0(2)$ are all less than 1 in modulus.

Figure 3 compares the distribution of the least squares estimators (LSE) in the strong and weak noise cases. The distributions of $\hat{a}_1^0(1)$, $\hat{a}_1^0(2)$ and $\hat{b}_1^0(2)$ are similar in the two cases, whereas the LSE of $\hat{b}_1^0(1)$ is more accurate in the weak case than in the strong one. Similar simulation experiments, not reported here, reveal that the situation is opposite, that is the

LSE is more accurate in the strong case than in the weak case, when the weak noise is defined by $\epsilon_t = \eta_t\eta_{t-1}$. This is in accordance with the results of [36] who showed that, with similar noises, the asymptotic variance of the sample autocorrelations can be greater or less than 1 as well (1 is the asymptotic variance for strong white noises).

Figure 4 compares the standard estimator $\hat{\Omega} = 2\hat{J}^{-1}$ and the sandwich estimator $\hat{\Omega} = \hat{J}^{-1}\hat{I}^{\mathrm{SP}}\hat{J}^{-1}$ of the LSE asymptotic variance Ω. We used the spectral estimator $\hat{I} := \hat{I}^{\mathrm{SP}}$ defined in Theorem 3.10, and the AR order r is automatically selected by AIC, using the function `VARselect()` of the **vars** R package. In the strong ARMA case we know that the two estimators are consistent. In view of the two top panels of Figure 4, it seems that the sandwich estimator is less accurate in the strong case. This is not surprising because the sandwich estimator is more robust, in the sense that this estimator continues to be consistent in the weak ARMA case, contrary to the standard estimator. It is clear that in the weak case $n\mathrm{Var}\left\{\hat{b}_1^0(1) - b_1^0(1)\right\}^2$ is better estimated by $\hat{\Omega}^{\mathrm{SP}}(3,3)$ (see the box-plot (c) of the right-bottom panel of Figure 4) than by $2\hat{J}^{-1}(3,3)$ (box-plot (c) of the left-bottom panel). The failure of the standard estimator of Ω in the weak RCARMA framework may have important consequences in terms of identification or hypothesis testing and validation.

Table 1 displays the relative percentages of rejection of the standard and modified Wald tests ($\mathbf{W}_S$ and $\mathbf{W}_M$) proposed in Section 3.5 for testing the null hypothesis $H_0 : b_1^0(1) = 0$. We simulated $N = 1,000$ independent trajectories of size $n = 500$, $n = 2,000$ and $n = 10,000$ of the strong RCARMA$(1,1)$ model (38)–(39) and of the weak RCARMA$(1,1)$ model (38)–(40). The nominal asymptotic level of the tests is $\alpha = 5\%$ and the empirical size over the $N = 1,000$ independent replications should vary between the significant limits 3.6% and 6.4% with probability 95%. The line in bold corresponds to the null hypothesis H_0. For the strong RCARMA model (38)–(39), the relative rejection frequencies of the $\mathbf{W}_S$ and $\mathbf{W}_M$ tests are close to the nominal 5% level when $b_1^0(1) = 0$, and are close to 100% under the alternative when n is large. In this strong RCARMA example, the $\mathbf{W}_S$ and $\mathbf{W}_M$ tests have very similar powers under the alternative for all sizes. As expected, for the weak RCARMA model (38)–(40), the relative rejection frequencies of the standard $\mathbf{W}_S$ Wald test is definitely outside the significant limits. Thus the error of first kind is well controlled by all the tests in the strong case, but only by the $\mathbf{W}_M$ modified version test in the weak case (Model (38)–(40)) when $b_1^0(1) = 0$. Note also that for Model (38)–(40), the relative rejection frequencies of the $\mathbf{W}_M$ test tend rapidly to 100% as n increases under the alternative. By contrast the empirical powers of the standard $\mathbf{W}_S$ test is hardly interpretable for Model (38)–(40), because we have already seen in Table 1 that the standard version of the $\mathbf{W}_S$ test do not well control the error of first kind in the weak RCARMA framework.

From these simulation experiments and from the asymptotic theory, we draw the conclusion that the standard methodology, based on the LSE, allows to fit RCARMA representations of a wide class of nonlinear time series. This standard methodology, including in particular the significance tests on the parameters, needs however to be adapted to take into account the possible lack of independence of the errors terms. In future works, we intent to study how the existing identification and diagnostic checking procedures should be adapted in the weak RCARMA framework considered in the present paper.

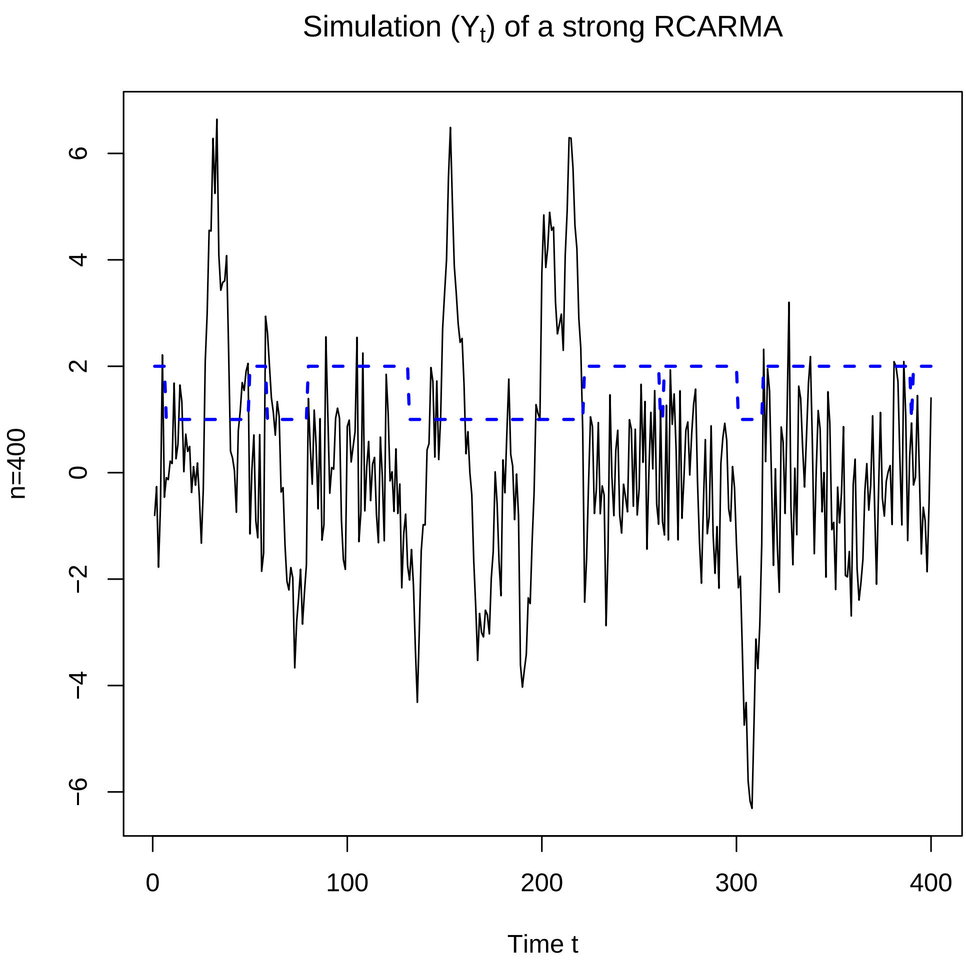

FIG 1. *Simulation of length 400 of Model (38)–(39) with* $\theta_0 = (a_1^0(1), a_1^0(2), b_1^0(1), b_1^0(2))' = (0.90, -0.45, 0.10, 0.85)'$, . *The process* (X_t) *is drawn in full line, the Markov chain* (Δ_t) *is plotted in dotted line.*

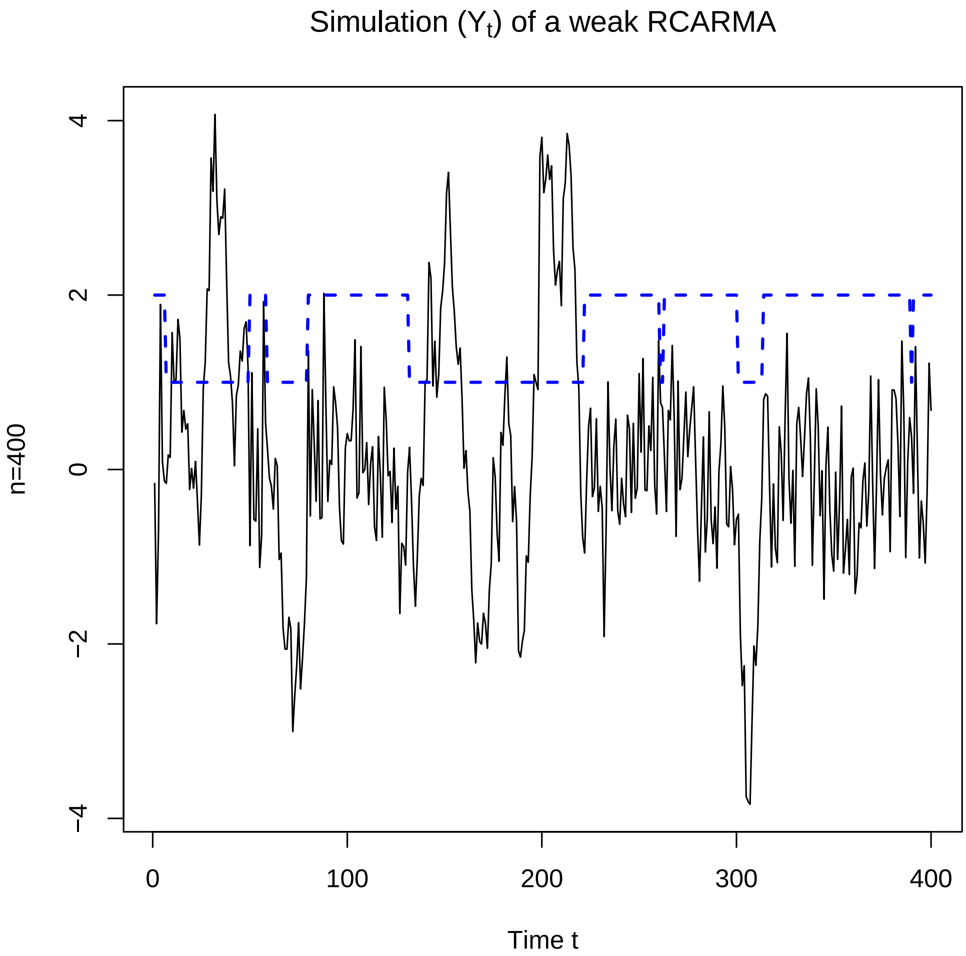

FIG 2. *Simulation of length 400 of Model (38)–(40) with* $\theta_0 = (a_1^0(1), a_1^0(2), b_1^0(1), b_1^0(2))' = (0.90, -0.45, 0.10, 0.85)'$, . *The process* (X_t) *is drawn in full line, the Markov chain* (Δ_t) *is plotted in dotted line.*

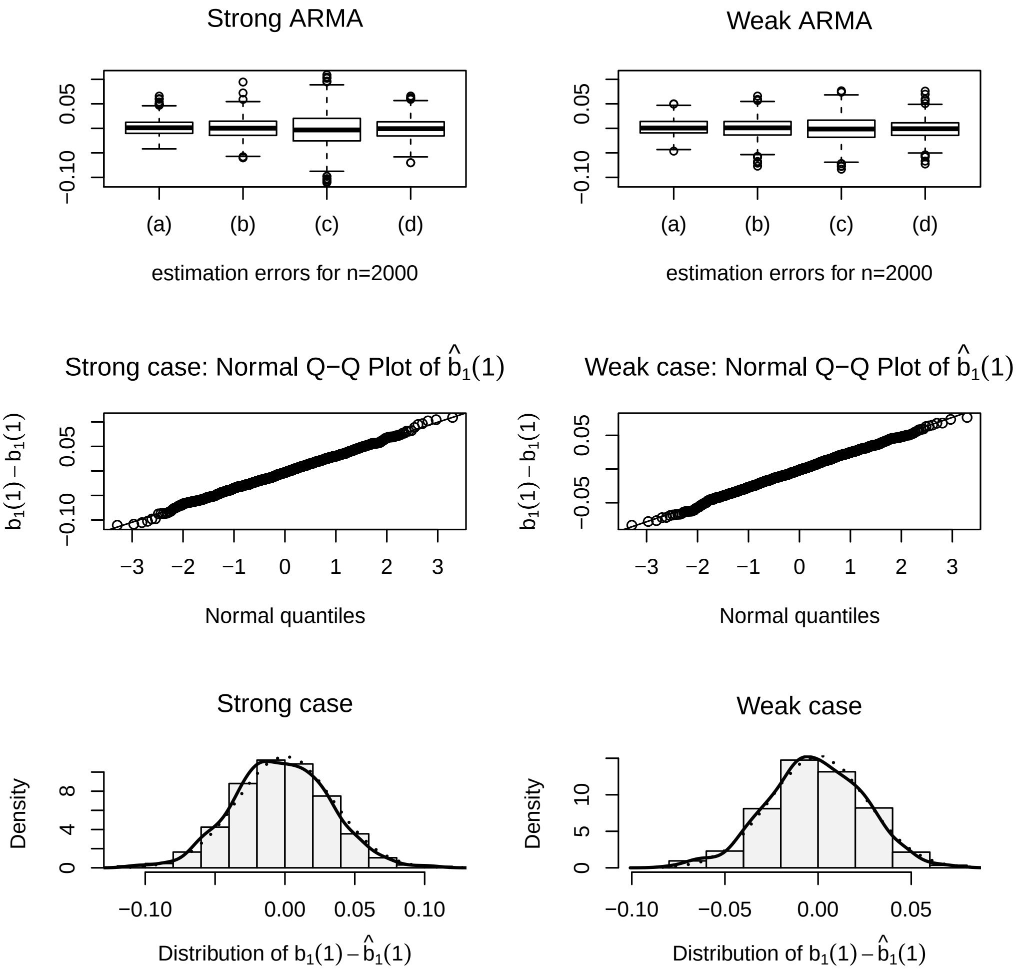

FIG 3. *LSE of $N = 1000$ independent simulations of the model (38) with size $n = 2000$ and unknown parameter $\theta_0 = (a_1^0(1), a_1^0(2), b_1^0(1), b_1^0(2))' = (0.90, -0.45, 0.10, 0.85)'$, when the noise is strong (39) (left panels) and when the noise is the weak noise (40) (right panels). Points (a)-(d), in the box-plots of the top panels, display the distribution of the estimation errors $\hat{\theta}(i) - \theta_0(i)$ for $i = 1, \ldots, 4$. The panels of the middle present the Q-Q plot of the estimates $\hat{\theta}(3) = \hat{b}_1^0(1)$ of the last parameter. The bottom panels display the distribution of the same estimates. The kernel density estimate is displayed in full line, and the centered Gaussian density with the same variance is plotted in dotted line.*

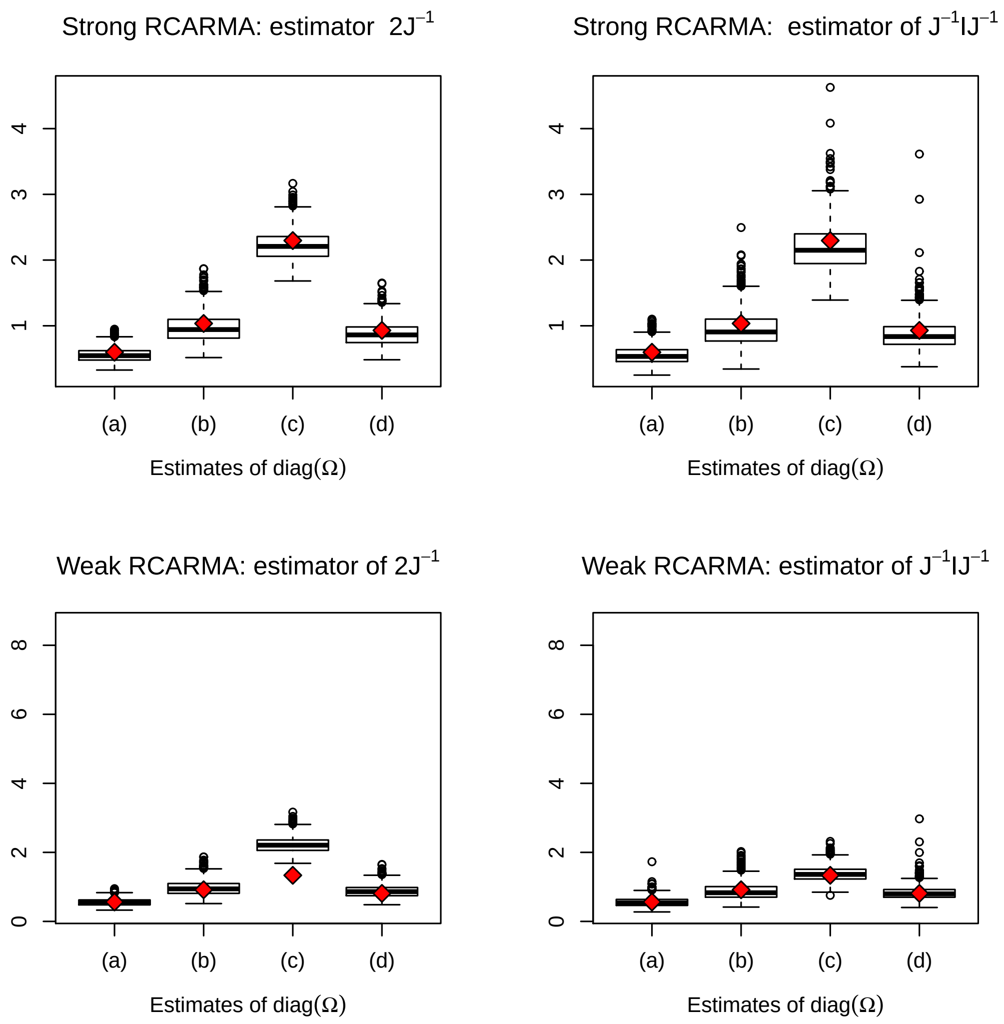

FIG 4. *Comparison of standard and modified estimates of the asymptotic variance* Ω *of the LSE, on the simulated models presented in Figure 3. The diamond symbols represent the mean, over the* $N = 1000$ *replications, of the standardized squared errors* $n\left\{\hat{a}_1^0(1) - 0.90\right\}^2$ *for (a) (0.58 in the strong case and 0.56 in the weak case),* $n\left\{\hat{a}_1^0(2) + 0.45\right\}^2$ *for (b) (0.94 in the strong case and 0.92 in the weak case),* $n\left\{\hat{b}_1^0(1) - 0.10\right\}^2$ *for (c) (2.33 in the strong case and 1.33 in the weak case) and* $n\left\{\hat{b}_1^0(2) - 0.85\right\}^2$ *for (d) (0.87 in the strong case and 0.81 in the weak case).*

Table 1

Percentages of rejection of standard $\mathbf{W}_S$ and modified $\mathbf{W}_M$ Wald tests for testing the null hypothesis $H_0 : b_1^0(1) = 0$, in the RCARMA$(1,1)$ model (38). The nominal asymptotic level of the tests is $\alpha = 5\%$. The number of replications is $N = 1000$. The line in bold corresponds to the null hypothesis H_0.

	$n = 500$		$n = 2000$		$n = 10000$	
$b_1^0(1)$	$\mathbf{W}_S$	$\mathbf{W}_M$	$\mathbf{W}_S$	$\mathbf{W}_M$	$\mathbf{W}_S$	$\mathbf{W}_M$
	Strong RCARMA-Model (38)–(39)					
0.9	100.0	100.0	100.0	100.0	100.0	100.0
0.4	100.0	100.0	100.0	100.0	100.0	100.0
0.2	86.1	86.9	100.0	100.0	100.0	100.0
0.1	32.3	35.3	85.7	86.6	100.0	100.0
0.0	**6.3**	**8.0**	**6.0**	**6.3**	**5.6**	**5.8**
-0.1	27.3	30.1	79.8	79.8	100.0	100.0
-0.2	76.5	77.0	100.0	100.0	100.0	100.0
-0.4	99.2	99.0	100.0	100.0	100.0	100.0
-0.9	84.7	85.9	99.6	99.6	100.0	100.0
	Weak RCARMA-Model (38)–(40)					
0.9	100.0	100.0	100.0	100.0	100.0	100.0
0.4	99.7	100.0	100.0	100.0	100.0	100.0
0.2	57.4	96.2	100.0	100.0	100.0	100.0
0.1	3.5	52.4	50.3	98.0	100.0	100.0
0.0	**0.2**	**5.8**	**0.0**	**4.7**	**0.0**	**5.6**
-0.1	2.8	39.5	37.6	93.8	100.0	100.0
-0.2	34.1	89.6	99.9	100.0	100.0	100.0
-0.4	96.0	99.6	100.0	100.0	100.0	100.0
-0.9	86.1	89.7	99.7	99.7	100.0	100.0

Appendix A: Proofs

A.1. Proofs of Proposition 3.1 and Lemma 3.3

Proof of Proposition 3.1. Let us first note that Condition (8) is equivalent to

$$\mathbb{E}\left(\sup_{\theta\in\Theta}\left\|\prod_{i=1}^{t}\Phi(\Delta_i,\theta)\right\|^8\right)\le C\rho^t,\quad \mathbb{E}\left(\left\|\prod_{i=1}^{t}\Psi(\Delta_i)\right\|^8\right)\le C\rho^t, \tag{41}$$

for some constant $C > 0$ and $0 < \rho < 1$ (independent from θ), and is akin to Condition (**A2**) in [16]. Let us first introduce processes $(\tilde{Z}_t)_{t\in\mathbb{Z}}$ and $(\tilde{\omega}_t)_{t\in\mathbb{Z}}$ by

$$\tilde{Z}_t = (X_t, \dots, X_{t-p+1}, \epsilon_t, \dots, \epsilon_{t-q+1})' \in \mathbb{R}^{(p+q)\times 1}, \quad \tilde{\omega}_t = (\epsilon_t, 0, \dots, \epsilon_t, \dots, 0)' \in \mathbb{R}^{(p+q)\times 1}$$

where ϵ_t in the latter is in $(p+1)$th position in $\tilde{\omega}_t$. Then it is clear that one has the following equation for $\tilde{Z}_t$:

$$\tilde{Z}_t = \Psi(\Delta_t)\tilde{Z}_{t-1} + \tilde{\omega}_t, \quad \forall t \in \mathbb{Z},$$

of which a candidate for the solution of the above equation is, with the usual convention $\prod_{j=0}^{-1} = 1$,

$$\tilde{Z}_t = \sum_{k=0}^{\infty} \prod_{j=0}^{k-1} \Psi(\Delta_{t-j})\tilde{\omega}_{t-k}, \quad t \in \mathbb{Z}, \tag{42}$$

a stationary process, provided that the series converges, which we strive to prove now. Let us pick for $||.||$ a subordinate norm on the set of matrices. By independence of processes $(\Delta_t)_{t\in\mathbb{Z}}$ and $(\epsilon_t)_{t\in\mathbb{Z}}$, and using the fact that the latter is square integrable, we easily get, for $k \geq 1$,

$$\begin{aligned}
\mathbb{E}\left(||\Psi(\Delta_t)\dots\Psi(\Delta_{t-k+1})\tilde{\omega}_{t-k}||^2\right) &\leq \mathbb{E}\left(||\Psi(\Delta_t)\dots\Psi(\Delta_{t-k+1})||^2 . ||\tilde{\omega}_{t-k}||^2\right)\\
&= \mathbb{E}\left(||\Psi(\Delta_t)\dots\Psi(\Delta_{t-k+1})||^2\right)\mathbb{E}\left(||\tilde{\omega}_{t-k}||^2\right) \leq C\mathbb{E}\left(||\tilde{\omega}_0||^2\right)\rho^k,
\end{aligned}$$

the last inequality stemming from (41), so that series (42) converges in L^2. Note that one proves that $\tilde{Z}_t$ (hence X_t) is in L^4 by replacing $||.||^2$ by $||.||^4$ in the above inequality, using again (41) and the fact that $(\epsilon)_{t\in\mathbb{Z}}$ is in L^4, see assumption (**A3**). Similarly, defining

$$Z_t(\theta) := (\epsilon_t(\theta), \dots, \epsilon_{t-q+1}(\theta), X_t, \dots, X_{t-p+1})', \quad \omega_t = (X_t, 0, \dots, X_t, \dots, 0)' \tag{43}$$

where X_t in the latter is in $(q+1)$th position, one also gets that $Z_t(\theta)$ satisfies

$$Z_t(\theta) = \Phi(\Delta_t, \theta)Z_{t-1}(\theta) + \omega_t.$$

A solution candidate to the above solution is

$$Z_t(\theta) = \sum_{k=0}^{\infty} \prod_{j=0}^{k-1} \Phi(\Delta_{t-j}, \theta)\omega_{t-k}, \quad t \in \mathbb{Z}. \tag{44}$$

Similarly to the proof leading to (42), convergence of (44) is obtained thanks to (41) as well as stationarity of $(X_t)_{t\in\mathbb{Z}}$ and the fact that $X_t \in L^4$.

One checks that $\omega_t = M\tilde{Z}_t$ and $\epsilon_t(\theta) = e_1 Z_t(\theta)$, which, plugged into (42) and (44) yields (9) with coefficients $c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1})$ given by (10). Finally, let us verify that $(c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1}))_{i\in\mathbb{N}}$ is the unique sequence verifying (9). Let us then pick a sequence of r.v. $(d_i)_{i\in\mathbb{N}}$ in $\mathcal{H}$ such that $\epsilon_t(\theta) = \sum_{i=0}^{\infty} c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1})\epsilon_{t-i} = \sum_{i=0}^{\infty} d_i\epsilon_{t-i}$. One then gets, by independence from $(\epsilon_t)_{t\in\mathbb{Z}}$ as well as by the fact that the latter is a weak white noise:

$$0 = \mathbb{E}\left(\left[\sum_{i=0}^{\infty}(c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1}) - d_i)\epsilon_{t-i}\right]^2\right) = \sigma^2\mathbb{E}\left(\sum_{i=0}^{\infty}(c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1}) - d_i)^2\right)$$

hence $(c_i(\theta, \Delta_t, \dots, \Delta_{t-i+1}))_{i\in\mathbb{N}} = (d_i)_{i\in\mathbb{N}}$ a.s. □

Proof of Lemma 3.3. The fact that the $\theta \mapsto c_i(\theta, \Delta_{t-1}, \dots, \Delta_{t-i})$, $\theta \mapsto \nabla[c_i(\theta, \Delta_{t-1}, \dots, \Delta_{t-i})]^2$ and $\theta \mapsto \nabla^2[c_i(\theta, \Delta_{t-1}, \dots, \Delta_{t-i})]^2$ are polynomial functions (of several variables) can be verified easily using the fact that, for all $s \in \mathcal{S}$, $\theta \mapsto \Phi(s, \theta)$ and $\theta \mapsto \Psi(\theta)$ are affine functions. We turn to (13). Using Minkovski's inequality, the fact that matrix norm $||.||$ is submulitplicative entails

$$\begin{aligned}
&\left|\left|\sup_{\theta\in\Theta}|c_i(\theta, \Delta_i, \dots, \Delta_1)|\right|\right|_{2\nu+4} \leq \sum_{k=0}^{i}\left|\left|\sup_{\theta\in\Theta}|e_1\Phi(\Delta_i, \theta)\dots\Phi(\Delta_{i-k+1}, \theta)M\Psi(\Delta_{i-k})\dots\Psi(\Delta_1)e_{p+1}'|\right|\right|_{2\nu+4}\\
&\leq C\sum_{k=0}^{i}\left[\mathbb{E}\left(\sup_{\theta\in\Theta}||\Phi(\Delta_i, \theta)\dots\Phi(\Delta_{i-k+1}, \theta)||^{2\nu+4}\,||\Psi(\Delta_{i-k})\dots\Psi(\Delta_1)||^{2\nu+4}\right)\right]^{1/(2\nu+4)} \qquad (45)
\end{aligned}$$

for some constant $C > 0$. The Cauchy Schwartz inequality as well as (8) yields

$$\left[\mathbb{E}\left(\sup_{\theta\in\Theta}||\Phi(\Delta_i,\theta)\dots\Phi(\Delta_{i-k+1},\theta)||^{2\nu+4}\,||\Psi(\Delta_{i-k})\dots\Psi(\Delta_1)||^{2\nu+4}\right)\right]^{1/(2\nu+4)}$$
$$\le \left[\mathbb{E}\left(\sup_{\theta\in\Theta}||\Phi(\Delta_i,\theta)\dots\Phi(\Delta_{i-k+1},\theta)||^{4\nu+8}\right)\right]^{\frac{1}{(4\nu+8)}}\left[\mathbb{E}\left(||\Psi(\Delta_{i-k})\dots\Psi(\Delta_1)||^{4\nu+8}\right)\right]^{\frac{1}{(4\nu+8)}} \le \kappa\rho^{\frac{i}{(2\nu+4)}}$$

which, plugged in (45), yields inequality (13) for $c_i(\theta,\Delta_i,\dots,\Delta_1)$. The inequalities for $\nabla^j[c_i(\theta,\Delta_i,\dots,\Delta_1)]$, $j=2,3$, are proved similarly. As to $c_i^e(t,\theta,\Delta_t,\dots,\Delta_{t-i+1})$, (12) yields the upper bound

$$\left|\left|\sup_{\theta\in\Theta}|c_i^e(t,\theta,\Delta_t,\dots,\Delta_{t-i+1})|\right|\right|_{2\nu+4}$$
$$\le\sum_{k=0}^{i}\left|\left|\sup_{\theta\in\Theta}|e_1\Phi(\Delta_t,\theta)\dots\Phi(\Delta_{t-k+1},\theta)M\Psi(\Delta_{t-k})\dots\Psi(\Delta_{t-i+1})e'_{p+1}|\right|\right|_{2\nu+4},$$

so that upper bound (14) for $c_i^e(t,\theta,\Delta_{t-1},\dots,\Delta_{t-i})$ follows again by a Cauchy Schwartz argument. The upper bound (14) for $\nabla c_i^e(t,\theta,\Delta_{t-1},\dots,\Delta_{t-i})$ is obtained similarly. $\square$

A.2. Proofs of Lemma 3.4 and Proposition 3.5

Proof of Lemma 3.4. We first prove Point 1. Using decomposition (9) of $\epsilon_t(\theta)$, independence of the white noise from the modulating process, as well as stationarity of the former, we obtain

$$\left|\left|\sup_{\theta\in\Theta}|\epsilon_0(\theta)|\right|\right|_4\le\sum_{i=0}^{\infty}\left|\left|\sup_{\theta\in\Theta}|c_i(\theta,\Delta_t,\dots,\Delta_{t-i+1})|\right|\right|_4.||\epsilon_0||_4$$

which is a converging series because of (13). As to $e_t(\theta)$, we use this time decomposition (11) as well as (14) in order to get

$$\sup_{t\ge0}\left|\left|\sup_{\theta\in\Theta}|e_t(\theta)|\right|\right|_4\le\sum_{i=0}^{\infty}\sup_{t\ge0}\left|\left|\sup_{\theta\in\Theta}|c_i^e(\theta,\Delta_t,\dots,\Delta_{t-i+1})|\right|\right|_4.||\epsilon_0||_4<+\infty.$$

In order to prove Point 2, we remind the following notations. From (4) and (5), we have

$$Z_t(\theta)=\omega_t+\Phi(\Delta_t,\theta)Z_{t-1}(\theta)\qquad\forall t\in\mathbb{Z},$$

and

$$Z_t^e(\theta)=\omega_t^e+\Phi(\Delta_t,\theta)Z_{t-1}^e(\theta)\qquad t=1,\dots,n,$$

where $Z_t^e(\theta):=(e_t(\theta),\dots,e_{t-q+1}(\theta),\tilde{X}_t,\dots,\tilde{X}_{t-p+1})'$, $\omega_t^e=(\tilde{X}_t,0,\dots,\tilde{X}_t,\dots,0)'$, so that $\omega_t^e=\omega_t$ for $t\ge r+1$ (where $r=\max(p,q)$), $\omega_t^e(\theta)=0_{p+q}$ for $t\le0$. We recall that processes $(\tilde{X}_t)_{t\in\mathbb{Z}}$ and $(e_t(\theta))_{t\in\mathbb{Z}}$ verify (5). Note that $||\sup_{\theta\in\Theta}|\epsilon_t(\theta)-e_t(\theta)|||_2\longrightarrow0$ is equivalent to $||\sup_{\theta\in\Theta}||Z_t^e(\theta)-Z_t(\theta)||||_2\longrightarrow0$ as $t\to\infty$. Now, since $\tilde{X}_t=X_t$ for $t\ge1$, one easily sees that

$$Z_t^e(\theta)-Z_t(\theta)=\Phi(\Delta_t,\theta)[Z_{t-1}^e(\theta)-Z_{t-1}(\theta)],\quad\forall t\ge r+1,\tag{46}$$
$$Z_t^e(\theta)-Z_t(\theta)=\omega_t^e-\omega_t+\Phi(\Delta_t,\theta)[Z_{t-1}^e(\theta)-Z_{t-1}(\theta)],\ \text{for } t=1,\dots,r.\tag{47}$$

Now, using (46) and (47) we obtain

$$
\begin{aligned}
Z_t^e(\theta) - Z_t(\theta) &= \prod_{j=0}^{t-r-1} \Phi(\Delta_{t-j},\theta)[Z_r^e(\theta) - Z_r(\theta)], \quad \forall t \geq r+1, \\
&= \prod_{j=0}^{t-r-1} \Phi(\Delta_{t-j},\theta) \left(\sum_{i=0}^{r-1} \prod_{j=0}^{i-1} \Phi(\Delta_{r-j},\theta)[\omega_{r-i}^e - \omega_{r-i}] \prod_{j=0}^{r-1} \Phi(\Delta_{r-j},\theta)\omega_0 \right). \qquad (48)
\end{aligned}
$$

Let us furthermore note that

$$
\left|\left|\sup_{\theta\in\Theta} |\tilde{X}_t - X_t|\right|\right|_4 = \left|\left|\sup_{\theta\in\Theta} |\sum_{i=t}^{r} g_i^a(\Delta_t,\theta)X_{t-i} + \sum_{j=t}^{r} g_j^b(\Delta_t,\theta)\epsilon_{t-i}(\theta)|\right|\right|_4 < +\infty \text{ for } t=1,\dots,r
$$

as indeed $X_t \in L^4$ (as proved in the proof of Proposition 3.1) and $||\sup_{\theta\in\Theta}\epsilon_t(\theta)||_4 < +\infty$ as proved in Point 1. In view of (48), using Minkowski's and Holder's inequalities and (8), we thus have

$$
\left|\left|\sup_{\theta\in\Theta} ||Z_t^e(\theta) - Z_t(\theta)||\right|\right|_2 \leq C\rho^t,
$$

for some constant $C>0$ and $0<\rho<1$ (independent from θ).
Let us turn to Point 3. This is due to

$$
\mathbb{P}\left(t^\alpha \sup_{\theta\in\Theta} |\epsilon_t(\theta) - e_t(\theta)| > \eta\right) \leq \frac{t^{2+2\alpha}\,||\sup_{\theta\in\Theta}|\epsilon_t(\theta) - e_t(\theta)|||_2^2}{t^2\eta^2} = o\left(\frac{1}{t^2}\right), \quad \forall \eta > 0,
$$

the last equality thanks to Point 2, and using Borel Cantelli's lemma.
We now turn to Point 4. The fact that $\left|\left|\sup_{\theta\in\Theta}||\nabla^j\epsilon_0(\theta)||\right|\right|_4$ and $\sup_{t\geq 0}\left|\left|\sup_{\theta\in\Theta}||\nabla^j e_t(\theta)||\right|\right|_4$ are finite is proved similarly to Point 1 and using estimates (13) and (14). We then pass on to the limit of $t^\alpha\,||\sup_{\theta\in\Theta}||\nabla(e_t-\epsilon_t)(\theta)||||_{4/3}$ as $t\to\infty$. Let $i\in\mathcal{S}$. Deriving (46) with respect to θ_i yields

$$
\frac{\partial}{\partial\theta_i}[Z_t^e(\theta) - Z_t(\theta)] = \Phi(\Delta_t,\theta)\frac{\partial}{\partial\theta_i}[Z_{t-1}^e(\theta) - Z_{t-1}(\theta)] + \frac{\partial}{\partial\theta_i}\Phi(\Delta_t,\theta)[Z_{t-1}^e(\theta) - Z_{t-1}(\theta)], \quad \forall t \geq p+1, \qquad (49)
$$

hence we may write

$$
\frac{\partial}{\partial\theta_i}[Z_t^e(\theta) - Z_t(\theta)] = \sum_{k=0}^{t-p}\prod_{j=0}^{k-1}\Phi(\Delta_{t-j},\theta)\frac{\partial}{\partial\theta_i}\Phi(\Delta_{t-k},\theta)[Z_{t-k}^e(\theta) - Z_{t-k}(\theta)],
$$

hence, using Minkovski's and Holder's inequalities, and letting $M_\Phi := \max_{s\in\mathcal{S},\theta\in\Theta}\left|\frac{\partial}{\partial\theta_i}\Phi(s,\theta)\right|$, we get

$$
\begin{aligned}
t^\alpha\left|\left|\sup_{\theta\in\Theta}||\frac{\partial}{\partial\theta_i}[Z_t^e(\theta) - Z_t(\theta)]||\right|\right|_{8/5} \leq M_\Phi \sum_{k=0}^{t-p}\left|\left|\sup_{\theta\in\Theta}|\prod_{j=0}^{k-1}\Phi(\Delta_{t-j},\theta)|\right|\right|_8 \\
.t^\alpha\left|\left|\sup_{\theta\in\Theta}||Z_{t-k}^e(\theta) - Z_{t-k}(\theta)||\right|\right|_2. \qquad (50)
\end{aligned}
$$

Now, since $\left|\left|\sup_{\theta\in\Theta}||\prod_{j=0}^{k-1}\Phi(\Delta_{t-j},\theta)||\right|\right|_8 \leq \kappa\rho^k$ for some $\kappa>0$ and $\rho<1$ thanks to (8), and since $t^\alpha\left|\left|\ \sup_{\theta\in\Theta}||Z_{t-k}^e(\theta)-Z_{t-k}(\theta)||\right|\right|_2$ is uniformly bounded in t and $k\leq t$, and tends to 0 as $t\to\infty$, the dominated convergence theorem yields that $t^\alpha\left|\left|\sup_{\theta\in\Theta}||\frac{\partial}{\partial\theta_i}[Z_t^e(\theta)-Z_t(\theta)]||\right|\right|_{8/5}\longrightarrow 0$ as $t\to\infty$, proving $t^\alpha\,||\sup_{\theta\in\Theta}||\nabla(e_t-\epsilon_t)(\theta)||||_{8/5}\longrightarrow 0$ as $t\to\infty$ in Point 4. Let us now prove that $t^\alpha\left|\left|\sup_{\theta\in\Theta}||\nabla^2(e_t-\epsilon_t)(\theta)||\right|\right|_{4/3}\longrightarrow 0$. Deriving again (49) with respect to θ_ℓ, $\ell\in\mathcal{S}$, we obtain

$$\frac{\partial^2}{\partial\theta_\ell\partial\theta_i}[Z_t^e(\theta)-Z_t(\theta)]=\Phi(\Delta_t,\theta)\frac{\partial^2}{\partial\theta_\ell\partial\theta_i}[Z_{t-1}^e(\theta)-Z_{t-1}(\theta)]+\frac{\partial}{\partial\theta_\ell}\Phi(\Delta_t,\theta)\frac{\partial}{\partial\theta_i}[Z_{t-1}^e(\theta)-Z_{t-1}(\theta)]$$
$$+\frac{\partial}{\partial\theta_i}\Phi(\Delta_t,\theta)\frac{\partial}{\partial\theta_\ell}[Z_{t-1}^e(\theta)-Z_{t-1}(\theta)]+\frac{\partial^2}{\partial\theta_\ell\partial\theta_i}\Phi(\Delta_t,\theta)[Z_{t-1}^e(\theta)-Z_{t-1}(\theta)],\quad\forall t\geq p+1,\tag{51}$$

so that, in the same spirit as (49), one obtains

$$\begin{aligned}t^\alpha\left|\left|\sup_{\theta\in\Theta}||\frac{\partial^2}{\partial\theta_\ell\partial\theta_i}[Z_t^e(\theta)-Z_t(\theta)]||\right|\right|_{4/3}&\leq M'_\Phi\sum_{k=0}^{t-p}\left|\left|\sup_{\theta\in\Theta}|\prod_{j=0}^{k-1}\Phi(\Delta_{t-j},\theta)|\right|\right|_8\\ .t^\alpha\Bigg[\left|\left|\ \sup_{\theta\in\Theta}||Z_{t-k}^e(\theta)-Z_{t-k}(\theta)||\right|\right|_{8/5}&+\left|\left|\ \sup_{\theta\in\Theta}||\frac{\partial}{\partial\theta_\ell}[Z_{t-k}^e(\theta)-Z_{t-k}(\theta)]||\right|\right|_{8/5}\\ &+\left|\left|\ \sup_{\theta\in\Theta}||\frac{\partial}{\partial\theta_i}[Z_{t-k}^e(\theta)-Z_{t-k}(\theta)]||\right|\right|_{8/5}\Bigg],\end{aligned}\tag{52}$$

for some positive constant M'_Φ. Using Point 2 (so that $t^\alpha\left|\left|\ \sup_{\theta\in\Theta}||Z_{t-k}^e(\theta)-Z_{t-k}(\theta)||\right|\right|_{8/5}$ tends to 0 as $t\to\infty$, since $8/5<2$) and the previous estimate

$$t^\alpha\left|\left|\sup_{\theta\in\Theta}||\frac{\partial}{\partial\theta_i}[Z_t^e(\theta)-Z_t(\theta)]||\right|\right|_{8/5}\longrightarrow 0$$

for all $i\in\mathcal{S}$, we conclude by a dominated convergence theorem that

$$t^\alpha\left|\left|\sup_{\theta\in\Theta}||\frac{\partial^2}{\partial\theta_\ell\partial\theta_i}[Z_t^e(\theta)-Z_t(\theta)]||\right|\right|_{4/3},\ \text{hence } t^\alpha\left|\left|\sup_{\theta\in\Theta}||\frac{\partial^2}{\partial\theta_\ell\partial\theta_i}(e_t-\epsilon_t)(\theta)||\right|\right|_{4/3},$$

tends to 0.

We finish by sketching the proof leading to $t^\alpha\left|\left|\sup_{\theta\in\Theta}||\nabla^3(e_t-\epsilon_t)(\theta)||\right|\right|_1\longrightarrow 0$. The starting point is again deriving (51) with respect to $\theta_{\ell'}$, $\ell'\in\mathcal{S}$, which yields, as in (52), the following estimate:

$$t^\alpha \left|\left|\sup_{\theta\in\Theta} || \frac{\partial^3}{\partial\theta'_\ell \partial\theta_\ell \partial\theta_i}[Z^e_t(\theta) - Z_t(\theta)]||\right|\right|_1 \le M''_\Phi \sum_{k=0}^{t-p} \left|\left|\sup_{\theta\in\Theta} | \prod_{j=0}^{k-1} \Phi(\Delta_{t-j},\theta)|\right|\right|_8$$
$$.t^\alpha \left[\left|\left| \sup_{\theta\in\Theta} ||Z^e_{t-k}(\theta) - Z_{t-k}(\theta)||\right|\right|_{4/3} + \left|\left| \sup_{\theta\in\Theta} || \frac{\partial}{\partial\theta_\ell}[Z^e_{t-k}(\theta) - Z_{t-k}(\theta)]||\right|\right|_{4/3} \right.$$
$$+ \left|\left| \sup_{\theta\in\Theta} || \frac{\partial}{\partial\theta_i}[Z^e_{t-k}(\theta) - Z_{t-k}(\theta)]||\right|\right|_{4/3} + \left|\left| \sup_{\theta\in\Theta} || \frac{\partial^2}{\partial\theta_\ell\partial\theta_i}[Z^e_{t-k}(\theta) - Z_{t-k}(\theta)]||\right|\right|_{4/3}$$
$$\left. + \left|\left| \sup_{\theta\in\Theta} || \frac{\partial^2}{\partial\theta'_\ell\partial\theta_i}[Z^e_{t-k}(\theta) - Z_{t-k}(\theta)]||\right|\right|_{4/3} + \left|\left| \sup_{\theta\in\Theta} || \frac{\partial^2}{\partial\theta'_\ell\partial\ell}[Z^e_{t-k}(\theta) - Z_{t-k}(\theta)]||\right|\right|_{4/3} \right],$$

for some constant M''_Φ, so that one concludes similarly. □

Proof of Proposition 3.5. Let us start with Point 1. The fact that $Q_n(\theta)$ converges a.s. to $O_\infty(\theta) = \mathbb{E}(\epsilon_0(\theta))$ as $n\to\infty$ is a consequence of the fact that $\sup_{\theta\in\Theta} |\epsilon_t(\theta) - e_t(\theta)|^2 \longrightarrow 0$ (itself a consequence of Point 3 of Lemma 3.4) and is justified by the same exact proof of Lemma 7 in [20]. We now prove that $n^\alpha\, ||\sup_{\theta\in\Theta}|Q_n(\theta) - O_n(\theta)|||_1$. Let $\alpha\in(0,1)$. Using the upper bound $\sup_{\theta\in\Theta}|e_t(\theta)^2 - \epsilon_t(\theta)^2| \le [\sup_{\theta\in\Theta}|e_t(\theta)| + \sup_{\theta\in\Theta}|\epsilon_t(\theta)|]\,.\sup_{\theta\in\Theta}|e_t(\theta)-\epsilon_t(\theta)|$, as well as Cauchy Schwartz and Minkovski's inequality, we get the following

$$n^\alpha \left|\left|\sup_{\theta\in\Theta}|Q_n(\theta) - O_n(\theta)|\right|\right|_1 \le \frac{1}{n^{1-\alpha}}\sum_{t=1}^n \left[\left|\left|\sup_{\theta\in\Theta}|e_t(\theta)|\right|\right|_2 + \left|\left|\sup_{\theta\in\Theta}|\epsilon_t(\theta)|\right|\right|_2 \right] . \left|\left|\sup_{\theta\in\Theta}|e_t(\theta) - \epsilon_t(\theta)|\right|\right|_2 .$$

Since $||\sup_{\theta\in\Theta}|e_t(\theta)|||_2$ is upper bounded by Point 1 of Lemma 3.4, and $||\sup_{\theta\in\Theta}|\epsilon_t(\theta)|||_2$ is constant in t and finite, there thus exists some constant $C>0$ such that

$$n^\alpha \left|\left|\sup_{\theta\in\Theta}|Q_n(\theta) - O_n(\theta)|\right|\right|_1 \le C\frac{1}{n^{1-\alpha}}\sum_{t=1}^n \left|\left|\sup_{\theta\in\Theta}|e_t(\theta) - \epsilon_t(\theta)|\right|\right|_2 . \tag{53}$$

Let us write the right hand side of the above inequality in the form $\frac{1}{n^{1-\alpha}}\sum_{t=1}^n [t^{1-\alpha} - (t-1)^{1-\alpha}]\frac{1}{t^{1-\alpha}-(t-1)^{1-\alpha}}\,||\sup_{\theta\in\Theta}|e_t(\theta) - \epsilon_t(\theta)|||_2$. Since

$$\frac{1}{t^{1-\alpha} - (t-1)^{1-\alpha}} \left|\left|\sup_{\theta\in\Theta}|e_t(\theta) - \epsilon_t(\theta)|\right|\right|_2 \sim_{t\to\infty} \frac{1}{(1-\alpha)t^{-\alpha}} \left|\left|\sup_{\theta\in\Theta}|e_t(\theta) - \epsilon_t(\theta)|\right|\right|_2 ,$$

which tends to 0 as $t\to\infty$ (a consequence of Point 2 of Lemma 3.4), Toeplitz's lemma implies that the right hand side of (53) tends to 0 as $n\to\infty$, and this proves Point 1.
We now prove Point 2. One has for all $\theta\in\Theta$

$$\begin{aligned} ||\nabla[e_t(\theta)^2 - \epsilon_t(\theta)^2]|| &= ||2e_t(\theta)\nabla[e_t(\theta) - \epsilon_t(\theta)] + 2[e_t(\theta) - \epsilon_t(\theta)]\nabla\epsilon_t(\theta)|| \\ &\le 2||e_t(\theta)\nabla[e_t(\theta) - \epsilon_t(\theta)]|| + 2|e_t(\theta) - \epsilon_t(\theta)|.||\nabla\epsilon_t(\theta)||. \end{aligned} \tag{54}$$

so that

$$\begin{aligned} \sup_{\theta\in\Theta}||\nabla(Q_n(\theta) - O_n(\theta))|| \le \frac{2}{n}\sum_{t=1}^n \sup_{\theta\in\Theta}|e_t(\theta)|.\sup_{\theta\in\Theta}||\nabla[e_t(\theta) - \epsilon_t(\theta)]|| \\ + \frac{2}{n}\sum_{t=1}^n \sup_{\theta\in\Theta}|e_t(\theta) - \epsilon_t(\theta)|.\sup_{\theta\in\Theta}||\nabla\epsilon_t(\theta)||. \end{aligned} \tag{55}$$

Lemma 3.4, Points 2 and 4, along with Borel Cantelli's lemma, yields that $\sup_{\theta\in\Theta}|\epsilon_t(\theta)-e_t(\theta)|$ and $\sup_{\theta\in\Theta}||\nabla(\epsilon_t-e_t)(\theta)||$ a.s. tend to 0 as $t\to\infty$. The second term on the right hand side of (55) if then a.s. upper bounded thanks to Cauchy Scwhartz inequality by

$$2\left[\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|e_t(\theta)-\epsilon_t(\theta)|^2\right]^{1/2}\cdot\left[\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}||\nabla\epsilon_t(\theta)||^2\right]^{1/2},$$

which tends to zero thanks to Cesaro's Lemma and the ergodic theorem. And since, by Minkowski's inequality,

$$\left[\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|e_t(\theta)|^2\right]^{1/2}\leq\left[\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|e_t(\theta)-\epsilon_t(\theta)|^2\right]^{1/2}+\left[\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|\epsilon_(\theta)|^2\right]^{1/2},$$

one has that $\left[\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|e_t(\theta)|^2\right]^{1/2}$ is a.s. upper bounded in $n\geq 1$, again by a Cesaro and ergodic theorem argument. The first term on the right hand side of (55) if then again a.s. upper bounded thanks to Cauchy Scwhartz inequality by

$$2\left[\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}||\nabla(e_t-\epsilon_t)(\theta)||^2\right]^{1/2}\cdot\left[\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|e_t(\theta)|^2\right]^{1/2},$$

which tends to zero as $t\to\infty$. Hence (55) implies that $\sup_{\theta\in\Theta}||\nabla(Q_n(\theta)-O_n(\theta))||$ a.s. tends to 0 as $n\to\infty$. Proof of a.s. convergence of $\sup_{\theta\in\Theta}||\nabla^j(Q_n(\theta)-O_n(\theta))||$ to 0 for $j=2,3$ is obtained similarly, using arguments related to Points 3 and 4 from Lemma 3.4.
Let us now prove Point 3. Let $\alpha\in(0,1)$. We deduce from (54), using Minkowski and Holder inequality, that

$$\begin{aligned}n^{\alpha}\left|\left|\sup_{\theta\in\Theta}||\nabla(Q_n(\theta)-O_n(\theta))||\right|\right|_1 &\leq 2\frac{1}{n^{1-\alpha}}\sum_{t=1}^{n}\left|\left|\sup_{\theta\in\Theta}|e_t(\theta)|\right|\right|_4\left|\left|\sup_{\theta\in\Theta}||\nabla[e_t(\theta)-\epsilon_t(\theta)]||\right|\right|_{4/3}\\ &\quad+2\frac{1}{n^{1-\alpha}}\sum_{t=1}^{n}\left|\left|\sup_{\theta\in\Theta}|e_t(\theta)-\epsilon_t(\theta)|\right|\right|_2\left|\left|\sup_{\theta\in\Theta}||\nabla\epsilon_t(\theta)||\right|\right|_2. \quad (56)\end{aligned}$$

Using Point 1 of Lemma 3.4, one has that $||\sup_{\theta\in\Theta}|e_t(\theta)|||_4$ is upper bounded by some constant C. The first term in the righthandside of (56) may thus be upper bounded by

$$2C\frac{1}{n^{1-\alpha}}\sum_{t=1}^{n}\left|\left|\sup_{\theta\in\Theta}||\nabla[e_t(\theta)-\epsilon_t(\theta)]||\right|\right|_{4/3}.$$

Noting that $||\sup_{\theta\in\Theta}||\nabla[e_t(\theta)-\epsilon_t(\theta)]||||_{4/3}\leq C'\,||\sup_{\theta\in\Theta}||\nabla[e_t(\theta)-\epsilon_t(\theta)]||||_{8/5}$ for some constant C', the above expression is, similarly to the argument in (53), a quantity that tends to 0 as $n\to\infty$ thanks to Point 4 in Lemma 3.4 coupled with Toeplitz's lemma. Hence the first term in the right hand side of (56) tends to 0 as $n\to\infty$. Again using Point 1 and Point 2 of the same lemma, and with the same argument, one also has that the second term in the right hand side of (56) tends to 0 as $n\to\infty$, which proves Point 2. □

A.3. Proofs of Proposition 3.6 and Theorem 3.7

Proof of Proposition 3.6. Independence of processes $(\Delta_t)_{t\in\mathbb{Z}}$ and $(\epsilon_t)_{t\in\mathbb{Z}}$ as well their ergodicity yields that, for fixed $j\in\mathbb{N}$, process $((\Delta_{t-1},...,\Delta_{t-j},\epsilon_{t-j}))$ is ergodic. One thus deduces from Expression (9), and using the fact that $(\epsilon_t)_{t\in\mathbb{Z}}$ is a weak white noise, that $O_n(\theta)$ defined by (15) verifies

$$O_n(\theta)\longrightarrow O_\infty(\theta):=\sigma^2\sum_{j=0}^{\infty}\mathbb{E}\left([c_j(\theta,\Delta_0,...,\Delta_{-j})]^2\right)=\sigma^2+\sigma^2\sum_{j=1}^{\infty}\mathbb{E}\left([c_j(\theta,\Delta_0,...,\Delta_{-j})]^2\right)\quad\text{a.s.}\tag{57}$$

as $n\to\infty$ (remember that $c_0(\theta,\Delta_0)=1$). By uniqueness of decomposition (9) in Proposition 3.1, and since $\epsilon_t(\theta_0)=\epsilon_t$, one has that $(c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i}))_{i\in\mathbb{N}}=(1,0,...)$ if and only if $\theta=\theta_0$, and that $O_\infty(\theta)$ given in (57) is minimum at $\theta=\theta_0$, with minimum given by $O_\infty(\theta_0)=\sigma^2$. Let us then deduce that estimator $\check{\theta}_n$ defined in (16) converges a.s. towards θ_0. For this we let a subsequence $(\check{\theta}_{n_k})_{k\in\mathbb{N}}$ converging to some θ^* in the compact set Θ and we prove that $\theta^*=\theta_0$. Indeed, by definition of estimator $\check{\theta}_{n_k}$ one has

$$O_{n_k}(\theta_0)\geq O_{n_k}(\check{\theta}_{n_k})\tag{58}$$

for all $k\in\mathbb{N}$. A Taylor expansion yields the inequality

$$|O_{n_k}(\check{\theta}_{n_k})-O_{n_k}(\theta^*)|\leq||\check{\theta}_{n_k}-\theta^*||.2\frac{1}{n_k}\sum_{t=1}^{n_k}\sup_{\theta\in\Theta}[|\epsilon_t(\theta)|.||\nabla\epsilon_t(\theta)||].\tag{59}$$

But, using the ergodic theorem, one has

$$\begin{aligned}\frac{2}{n_k}\sum_{t=1}^{n_k}\sup_{\theta\in\Theta}[|\epsilon_t(\theta)|.||\nabla\epsilon_t(\theta)||]&\leq\frac{1}{n_k}\sum_{t=1}^{n_k}\left[\sup_{\theta\in\Theta}|\epsilon_t(\theta)|^2+\sup_{\theta\in\Theta}||\nabla\epsilon_t(\theta)||^2\right]\\&\longrightarrow\left|\left|\sup_{\theta\in\Theta}|\epsilon_0(\theta)|\right|\right|_2^2+\left|\left|\sup_{\theta\in\Theta}||\nabla\epsilon_0(\theta)||\right|\right|_2^2<+\infty,\end{aligned}$$

so that one gets from (59) that $O_{n_k}(\check{\theta}_{n_k})-O_{n_k}(\theta^*)\longrightarrow 0$ as $k\to\infty$. Since $O_{n_k}(\theta^*)\longrightarrow O_\infty(\theta^*)$, we obtain, passing to the limit in (58), that

$$O_\infty(\theta_0)\geq O_\infty(\theta^*),$$

hence $\theta^*=\theta_0$ thank to uniqueness of the minimum of $O_\infty(\theta)$. □

Proof of Theorem 3.7. Similarly to the proof of the previous theorem, we let a subsequence $(\hat{\theta}_{n_k})_{k\in\mathbb{N}}$ converging to some θ_* in the compact set Θ and we prove that $\theta_*=\theta_0$ by proving that $O_\infty(\theta_0)=O_\infty(\theta_*)$. By definition of $\hat{\theta}_{n_k}$ we have

$$Q_{n_k}(\theta_0)\geq Q_{n_k}(\hat{\theta}_{n_k}),\quad\forall k\geq 0.\tag{60}$$

Now, a Taylor expansion yields, for all θ' and θ'' in Θ, similarly to the argument in the proof of Proposition 3.6,

$$|Q_{n_k}(\theta')-Q_{n_k}(\theta'')|\leq||\theta'-\theta''||.\frac{1}{n_k}\sum_{t=1}^{n_k}\left[\sup_{\theta\in\Theta}|e_t(\theta)|^2+\sup_{\theta\in\Theta}||\nabla e_t(\theta)||^2\right].\tag{61}$$

Using inequality $(a+b)^2 \leq 2(a^2+b^2)$ for all a and b, we deduce that $\sup_{\theta\in\Theta}|e_t(\theta)|^2 \leq 2(\sup_{\theta\in\Theta}|e_t(\theta)-\epsilon_t(\theta)|^2)+\sup_{\theta\in\Theta}|\epsilon_t(\theta)|^2$. Since a consequence of Point 3 of Lemma 3.4 is that $\sup_{\theta\in\Theta}|e_t(\theta)-\epsilon_t(\theta)|^2$ tends to 0 as $t\to\infty$, the ergodic theorem yields that

$$\frac{1}{n_k}\sum_{t=1}^{n_k}\left[\sup_{\theta\in\Theta}|e_t(\theta)|^2+\sup_{\theta\in\Theta}||\nabla e_t(\theta)||^2\right] \longrightarrow \left|\left|\sup_{\theta\in\Theta}|\epsilon_0(\theta)|\right|\right|_2^2+\left|\left|\sup_{\theta\in\Theta}||\nabla\epsilon_0(\theta)||\right|\right|_2^2 < +\infty$$

as $k\to\infty$. Thanks to (61) and Point 1 of Proposition 3.5, we thus deduce that $Q_{n_k}(\theta_0)\longrightarrow O_\infty(\theta_0)$ and $Q_{n_k}(\hat{\theta}_{n_k})\longrightarrow O_\infty(\theta_*)$ as $k\to\infty$, and we conclude in the same way as in proof of Theorem 3.6. $\square$

A.4. Proofs of Theorem 3.8

Let us introduce the following matrices and vectors

$$I_n(\theta) := \mathrm{Var}\left(\sqrt{n}\nabla O_n(\theta)\right) = (I_n(l,r)(\theta))_{l,r=1\dots(p+q)K} \in \mathbb{R}^{(p+q)K\times(p+q)K}, \quad n\in\mathbb{N},, \tag{62}$$

$$Y_k(\theta) := \epsilon_k(\theta)\nabla\epsilon_k(\theta) = (Y_k(l)(\theta))_{l=1\dots(p+q)K} \in \mathbb{R}^{(p+q)K\times 1}, \quad k\in\mathbb{Z}, \tag{63}$$

Theorem 3.8 can be established using the following lemmas.

Lemma A.1 (Davydov (1968))**.** *Let p, q and r three positive numbers such that $p^{-1}+q^{-1}+r^{-1}=1$. Davydov [14] showed that*

$$|Cov(X,Y)| \leq K_0\|X\|_p\|Y\|_q\left[\alpha\left\{\sigma(X),\sigma(Y)\right\}\right]^{1/r}, \tag{64}$$

where $\|X\|_p^p=\mathbb{E}(X^p)$, K_0 is an universal constant, and $\alpha\{\sigma(X),\sigma(Y)\}$ denotes the strong mixing coefficient between the σ-fields $\sigma(X)$ and $\sigma(Y)$ generated by the random variables X and Y, respectively.

Lemma A.2. *Let the assumptions of Theorem 3.8 be satisfied. For all l, r in $1,\dots,(p+q)K$ and $\theta\in\Theta$ we have*

$$I_n(l,r)(\theta)\longrightarrow I(l,r)(\theta) := \sum_{k=-\infty}^{\infty}c_k(l,r)(\theta), \quad n\to+\infty,$$

where $c_k(l,r)(\theta)=\mathrm{Cov}\left(Y_t(l)(\theta),Y_{t-k}(r)(\theta)\right)$, $k\in\mathbb{Z}$, the former being a convergent series.

Proof of Lemma A.2: Let us write

$$\nabla\epsilon_t(\theta)=\left(\frac{\partial\epsilon_t(\theta)}{\partial\theta_1},\dots,\frac{\partial\epsilon_t(\theta)}{\partial\theta_{(p+q)K}}\right)',$$

where $\epsilon_t(\theta)$ is given by (9). The process $(Y_k(\theta))_k$ is strictly stationary and ergodic. Moreover, we have

$$\begin{aligned} I_n(\theta)=\mathrm{Var}\left(\sqrt{n}\frac{\partial}{\partial\theta}O_n(\theta)\right) &= \mathrm{Var}\left(\frac{2}{\sqrt{n}}\sum_{t=1}^{n}Y_t(\theta)\right)=\frac{4}{n}\sum_{t,s=1}^{n}\mathrm{Cov}\left(Y_t(\theta),Y_s(\theta)\right) \\ &= \frac{4}{n}\sum_{k=-n+1}^{n-1}(n-|k|)\mathrm{Cov}\left(Y_t(\theta),Y_{t-k}(\theta)\right). \end{aligned}$$

From Proposition 3.1 and Lemma 13, we have

$$\epsilon_t(\theta)=\sum_{i=0}^{\infty}c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})\epsilon_{t-i}\text{ and }\frac{\partial\epsilon_t(\theta)}{\partial\theta_l}=\sum_{i=0}^{\infty}c_{i,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-i})\epsilon_{t-i},\text{ for }l=1,\dots,(p+q)K,$$

where we recall that $c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})$ is defined by (10), and

$$\begin{aligned}c_{i,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-i}) &= \frac{\partial}{\partial\theta_l}c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})\\ &= \frac{\partial}{\partial\theta_l}\left(\sum_{k=0}^{i}e_1\Phi(\Delta_{t-1},\theta)\dots\Phi(\Delta_{t-k+1},\theta)M\Psi(\Delta_{t-k})\dots\Psi(\Delta_{t-i})e'_{p+1}\right),\end{aligned}$$

with the following upper bound holding thanks to (14):

$$\mathbb{E}\sup_{\theta\in\Theta}(c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i}))^2\le C\rho^i\text{ and }\mathbb{E}\sup_{\theta\in\Theta}(c_{i,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-i}))^2\le C\rho^i,\quad\forall i.$$

Let

$$\begin{aligned}\beta_{i,j,i',j',k}(l,r)(\theta) &= \mathbb{E}\left[c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})c_{j,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-j})c_{i'}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-i'})\right.\\ &\quad\left.c_{j',r}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-j'})\right]\mathbb{E}\left[\epsilon_{t-i}\epsilon_{t-j}\epsilon_{t-k-i'}\epsilon_{t-k-j'}\right]\\ &\quad-\mathbb{E}\left[c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})c_{j,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-j})\right]\\ &\quad\times\mathbb{E}\left[c_{i'}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-i'})c_{j',r}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-j'})\right]\mathbb{E}\left[\epsilon_{t-i}\epsilon_{t-j}\right]\\ &\quad\times\mathbb{E}\left[\epsilon_{t-k-i'}\epsilon_{t-k-j'}\right]\\ &= \mathbb{E}\left[c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})c_{j,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-j})c_{i'}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-i'})\right.\\ &\quad\left.c_{j',r}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-j'})\right]\mathrm{Cov}\left(\epsilon_{t-i}\epsilon_{t-j},\epsilon_{t-k-i'}\epsilon_{t-k-j'}\right)\\ &\quad+\mathrm{Cov}\left(c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})c_{j,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-j}),c_{i'}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-i'})\right.\\ &\quad\left.c_{j',r}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-j'})\right)\mathbb{E}\left[\epsilon_{t-i}\epsilon_{t-j}\right]\mathbb{E}\left[\epsilon_{t-k-i'}\epsilon_{t-k-j'}\right].\end{aligned}$$

We then obtain

$$c_k(l,r)(\theta)=\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty}\beta_{i,j,i',j',k}(l,r)(\theta),\quad k\in\mathbb{Z}.$$

The Cauchy Schwarz inequality implies that

$$\begin{aligned}&|\mathbb{E}[c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})c_{j,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-j})c_{i'}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-i'})\\ \times\ &c_{j',r}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-j'})]|\le\left(\mathbb{E}[c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})c_{j,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-j})]^2\right)^{1/2}\\ \times\ &\left(\mathbb{E}[c_{i'}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-i'})c_{j',r}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-j'})]^2\right)^{1/2}\le\left(\mathbb{E}[c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})]^4\right.\\ \times\ &\left.\mathbb{E}[c_{j,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-j})]^4\right)^{1/4}\left(\mathbb{E}[c_{i'}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-i'})]^4\mathbb{E}[c_{j',r}(\theta,\Delta_{t-k-1},\dots,\Delta_{t-k-j'})]^4\right)^{1/4}\\ \le\ &C\rho^{i+j+i'+j'}.\end{aligned}\tag{65}$$

First, suppose that $k\ge0$, for all l, r in $1,\dots,(p+q)K$ and $\theta\in\Theta$, in view of (65) it follows that

$$\begin{aligned}|c_k(l,r)(\theta)| &= |\mathrm{cov}\left(Y_t(l)(\theta),Y_{t-k}(r)(\theta)\right)|=4\left|\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty}\beta_{i,j,i',j',k}(l,r)(\theta)\right|\\ &\le 4(g_1+g_2+g_3+g_4+g_5+h_1+h_2+h_3),\end{aligned}$$

where

$$
\begin{aligned}
g_1 &= \sum_{i>[k/2]} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \kappa\rho^{i+j+i'+j'} \left|\mathrm{Cov}\left(\epsilon_{t-i}\epsilon_{t-j}, \epsilon_{t-k-i'}\epsilon_{t-k-j'}\right)\right|, \\
g_2 &= \sum_{i=0}^{\infty} \sum_{j>[k/2]} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \kappa\rho^{i+j+i'+j'} \left|\mathrm{Cov}\left(\epsilon_{t-i}\epsilon_{t-j}, \epsilon_{t-k-i'}\epsilon_{t-k-j'}\right)\right| \\
g_3 &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{i'>[k/2]} \sum_{j'=0}^{\infty} \kappa\rho^{i+j+i'+j'} \left|\mathrm{Cov}\left(\epsilon_{t-i}\epsilon_{t-j}, \epsilon_{t-k-i'}\epsilon_{t-k-j'}\right)\right|, \\
g_4 &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'>[k/2]} \kappa\rho^{i+j+i'+j'} \left|\mathrm{Cov}\left(\epsilon_{t-i}\epsilon_{t-j}, \epsilon_{t-k-i'}\epsilon_{t-k-j'}\right)\right| \\
g_5 &= \sum_{i=0}^{[k/2]} \sum_{j=0}^{[k/2]} \sum_{i'=0}^{[k/2]} \sum_{j'=0}^{[k/2]} \kappa\rho^{i+j+i'+j'} \left|\mathrm{Cov}\left(\epsilon_{t-i}\epsilon_{t-j}, \epsilon_{t-k-i'}\epsilon_{t-k-j'}\right)\right|, \\
h_1 &= \sigma^4 \sum_{i>[k/2]} \sum_{i'=0}^{\infty} |\mathrm{Cov}(c_i(\theta, \Delta_{t-1}, \dots, \Delta_{t-i})c_{i,l}(\theta, \Delta_{t-1}, \dots, \Delta_{t-i}), \\
&\quad c_{i'}(\theta, \Delta_{t-k-1}, \dots, \Delta_{t-k-i'})c_{i',r}(\theta, \Delta_{t-k-1}, \dots, \Delta_{t-k-i'}))\big|, \\
h_2 &= \sigma^4 \sum_{i=0}^{\infty} \sum_{i'>[k/2]} |\mathrm{Cov}(c_i(\theta, \Delta_{t-1}, \dots, \Delta_{t-i})c_{i,l}(\theta, \Delta_{t-1}, \dots, \Delta_{t-i}), \\
&\quad c_{i'}(\theta, \Delta_{t-k-1}, \dots, \Delta_{t-k-i'})c_{i',r}(\theta, \Delta_{t-k-1}, \dots, \Delta_{t-k-i'}))\big|, \\
h_3 &= \sigma^4 \sum_{i=0}^{[k/2]} \sum_{i'=0}^{[k/2]} |\mathrm{Cov}(c_i(\theta, \Delta_{t-1}, \dots, \Delta_{t-i})c_{i,l}(\theta, \Delta_{t-1}, \dots, \Delta_{t-i}), \\
&\quad c_{i'}(\theta, \Delta_{t-k-1}, \dots, \Delta_{t-k-i'})c_{i',r}(\theta, \Delta_{t-k-1}, \dots, \Delta_{t-k-i'}))\big|.
\end{aligned}
$$

Because

$$
\left|\mathrm{Cov}\left(\epsilon_{t-i}\epsilon_{t-j}, \epsilon_{t-k-i'}\epsilon_{t-k-j'}\right)\right| \leq \sqrt{\mathbb{E}\left[\epsilon_{t-i}\epsilon_{t-j}\right]^2 \mathbb{E}\left[\epsilon_{t-k-i'}\epsilon_{t-k-j'}\right]^2} \leq \mathbb{E}\left|\epsilon_t\right|^4 < \infty
$$

by Assumption (**A3**), we have It follows that

$$
g_1 = \sum_{i>[k/2]} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \kappa\rho^{i+j+i'+j'} \left|\mathrm{Cov}\left(\epsilon_{t-i}\epsilon_{t-j}, \epsilon_{t-k-i'}\epsilon_{t-k-j'}\right)\right| \leq \kappa_1\rho^{k/2},
$$

for some positive constant κ_1. Using the same arguments we obtain that $g_i \quad (i = 2, 3, 4)$ is bounded by $\kappa_i\rho^{k/2}$. Furthermore, **(A3)** and the Cauchy Schwartz inequality yields that $\|\epsilon_i\epsilon_{i'}\|_{2+\nu} < +\infty$ for any i and i' in $\mathbb{Z}$. Lemma A.1 thus entails that

$$
\begin{aligned}
g_5 &= \sum_{i=0}^{[k/2]} \sum_{j=0}^{[k/2]} \sum_{i'=0}^{[k/2]} \sum_{j'=0}^{[k/2]} \kappa\rho^{i+j+i'+j'} \left|\mathrm{Cov}\left(\epsilon_{t-i}\epsilon_{t-j}, \epsilon_{t-k-i'}\epsilon_{t-k-j'}\right)\right| \\
&\leq \sum_{i=0}^{[k/2]} \sum_{j=0}^{[k/2]} \sum_{i'=0}^{[k/2]} \sum_{j'=0}^{[k/2]} \kappa_5\rho^{i+j+i'+j'} \left\|\epsilon_{t-i}\epsilon_{t-j}\right\|_{2+\nu} \left\|\epsilon_{t-k-i'}\epsilon_{t-k-j'}\right\|_{2+\nu} \\
&\quad \times \left\{\alpha_\epsilon\left(\min\left[k+j'-i, k+i'-i, k+j'-j, k+i'-j\right]\right)\right\}^{\nu/(2+\nu)} \leq \kappa'\alpha_\epsilon^{\nu/(2+\nu)}\left([k/2]\right).
\end{aligned}
$$

Since

$$\begin{aligned} & \left|\operatorname{Cov}(c_i(\theta,\Delta_{t-1},\ldots,\Delta_{t-i})c_{i,l}(\theta,\Delta_{t-1},\ldots,\Delta_{t-i}),c_{i'}(\theta,\Delta_{t-k-1},\ldots,\Delta_{t-k-i'})\right. \\ \times\quad & \left. c_{i',r}(\theta,\Delta_{t-k-1},\ldots,\Delta_{t-k-i'}))\right|\leq C\rho^{i+i'}, \end{aligned}$$

we have

$$\begin{aligned} h_1 &= \sigma^4\sum_{i>[k/2]}\sum_{i'=0}^{\infty}\left|\operatorname{Cov}(c_i(\theta,\Delta_{t-1},\ldots,\Delta_{t-i})c_{i,l}(\theta,\Delta_{t-1},\ldots,\Delta_{t-i}),\right. \\ & \qquad \left. c_{i'}(\theta,\Delta_{t-k-1},\ldots,\Delta_{t-k-i'})c_{i',r}(\theta,\Delta_{t-k-1},\ldots,\Delta_{t-k-i'}))\right|\leq\kappa_1'\rho^{k/2}, \end{aligned}$$

for some positive constant κ_1'. Using the same arguments we obtain that h_2 is bounded by $\kappa_2'\rho^{k/2}$. The $\alpha-$mixing property (see Theorem 14.1 in [13], p. 210) and Lemma A.1, along with (13), entail that

$$\begin{aligned} h_3 &= \sigma^4\sum_{i=0}^{[k/2]}\sum_{i'=0}^{[k/2]}\left|\operatorname{Cov}(c_i(\theta,\Delta_{t-1},\ldots,\Delta_{t-i})c_{i,l}(\theta,\Delta_{t-1},\ldots,\Delta_{t-i}),\right. \\ & \qquad \left. c_{i'}(\theta,\Delta_{t-k-1},\ldots,\Delta_{t-k-i'})c_{i',r}(\theta,\Delta_{t-k-1},\ldots,\Delta_{t-k-i'}))\right| \\ &\leq \sum_{i=0}^{[k/2]}\sum_{i'=0}^{[k/2]}\kappa_6\left\|c_i(\theta,\Delta_{t-1},\ldots,\Delta_{t-i})c_{i,l}(\theta,\Delta_{t-1},\ldots,\Delta_{t-i})\right\|_{2+\nu} \\ & \qquad \times\left\|c_{i'}(\theta,\Delta_{t-k-1},\ldots,\Delta_{t-k-i'})c_{i',r}(\theta,\Delta_{t-k-1},\ldots,\Delta_{t-k-i'}))\right\|_{2+\nu} \\ & \qquad \times\left\{\alpha_\Delta\left(k+1-i\right)\right\}^{\nu/(2+\nu)}\leq\kappa_3'\alpha_\Delta^{\nu/(2+\nu)}\left([k/2]\right). \end{aligned}$$

It follows that

$$\sum_{k=0}^{\infty}|c_k(l,r)(\theta)|\leq\kappa\sum_{k=0}^{\infty}\rho^{|k|/2}+\kappa'\sum_{k=0}^{\infty}\alpha_\epsilon^{\nu/(2+\nu)}\left([k/2]\right)+\kappa''\sum_{k=0}^{\infty}\alpha_\Delta^{\nu/(2+\nu)}\left([k/2]\right)<\infty,$$

by Assumption (**A2**). The same bounds clearly holds for

$$\sum_{k=-\infty}^{0}|c_k(l,r)(\theta)|,$$

which shows that

$$\sum_{k=-\infty}^{\infty}|c_k(l,r)(\theta)|<\infty.$$

Then, the dominated convergence theorem gives

$$I_n(l,r)(\theta)=\frac{1}{n}\sum_{k=-n+1}^{n-1}(n-|k|)c_k(l,r)(\theta)\longrightarrow I(l,r)(\theta):=\sum_{k=-\infty}^{\infty}c_k(l,r)(\theta),\quad n\to+\infty,$$

and completes the proof. $\square$

Lemma A.3. *Under the assumptions of Theorem 3.8, one has convergence in distribution of the random vector*

$$\sqrt{n}\nabla Q_n(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,I), \text{ as } n\to\infty$$

where we recall that matrix I is given by (19).

Proof of Lemma A.3: In view of Proposition 3.5, it is easy to see that

$$\sqrt{n}\nabla\left(Q_n - O_n\right)(\theta_0) = o_{\mathbb{P}}(1).$$

Thus $\nabla Q_n(\theta_0)$ and $\nabla O_n(\theta_0)$ have the same asymptotic distribution. Therefore, it remains to show that

$$\sqrt{n}\nabla O_n(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,I), \text{ as } n\to\infty.$$

For l, in $1,\dots,(p+q)K$ and $\theta\in\Theta$, we have

$$\frac{\partial\epsilon_t(\theta)}{\partial\theta_l} = \sum_{i=1}^{\infty} c_{i,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-i})\epsilon_{t-i}, \tag{66}$$

where the sequence $c_{i,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-i})$ is such that $\mathbb{E}\sup_{\theta\in\Theta}|(c_{i,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-i}))^2\to 0$ at a geometric rate as $i\to\infty$ (see Lemma 3.3). Moreover, note that

$$\sqrt{n}\frac{\partial O_n(\theta)}{\partial\theta_l} = \frac{2}{\sqrt{n}}\sum_{t=1}^{n} Y_t(l)(\theta) = \frac{2}{\sqrt{n}}\sum_{t=1}^{n}\sum_{i=0}^{\infty} c_i(\theta,\Delta_{t-1},\dots,\Delta_{t-i})\epsilon_{t-i}\sum_{j=1}^{\infty} c_{j,l}(\theta,\Delta_{t-1},\dots,\Delta_{t-j})\epsilon_{t-j}.$$

Since $\nabla\epsilon_t(\theta_0)$ belongs to the Hilbert space $\mathcal{H}_\epsilon(t-1)$, the random variables $\epsilon_t(\theta_0)$ and $\nabla\epsilon_t(\theta_0)$ are orthogonal and it is easy to verify that $\mathbb{E}\left[\sqrt{n}\nabla O_n(\theta_0)\right]=0$. Now, we have for all m

$$\sqrt{n}\frac{\partial O_n(\theta_0)}{\partial\theta_l} = \frac{2}{\sqrt{n}}\sum_{t=1}^{n} Y_{t,m}(l) + \frac{2}{\sqrt{n}}\sum_{t=1}^{n} Z_{t,m}(l)$$

where

$$\begin{aligned} Y_{t,m}(l) &= \sum_{j=1}^{m} c_{j,l}(\theta_0,\Delta_{t-1},\dots,\Delta_{t-j})\epsilon_t\epsilon_{t-j} \\ Z_{t,m}(l) &= \sum_{j=m+1}^{\infty} c_{j,l}(\theta_0,\Delta_{t-1},\dots,\Delta_{t-j})\epsilon_t\epsilon_{t-j}. \end{aligned}$$

Let

$$\begin{aligned} Y_{t,m} &:= Y_{t,m}(\theta_0) = (Y_{t,m}(1),\dots,Y_{t,m}((p+q)K))' \text{ and} \\ Z_{t,m} &:= Z_{t,m}(\theta_0) = (Z_{t,m,}(1),\dots,Z_{t,m}((p+q)K))'. \end{aligned}$$

The processes $(Y_{t,m})_t$ and $(Z_{t,m})_t$ are stationary and centered. Moreover, under Assumption **(A2)** and m fixed, the process $Y=(Y_{t,m})_t$ is strongly mixing (see [13], Theorem 14.1 p. 210), with mixing coefficients $\alpha_Y(h)\le\alpha_{\Delta,\epsilon}\left(\max\{0,h-m\}\right)\le\alpha_\Delta\left(\max\{0,h-m+1\}\right)+\alpha_\epsilon\left(\max\{0,h-m\}\right)$, by independence of $(\Delta_t)_{t\in\mathbb{Z}}$ and $(\epsilon_t)_{t\in\mathbb{Z}}$. Applying the central limit theorem (CLT) for mixing processes (see [31]) we directly obtain

$$\frac{2}{\sqrt{n}}\sum_{t=1}^{n} Y_{t,m} \xrightarrow{\mathcal{D}} \mathcal{N}(0,I_m), \quad I_m = 4\sum_{h=-\infty}^{\infty}\operatorname{Cov}\left(Y_{t,m},Y_{t-h,m}\right).$$

As in [20] (see Lemma 3), one can show that $I = \lim_{m\to\infty} I_m$ exists. Since $\|Z_{t,m}\|_2 \to 0$ at an exponential rate when $m \to \infty$, using the arguments given in [20] (see Lemma 4), one show that

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left\{ \left\| n^{-1/2} \sum_{t=1}^{n} Z_{t,m} \right\| > \varepsilon \right\} = 0 \tag{67}$$

for every $\varepsilon > 0$ (see the following lemma A.4). From a standard result (see *e.g.* [11], Proposition 6.3.9), we deduce that

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \nabla O_n(\theta_0) = \frac{2}{\sqrt{n}} \sum_{t=1}^{n} Y_{t,m} + \frac{2}{\sqrt{n}} \sum_{t=1}^{n} Z_{t,m} \xrightarrow{\mathcal{D}} \mathcal{N}(0, I),$$

which completes the proof. □

Lemma A.4. *Under the assumptions of Theorem 3.8, (67) holds, that is*

$$\lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{P}\left\{ \left\| n^{-1/2} \sum_{t=1}^{n} Z_{t,m} \right\| > \varepsilon \right\} = 0.$$

Proof of Lemma A.4: For $l = 1, \dots, (p+q)K$, by stationarity we have

$$\begin{aligned}
\mathrm{Var}\left(\frac{2}{\sqrt{n}} \sum_{t=1}^{n} Z_{t,m}(l) \right) &= \frac{4}{n} \sum_{t,s=1}^{n} \mathrm{Cov}(Z_{t,m}(l), Z_{s,m}(l)) \\
&= \frac{4}{n} \sum_{|h|<n} (n - |h|) \mathrm{Cov}(Z_{t,m}(l), Z_{t-h,m}(l)) \\
&\le 4 \sum_{h=-\infty}^{\infty} \left| \mathrm{Cov}(Z_{t,m}(l), Z_{t-h,m}(l)) \right|.
\end{aligned}$$

Consider first the case $h \ge 0$. Because $\mathbb{E} \sup_{\theta\in\Theta} (c_{j,l}(\theta_0, \Delta_{t-1}, \dots, \Delta_{t-j}))^2 \le \kappa\rho^j$ (see 13), using also $\mathbb{E}|\epsilon_t|^4 < \infty$, for $[h/2] \le m$, it follows from the Hölder inequality that

$$\sup_h \left| \mathrm{Cov}(Z_{t,m}(l), Z_{t-h,m}(l)) \right| = \sup_h \left| \mathbb{E}(Z_{t,m}(l) Z_{t-h,m}(l)) \right| \le \kappa\rho^m. \tag{68}$$

Let $h > 0$ such that $[h/2] > m$. Write

$$Z_{t,m} = Z_{t,m}^{h-}(l) + Z_{t,m}^{h+}(l),$$

where

$$Z_{t,m}^{h-}(l) = \sum_{j=m+1}^{[h/2]} c_{j,l}(\theta_0, \Delta_{t-1}, \dots, \Delta_{t-j}) \epsilon_t \epsilon_{t-j}, \quad Z_{t,m}^{h+}(l) = \sum_{j=[h/2]+1}^{\infty} c_{j,l}(\theta_0, \Delta_{t-1}, \dots, \Delta_{t-j}) \epsilon_t \epsilon_{t-j}.$$

Note that $Z_{t,m}^{h-}(l)$ belongs to the σ-field generated by $\{\Delta_{t-1}, \dots, \Delta_{t-[h/2]}, \epsilon_t, \epsilon_{t-1}, \dots, \epsilon_{t-[h/2]}\}$ and that $Z_{t-h,m}(l)$ belongs to the σ-field generated by $\{\Delta_{t-h-1}, \Delta_{t-h-2}, \dots, \epsilon_{t-h}, \epsilon_{t-h-1}, \dots\}$.

Note also that, by **(A3)**, $\mathbb{E}|Z_{t,m}^{h-}(l)|^{2+\nu} < \infty$ and $\mathbb{E}|Z_{t-h,m}(l)|^{2+\nu} < \infty$. The $\alpha-$mixing property and Lemma A.1 then entail that

$$\begin{aligned}
\left|\mathrm{Cov}(Z_{t,m}^{h-}(l), Z_{t-h,m}(l))\right| &\leq \kappa_1 \sum_{j=m+1}^{[h/2]} \sum_{j'=m+1}^{\infty} \left\|c_{j',l}(\theta_0, \Delta_{t-h-1}, \dots, \Delta_{t-h-j'})\epsilon_t\epsilon_{t-j'}\right\|_{2+\nu} \\
&\quad \times \left\|c_{j,l}(\theta_0, \Delta_{t-1}, \dots, \Delta_{t-j})\epsilon_t\epsilon_{t-j}\right\|_{2+\nu} \left[\alpha_{\Delta,\epsilon}([h/2])\right]^{\nu/(2+\nu)} \\
&\leq \kappa_2 \sum_{j=m+1}^{[h/2]} \sum_{j'=m+1}^{\infty} \rho^j\rho^{j'} \left[\alpha_\epsilon^{\nu/(2+\nu)}([h/2]) + \alpha_\Delta^{\nu/(2+\nu)}([h/2])\right] \\
&\leq \kappa\rho^m \left[\alpha_\epsilon^{\nu/(2+\nu)}([h/2]) + \alpha_\Delta^{\nu/(2+\nu)}([h/2])\right]. \qquad (69)
\end{aligned}$$

By the argument used to show (68), we also have

$$\left|\mathrm{Cov}(Z_{t,m}^{h+}(l), Z_{t-h,m}(l))\right| \leq \kappa\rho^h\rho^m. \qquad (70)$$

In view of (68), (69) and (70), we obtain

$$\sum_{h=0}^{\infty} |\mathrm{Cov}(Z_{t,m}(l), Z_{t-h,m}(l))| \leq \kappa m\rho^m + \sum_{h=m}^{\infty} \left\{\kappa\rho^h\rho^m + \kappa\rho^m \left[\alpha_\epsilon^{\nu/(2+\nu)}([h/2]) + \alpha_\Delta^{\nu/(2+\nu)}([h/2])\right]\right\} \to 0$$

as $m \to \infty$ by **(A2)**. This implies that

$$\sup_n \mathrm{Var}\left(\frac{2}{\sqrt{n}} \sum_{t=1}^{n} Z_{t,m}(l)\right) \xrightarrow[m\to\infty]{} 0. \qquad (71)$$

We have the same bound for $h < 0$. The conclusion follows from (71). □

Lemma A.5. *Under the assumptions of Theorem 3.8, almost surely*

$$\nabla^2 Q_n(\theta_0) \longrightarrow J, \quad n \to \infty,$$

where J exists (is given by (18)) and is invertible.

Proof of Lemma A.5: For all l, r in $1, \dots, (p+q)K$, in view of Proposition 3.5, we have almost surely

$$\left|\frac{\partial^2}{\partial\theta_l\partial\theta_r}\left(Q_n(\theta_0) - O_n(\theta_0)\right)\right| \to 0, \text{ as } t \to \infty.$$

Thus $\partial^2 Q_n(\theta_0)/\partial\theta_l\partial\theta_r$ and $\partial^2 O_n(\theta_0)/\partial\theta_l\partial\theta_r$ have almost surely the same asymptotic distribution. From (9) and (13), there exists a sequence $(c_{i,l,r}(\theta, \Delta_{t-1}, \dots, \Delta_{t-i}))_{i\in\mathbb{N}}$ such that

$$\frac{\partial^2\epsilon_t(\theta)}{\partial\theta_l\partial\theta_r} = \sum_{i=1}^{\infty} c_{i,l,r}(\theta, \Delta_{t-1}, \dots, \Delta_{t-i})\epsilon_{t-i} \text{ with } \mathbb{E}(c_{i,l,r}(\theta, \Delta_{t-1}, \dots, \Delta_{t-i}))^2 \leq C\rho^i, \forall i. \quad (72)$$

This implies that $\partial^2\epsilon_t(\theta)/\partial\theta_l\partial\theta_r$ belongs to L^2. In the other hand, we have

$$\begin{aligned}
\frac{\partial^2 O_n(\theta)}{\partial\theta_l\partial\theta_r} &= \frac{2}{n}\sum_{t=1}^{n} \epsilon_t(\theta)\frac{\partial^2\epsilon_t(\theta)}{\partial\theta_l\partial\theta_r} + \frac{2}{n}\sum_{t=1}^{n} \frac{\partial\epsilon_t(\theta)}{\partial\theta_l}\frac{\partial\epsilon_t(\theta)}{\partial\theta_r} \\
&\longrightarrow 2\mathbb{E}\left(\epsilon_t(\theta)\frac{\partial^2\epsilon_t(\theta)}{\partial\theta_l\partial\theta_r}\right) + 2\mathbb{E}\left(\frac{\partial\epsilon_t(\theta)}{\partial\theta_l}\frac{\partial\epsilon_t(\theta)}{\partial\theta_r}\right), \text{ as } n \to \infty,
\end{aligned}$$

by ergodic theorem. Using the uncorrelatedness between $\epsilon_t(\theta_0)$ and the linear past $\mathcal{H}_\epsilon(t-1)$, $\partial\epsilon_t(\theta_0)/\partial\theta_l \in \mathcal{H}_\epsilon(t-1)$, and $\partial^2\epsilon_t(\theta_0)/\partial\theta_l\partial\theta_r \in \mathcal{H}_\epsilon(t-1)$, we have

$$\mathbb{E}\left(\frac{\partial^2 O_n(\theta_0)}{\partial\theta_l\partial\theta_r}\right) = 2\mathbb{E}\left(\frac{\partial\epsilon_t(\theta_0)}{\partial\theta_l}\frac{\partial\epsilon_t(\theta_0)}{\partial\theta_r}\right) = J(l,r). \tag{73}$$

Therefore, J is the covariance matrix of $\sqrt{2}\partial\epsilon_t(\theta_0)/\partial\theta$. If J is singular, then there exists a vector $\boldsymbol{c} = (c_1,\dots,c_{(p+q)K})' \neq 0$ such that $\boldsymbol{c}'J\boldsymbol{c} = 0$. Thus we have

$$\sum_{k=1}^{(p+q)K} c_k \frac{\partial\epsilon_t(\theta_0)}{\partial\theta_k} = 0,\ a.s. \tag{74}$$

Differentiating the two sides of (4) yields

$$-\sum_{i=1}^{p}(g_i^a)^*(\Delta_t,\theta_0)X_{t-i} = \sum_{k=1}^{(p+q)K} c_k\frac{\partial\epsilon_t(\theta_0)}{\partial\theta_k} - \sum_{j=1}^{q} g_j^b(\Delta_t,\theta_0)\sum_{k=1}^{(p+q)K} c_k\frac{\partial\epsilon_{t-j}(\theta_0)}{\partial\theta_k} - \sum_{j=1}^{q}(g_j^b)^*(\Delta_t,\theta_0)\epsilon_{t-j}(\theta_0)$$

where

$$(g_i^a)^*(\Delta_t,\theta_0) = \sum_{k=1}^{(p+q)K} c_k\frac{\partial g_i^a(\Delta_t,\theta_0)}{\partial\theta_k} \text{ and } (g_j^b)^*(\Delta_t,\theta_0) = \sum_{k=1}^{(p+q)K} c_k\frac{\partial g_j^b(\Delta_t,\theta_0)}{\partial\theta_k}.$$

Because (74) is satisfied for all t, we have

$$\sum_{i=1}^{p}(g_i^a)^*(\Delta_t,\theta_0)X_{t-i} = \sum_{j=1}^{q}(g_j^b)^*(\Delta_t,\theta_0)\epsilon_{t-j}(\theta_0).$$

The latter equation yields a RCARMA$(p-1,q-1)$ representation at best. The identifiability assumption (see Proposition 3.1) excludes the existence of such representation. Thus

$$(g_i^a)^*(\Delta_t,\theta_0) = \sum_{k=1}^{(p+q)K} c_k\frac{\partial g_i^a(\Delta_t,\theta_0)}{\partial\theta_k} = 0 \text{ and } (g_j^b)^*(\Delta_t,\theta_0) = \sum_{k=1}^{(p+q)K} c_k\frac{\partial g_j^b(\Delta_t,\theta_0)}{\partial\theta_k} = 0$$

and the conclusion follows. □

Proof of Theorem 3.8: For all $i,j,k = 1,\dots,K(p+q)$ we have

$$\begin{aligned}\frac{\partial^3 O_n(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k} &= \frac{2}{n}\sum_{t=1}^{n}\left\{\epsilon_t(\theta)\frac{\partial^3\epsilon_t(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right\} + \frac{2}{n}\sum_{t=1}^{n}\left\{\frac{\partial\epsilon_t(\theta)}{\partial\theta_i}\frac{\partial^2\epsilon_t(\theta)}{\partial\theta_j\partial\theta_k}\right\}\\ &+ \frac{2}{n}\sum_{t=1}^{n}\left\{\frac{\partial^2\epsilon_t(\theta)}{\partial\theta_i\partial\theta_j}\frac{\partial\epsilon_t(\theta)}{\partial\theta_k}\right\} + \frac{2}{n}\sum_{t=1}^{n}\left\{\frac{\partial\epsilon_t(\theta)}{\partial\theta_j}\frac{\partial^2\epsilon_t(\theta)}{\partial\theta_i\partial\theta_k}\right\}.\end{aligned}$$

Using the ergodic theorem, the Cauchy Schwarz inequality and Lemma 3.4, we obtain

$$\sup_n\sup_{\theta\in\Theta}\left|\frac{\partial^3 O_n(\theta)}{\partial\theta_i\partial\theta_j\partial\theta_k}\right| < +\infty. \tag{75}$$

In view of Proposition 3.5, we have almost surely

$$\sup_{\theta\in\Theta}\left|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_k}\left(Q_n(\theta)-O_n(\theta)\right)\right|\longrightarrow 0,\ \text{as } n\to\infty.$$

Thus $\partial^3 Q_n(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k$ and $\partial^2 O_n(\theta)/\partial\theta_i\partial\theta_j\partial\theta_k$ have almost surely the same asymptotic distribution. In view of Theorem 3.6 and **(A4)**, we have almost surely $\hat{\theta}_n\longrightarrow\theta_0\in\overset{\circ}{\Theta}$. Thus $\nabla Q_n(\hat{\theta}_n)=0_{\mathbb{R}^{(p+q)K}}$ for sufficiently large n, and a Taylor expansion gives for all $r\in\{1,...,(p+q)K\}$,

$$0=\sqrt{n}\frac{\partial}{\partial\theta_r}Q_n(\theta_0)+\nabla\frac{\partial}{\partial\theta_r}Q_n(\theta^*_{n,r})\sqrt{n}\left(\hat{\theta}_n-\theta_0\right),\tag{76}$$

where $\theta^*_{n,r}$ lies on the segment in $\mathbb{R}^{(p+q)K}$ with endpoints $\hat{\theta}_n$ and θ_0. Using again a Taylor expansion, Theorem 3.7 and (75), we obtain for all $l=1,\dots,(p+q)K$,

$$\begin{aligned}\left|\frac{\partial^2 Q_n(\theta^*_{n,r})}{\partial\theta_l\partial\theta_r}-\frac{\partial^2 Q_n(\theta_0)}{\partial\theta_l\partial\theta_r}\right| &\leq \sup_n\sup_{\theta\in\Theta}\left\|\nabla\left(\frac{\partial^2}{\partial\theta_l\partial\theta_r}Q_n(\theta)\right)\right\|\left\|\theta^*_{n,r}-\theta_0\right\|\\ &\longrightarrow 0 \text{ a.s. as } n\to\infty.\end{aligned}$$

This, along with (76), implies that, as $n\to\infty$

$$\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)=-\left[\nabla^2 Q_n(\theta_0)\right]^{-1}\sqrt{n}\frac{\partial Q_n(\theta_0)}{\partial\theta}+o_{\mathbb{P}}(1).$$

From Lemma A.3 and Lemma A.4, we obtain that $\sqrt{n}(\hat{\theta}_n-\theta_0)$ has a limiting normal distribution with mean 0 and covariance matrix $J^{-1}IJ^{-1}$. □

A.5. Proofs of Theorem 3.10

The proof of Theorem 3.10 is based on a series of lemmas.

Consider the regression of Υ_t on $\Upsilon_{t-1},\dots,\Upsilon_{t-r}$ defined by

$$\Upsilon_t=\sum_{i=1}^{r}\Phi_{r,i}\Upsilon_{t-i}+u_{r,t},\qquad u_{r,t}\perp\left\{\Upsilon_{t-1}\cdots\Upsilon_{t-r}\right\}.\tag{77}$$

If $\Upsilon_1,\dots,\Upsilon_n$ were observed, the least squares estimators of $\underline{\mathbf{\Phi}}_r=(\Phi_{r,1}\cdots\Phi_{r,r})$ and $\Sigma_{u_r}=\operatorname{Var}(u_{r,t})$ would be given by

$$\underline{\breve{\mathbf{\Phi}}}_r=\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r}\hat{\Sigma}^{-1}_{\underline{\Upsilon}_r}\qquad\text{and}\qquad\hat{\Sigma}_{\breve{u}_r}=\frac{1}{n}\sum_{t=1}^{n}\left(\Upsilon_t-\underline{\breve{\mathbf{\Phi}}}_r\underline{\Upsilon}_{r,t}\right)\left(\Upsilon_t-\underline{\breve{\mathbf{\Phi}}}_r\underline{\Upsilon}_{r,t}\right)'$$

where $\underline{\Upsilon}_{r,t}=(\Upsilon'_{t-1}\cdots\Upsilon'_{t-r})'$,

$$\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r}=\frac{1}{n}\sum_{t=1}^{n}\Upsilon_t\underline{\Upsilon}'_{r,t},\qquad\hat{\Sigma}_{\underline{\Upsilon}_r}=\frac{1}{n}\sum_{t=1}^{n}\underline{\Upsilon}_{r,t}\underline{\Upsilon}'_{r,t},$$

with by convention $\Upsilon_t=0$ when $t\leq 0$, and assuming $\hat{\Sigma}_{\underline{\Upsilon}_r}$ is non singular (which holds true asymptotically).

Actually, we just observe $X_1,\dots,X_n$. The residuals $\hat\epsilon_t := e_t(\hat\theta_n)$ are then available for $t=1,\dots,n$ and the vectors $\hat\Upsilon_t$ obtained by replacing θ_0 by $\hat\theta_n$ in (20) are available for $t=1,\dots,n$. We therefore define the least squares estimators of $\underline{\mathbf{\Phi}}_r = (\Phi_{r,1}\cdots\Phi_{r,r})$ and $\Sigma_{u_r}=\mathrm{Var}(u_{r,t})$ by

$$\hat{\underline{\mathbf{\Phi}}}_r = \hat\Sigma_{\hat\Upsilon,\hat{\underline{\Upsilon}}_r}\hat\Sigma^{-1}_{\hat{\underline{\Upsilon}}_r} \qquad \text{and} \qquad \hat\Sigma_{\hat u_r} = \frac{1}{n}\sum_{t=1}^n\left(\hat\Upsilon_t - \hat{\underline{\mathbf{\Phi}}}_r\hat{\underline{\Upsilon}}_{r,t}\right)\left(\hat\Upsilon_t - \hat{\underline{\mathbf{\Phi}}}_r\hat{\underline{\Upsilon}}_{r,t}\right)'$$

where $\hat{\underline{\Upsilon}}_{r,t} = (\hat\Upsilon'_{t-1}\cdots\hat\Upsilon'_{t-r})'$,

$$\hat\Sigma_{\hat\Upsilon,\hat{\underline{\Upsilon}}_r} = \frac{1}{n}\sum_{t=1}^n\hat\Upsilon_t\hat{\underline{\Upsilon}}'_{r,t}, \qquad \hat\Sigma_{\hat{\underline{\Upsilon}}_r} = \frac{1}{n}\sum_{t=1}^n\hat{\underline{\Upsilon}}_{r,t}\hat{\underline{\Upsilon}}'_{r,t},$$

with by convention $\hat\Upsilon_t=0$ when $t\le 0$, and assuming $\hat\Sigma_{\hat{\underline{\Upsilon}}_r}$ is non singular (which holds true asymptotically).

In the sequel, we use the multiplicative matrix norm defined by

$$\|A\| = \sup_{\|x\|\le 1}\|Ax\| = \varrho^{1/2}(A'\bar A), \tag{78}$$

where A is a $\mathbb{C}^{d_1\times d_2}$ matrix, $\|x\|^2 = x'\bar x$ is the Euclidean norm of the vector $x\in\mathbb{C}^{d_2\times 1}$, and $\varrho(\cdot)$ denotes the spectral radius. This norm satisfies

$$\|A\|^2 \le \sum_{i,j} a_{i,j}^2, \text{ when } A \text{ is a } \mathbb{R}^{d_1\times d_2} \text{ matrix} \tag{79}$$

with obvious notations. This choice of the norm is crucial for the following lemma to hold (with e.g. the Euclidean norm, this result is not valid). Let

$$\Sigma_{\Upsilon,\underline{\Upsilon}_r} = \mathbb{E}\Upsilon_t\underline{\Upsilon}'_{r,t}, \quad \Sigma_\Upsilon = \mathbb{E}\Upsilon_t\Upsilon'_t, \quad \Sigma_{\underline{\Upsilon}_r} = \mathbb{E}\underline{\Upsilon}_{r,t}\underline{\Upsilon}'_{r,t}, \quad \hat\Sigma_{\hat\Upsilon} = \frac{1}{n}\sum_{t=1}^n\hat\Upsilon_t\hat\Upsilon'_t.$$

In the sequel, C and ρ denote generic constant such as $K>0$ and $\rho\in(0,1)$, whose exact values are unimportant.

Lemma A.6. *Under the assumptions of Theorem 3.10,*

$$\sup_{r\ge 1}\max\left\{\left\|\Sigma_{\Upsilon,\underline{\Upsilon}_r}\right\|,\left\|\Sigma_{\underline{\Upsilon}_r}\right\|,\left\|\Sigma^{-1}_{\underline{\Upsilon}_r}\right\|\right\}<\infty.$$

Proof. The proof is an extension of Section 5.2 of [25]. We readily have

$$\|\Sigma_{\underline{\Upsilon}_r}x\| \le \|\Sigma_{\underline{\Upsilon}_{r+1}}(x',0'_{(p+q)K})'\| \quad \text{and} \quad \|\Sigma_{\underline{\Upsilon}_r}x\| \le \|\Sigma_{\underline{\Upsilon}_{r+1}}(0'_{(p+q)K},x')'\|$$

for any $x\in\mathbb{R}^{K(p+q)r}$ and $0_{(p+q)K}=(0,\dots,0)'\in\mathbb{R}^{(p+q)K}$. Therefore

$$0<\|\mathrm{Var}(\Upsilon_t)\| = \left\|\Sigma_{\underline{\Upsilon}_1}\right\| \le \left\|\Sigma_{\underline{\Upsilon}_2}\right\| \le \cdots$$

and

$$\left\|\Sigma_{\Upsilon,\underline{\Upsilon}_r}\right\| \le \left\|\Sigma_{\underline{\Upsilon}_{r+1}}\right\|,$$

so that it suffices to prove that $\sup_{r\geq 1}\left\|\Sigma_{\underline{\Upsilon}_r}\right\|$ and $\sup_{r\geq 1}\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|$ are finite to prove the result. Let us write matrix $\Sigma_{\underline{\Upsilon}_r}$ in blockwise form

$$\Sigma_{\underline{\Upsilon}_r} = [C(i-j)]_{i,j=1,\dots,r}\,,\quad C(k)=\mathbb{E}(\Upsilon_0\Upsilon_k')\in\mathbb{R}^{K(p+q)\times K(p+q)},\ k\in\mathbb{Z}.$$

Let now $f:\mathbb{R}\longrightarrow\mathbb{C}^{K(p+q)\times K(p+q)}$ be the spectral density of $(\Upsilon_t)_{t\in\mathbb{Z}}$ defined by

$$f(\omega)=\frac{1}{2\pi}\sum_{k=-\infty}^{\infty}C(k)e^{i\omega k},\quad \omega\in\mathbb{R}.$$

A direct consequence of (20) and Lemma A.2 is that $f(\omega)$ is absolutely summable, and that $\sup_{\omega\in\mathbb{R}}\|f(\omega)\|<+\infty$, for any norm $\|.\|$ on $\mathbb{C}^{K(p+q)\times K(p+q)}$ (in particular, one which is independent from $r\geq 1$). Another consequence is that one has the inversion formula

$$C(k)=\int_{-\pi}^{\pi}f(x)e^{-ikx}dx,\quad \forall k\in\mathbb{Z}. \tag{80}$$

Last, it is easy to check that $f(\omega)$ is an hermitian matrix for all $\omega\in\mathbb{R}$, i.e. $\overline{f(\omega)}=f(\omega)'$, where $\bar{z}$ is the conjugate of any vector or matrix z with entries in $\mathbb{C}$. Let then $\delta^{(r)}=\left(\delta_1^{(r)'},\dots,\delta_r^{(r)'}\right)\in\mathbb{R}^{rK(p+q)\times 1}$ be an eigenvector for $\Sigma_{\underline{\Upsilon}_r}$, with $\delta_j^{(r)}\in\mathbb{R}^{K(p+q)\times 1}$, $j=1,\dots,r$, such that $\|\delta^{(r)}\|=1$ and

$$\delta^{(r)'}\Sigma_{\underline{\Upsilon}_r}\delta^{(r)}=\|\Sigma_{\underline{\Upsilon}_r}\|=\varrho\left(\Sigma_{\underline{\Upsilon}_r}\right), \tag{81}$$

where $\|\Sigma_{\underline{\Upsilon}_r}\|$ is the norm of matrix $\Sigma_{\underline{\Upsilon}_r}$ defined in (78). One then checks that

$$\delta^{(r)'}\Sigma_{\underline{\Upsilon}_r}\delta^{(r)}=\sum_{i,j=1}^{r}\delta_i^{(r)'}C(i-j)\delta_j^{(r)}=\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{m=1}^{r}\delta_m^{(r)}e^{i(m-1)x}\right)'f(x)\overline{\left(\sum_{m=1}^{r}\delta_m^{(r)}e^{i(m-1)x}\right)}dx, \tag{82}$$

the last equality a direct consequence of (80). $f(x)$ being hermitian, $(X,Y)\in\mathbb{C}^{K(p+q)\times 1}\times\mathbb{C}^{K(p+q)\times 1}\mapsto X'f(x)\bar{Y}$ defines a semi definite non negative bilinear form, hence one has for all $x\in\mathbb{R}$ and $X\in\mathbb{C}^{K(p+q)\times 1}$:

$$0\leq X'f(x)\bar{X}\leq\|f(x)\|.X'\bar{X}\leq\sup_{\omega\in\mathbb{R}}\|f(\omega)\|.X'\bar{X}.$$

Let us point out that $\sup_{\omega\in\mathbb{R}}\|f(\omega)\|$ is a quantity which is independent from $r\geq 1$. We deduce from (82) and the previous inequality that

$$\delta^{(r)'}\Sigma_{\underline{\Upsilon}_r}\delta^{(r)}\leq\sup_{\omega\in\mathbb{R}}\|f(\omega)\|\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{m=1}^{r}\delta_m^{(r)}e^{i(m-1)x}\right)'\overline{\left(\sum_{m=1}^{r}\delta_m^{(r)}e^{i(m-1)x}\right)}dx. \tag{83}$$

A short computation yields that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left(\sum_{m=1}^{r}\delta_m^{(r)}e^{i(m-1)x}\right)'\overline{\left(\sum_{m=1}^{r}\delta_m^{(r)}e^{i(m-1)x}\right)}dx=\sum_{m=1}^{r}\delta_m^{(r)'}\delta_m^{(r)}=\|\delta^{(r)}\|^2=1,$$

which, coupled with (81) and (83), yields that $\|\Sigma_{\underline{\Upsilon}_r}\|\leq\sup_{\omega\in\mathbb{R}}\|f(\omega)\|<+\infty$, an upper bound independent from $r\geq 1$. By similar arguments, the smallest eigenvalue of $\Sigma_{\underline{\Upsilon}_r}$ is greater than a positive constant independent of r. Using the fact that $\|\Sigma_{\underline{\Upsilon}_r}^{-1}\|$ is equal to the inverse of the smallest eigenvalue of $\Sigma_{\underline{\Upsilon}_r}$, the proof is completed. □

The following lemma is necessary in the sequel.

Lemma A.7. *Let us suppose that* **(A1)** *and that Stationarity condition* (8) *for* $\nu = 6$

$$(\mathbf{A6}) \limsup_{t\to\infty} \frac{1}{t} \ln \mathbb{E}\left(\sup_{\theta\in\Theta} \left|\left| \prod_{i=1}^{t} \Phi(\Delta_i, \theta) \right|\right|^{32} \right) < 0, \quad \limsup_{t\to\infty} \frac{1}{t} \ln \mathbb{E}\left(\left|\left| \prod_{i=1}^{t} \Psi(\Delta_i) \right|\right|^{32} \right) < 0$$

hold. We assume that $\epsilon_t \in L^{4\nu+8}$. *Sequences* $(\epsilon_t(\theta))_{t\in\mathbb{Z}}$ *and* $(e_t(\theta))_{t\in\mathbb{Z}}$ *satisfy*

1. $||\sup_{\theta\in\Theta} |\epsilon_0(\theta)|||_{16} < +\infty$ *and* $\sup_{t\geq 0} ||\sup_{\theta\in\Theta} |e_t(\theta)|||_{16} < +\infty$,
2. $||\sup_{\theta\in\Theta} |\epsilon_t(\theta) - e_t(\theta)|||_4$ *tends to* 0 *exponentially fast as* $t \to \infty$,
3. *For all* $\alpha > 0$, $t^\alpha \sup_{\theta\in\Theta} |\epsilon_t(\theta) - e_t(\theta)| \longrightarrow 0$ *a.s. as* $t \to \infty$,
4. *For all* $j = 1, 2, 3$, $\left|\left|\sup_{\theta\in\Theta} ||\nabla^j \epsilon_0(\theta)||\right|\right|_{16} < +\infty$, $\sup_{t\geq 0} \left|\left|\sup_{\theta\in\Theta} ||\nabla^j e_t(\theta)||\right|\right|_{16} < +\infty$ *and one has* $t^\alpha ||\sup_{\theta\in\Theta} ||\nabla(e_t - \epsilon_t)(\theta)||||_{16/5} \longrightarrow 0$, *as* $t \to \infty$ *for all* $\alpha > 0$.

Proof of Lemma A.7 is similar to the proofs of Lemmas 3.3 and 3.4. □

Denote by $\Upsilon_t(i)$ the i-th element of Υ_t.

Lemma A.8. *Let* (ϵ_t) *be a sequence of centered and uncorrelated variables, with* $\mathbb{E}|\epsilon_t|^{8+4\nu} < \infty$ *and* $\sum_{h=0}^{\infty} [\alpha_\epsilon(h)]^{\nu/(2+\nu)} < \infty$ *for some* $\nu > 0$. *Then there exits a finite constant* C_1 *such that for* $m_1, m_2 = 1, \ldots, (p+q)K$ *and all* $s \in \mathbb{Z}$,

$$\sum_{h=-\infty}^{\infty} |Cov\{\Upsilon_1(m_1)\Upsilon_{1+s}(m_2), \Upsilon_{1+h}(m_1)\Upsilon_{1+s+h}(m_2)\}| < C_1.$$

Proof. Recall that

$$\frac{\partial \epsilon_t(\theta_0)}{\partial \theta_l} = \sum_{i=0}^{\infty} c_{i,l}(\theta_0, \Delta_{t-1}, \ldots, \Delta_{t-i})\epsilon_{t-i}, \text{ for } l = 1, \ldots, (p+q)K, \tag{84}$$

where $c_i(\theta_0, \Delta_{t-1}, \ldots, \Delta_{t-i})$ is defined by (10) and $c_{i,l}(\theta_0, \Delta_{t-1}, \ldots, \Delta_{t-i}) = \partial c_i(\theta_0, \Delta_{t-1}, \ldots, \Delta_{t-i})/\partial\theta_l$, and with the following upper bound holding thanks to (14):

$$\mathbb{E} \sup_{\theta\in\Theta} (c_i(\theta, \Delta_{t-1}, \ldots, \Delta_{t-i}))^2 \leq C\rho^i \text{ and } \mathbb{E} \sup_{\theta\in\Theta} (c_{i,l}(\theta, \Delta_{t-1}, \ldots, \Delta_{t-i}))^2 \leq C\rho^i, \quad \forall i.$$

Let

$$\begin{aligned} \gamma_{i,j,i',j',s,h}(m_1, m_2)(\theta_0) = \; & \mathbb{E}\left[c_{i,m_1}(\theta_0, \Delta_{t-1}, \ldots, \Delta_{t-i}) c_{j,m_2}(\theta_0, \Delta_{t+s-1}, \ldots, \Delta_{t+s-j})\right. \\ & \left.\times c_{i',m_1}(\theta_0, \Delta_{t+h-1}, \ldots, \Delta_{t+h-i'}) c_{j',m_2}(\theta_0, \Delta_{t+s+h-1}, \ldots, \Delta_{t+s+h-j'})\right] \\ & \times \mathrm{Cov}\left(\epsilon_t\epsilon_{t-i}\epsilon_{t+s}\epsilon_{t+s-j}, \epsilon_{t+h}\epsilon_{t+h-i'}\epsilon_{t+s+h}\epsilon_{t+s+h-j'}\right) \\ & + \mathrm{Cov}\left(c_{i,m_1}(\theta_0, \Delta_{t-1}, \ldots, \Delta_{t-i}) c_{j,m_2}(\theta_0, \Delta_{t+s-1}, \ldots, \Delta_{t+s-j}),\right. \\ & \left. c_{i',m_1}(\theta_0, \Delta_{t+h-1}, \ldots, \Delta_{t+h-i'}) c_{j',m_2}(\theta_0, \Delta_{t+s+h-1}, \ldots, \Delta_{t+s+h-j'})\right) \\ & \times \mathbb{E}\left[\epsilon_t\epsilon_{t-i}\epsilon_{t+s}\epsilon_{t+s-j}\right] \mathbb{E}\left[\epsilon_{t+h}\epsilon_{t+h-i'}\epsilon_{t+s+h}\epsilon_{t+s+h-j'}\right]. \end{aligned} \tag{85}$$

The Cauchy Schwarz inequality implies that

$$\begin{aligned} & \left|\mathbb{E}[c_{i,m_1}(\theta_0, \Delta_{t-1}, \ldots, \Delta_{t-i}) c_{j,m_2}(\theta_0, \Delta_{t+s-1}, \ldots, \Delta_{t+s-j})\right. \\ & \left.\times c_{i',m_1}(\theta_0, \Delta_{t+h-1}, \ldots, \Delta_{t+h-i'}) c_{j',m_2}(\theta_0, \Delta_{t+s+h-1}, \ldots, \Delta_{t+s+h-j'})]\right| \leq C\rho^{i+j+i'+j'} \end{aligned} \tag{86}$$

In view of (84) and (85), we have

$$\begin{aligned}
&\sum_{h=-\infty}^{\infty} \operatorname{Cov}\left\{\Upsilon_1(m_1)\Upsilon_{1+s}(m_2), \Upsilon_{1+h}(m_1)\Upsilon_{1+s+h}(m_2)\right\} \\
&= 16 \sum_{h=-\infty}^{\infty} \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty} \gamma_{i,j,i',j',s,h}(m_1,m_2)(\theta_0).
\end{aligned}$$

Without loss of generality, we can take the supremum over the integers $s > 0$, and consider the sum for positive h. Let $m_0 = m_1 \wedge m_2$ and $Y_{t,h_1} = \epsilon_t\epsilon_{t-h_1} - \mathbb{E}(\epsilon_t\epsilon_{t-h_1})$. We first suppose that $h \geq 0$. It follows that

$$\begin{aligned}
&\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty} \left|\operatorname{Cov}\left(c_{i,m_1}(\theta_0,\Delta_{t-1},\dots,\Delta_{t-i})c_{j,m_2}(\theta_0,\Delta_{t+s-1},\dots,\Delta_{t+s-j}),\right.\right. \\
&\qquad \left.\left. c_{i',m_1}(\theta_0,\Delta_{t+h-1},\dots,\Delta_{t+h-i'})c_{j',m_2}(\theta_0,\Delta_{t+s+h-1},\dots,\Delta_{t+s+h-j'})\right)\right| \\
&\leq v_1+v_2+v_3+v_4+v_5,
\end{aligned}$$

where

$$\begin{aligned}
v_1 = v_1(h) &= \sum_{i>[h/2]}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty} \left|\operatorname{Cov}\left(\mathbf{c}_{i,m_1}^{t}\mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h}\mathbf{c}_{j',m_2}^{t+h+s}\right)\right|, \\
v_2 = v_2(h) &= \sum_{i=0}^{\infty}\sum_{j>[h/2]}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty} \left|\operatorname{Cov}\left(\mathbf{c}_{i,m_1}^{t}\mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h}\mathbf{c}_{j',m_2}^{t+h+s}\right)\right| \\
v_3 = v_3(h) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'>[h/2]}\sum_{j'=0}^{\infty} \left|\operatorname{Cov}\left(\mathbf{c}_{i,m_1}^{t}\mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h}\mathbf{c}_{j',m_2}^{t+h+s}\right)\right|, \\
v_4 = v_4(h) &= \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'>[h/2]} \left|\operatorname{Cov}\left(\mathbf{c}_{i,m_1}^{t}\mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h}\mathbf{c}_{j',m_2}^{t+h+s}\right)\right|, \\
v_5 = v_5(h) &= \sum_{i=0}^{[h/2]}\sum_{j=0}^{[h/2]}\sum_{i'=0}^{[h/2]}\sum_{j'=0}^{[h/2]} \left|\operatorname{Cov}\left(\mathbf{c}_{i,m_1}^{t}\mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h}\mathbf{c}_{j',m_2}^{t+h+s}\right)\right|,
\end{aligned}$$

where

$$\mathbf{c}_{i_1,m}^{t} = c_{i_1,m}(\theta_0,\Delta_{t-1},\dots,\Delta_{t-i_1}).$$

One immediate remark is that $\mathbf{c}_{i_1,m}^{t}$ is measurable with respect to Δ_r, $r \in \{t-1,...,t-i_1\}$. Since

$$\left|\operatorname{Cov}\left(\mathbf{c}_{i,m_1}^{t}\mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h}\mathbf{c}_{j',m_2}^{t+h+s}\right)\right| \leq C\rho^{i+i'+j+j'},$$

we have

$$v_1 = \sum_{i>[h/2]}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty} \left|\operatorname{Cov}\left(\mathbf{c}_{i,m_1}^{t}\mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h}\mathbf{c}_{j',m_2}^{t+h+s}\right)\right| \leq \kappa_1\rho^{h/2},$$

for some positive constant κ_1. Using the same arguments we obtain that v_i, $i = 2, 3, 4$ are bounded by $\kappa_i \rho^{h/2}$. The $\alpha-$mixing property (see Theorem 14.1 in [13], p. 210) and Lemmas A.1 and A.7, entail that

$$\begin{aligned} v_5 &= \sum_{i=0}^{[h/2]} \sum_{j=0}^{[h/2]} \sum_{i'=0}^{[h/2]} \sum_{j'=0}^{[h/2]} \left| \operatorname{Cov}\left(\mathbf{c}_{i,m_1}^{t} \mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h} \mathbf{c}_{j',m_2}^{t+h+s} \right) \right| \\ &\leq \sum_{k=1}^{4} \sum_{(i,j,i',j') \in \mathcal{C}_k} \kappa_6 \left\| \mathbf{c}_{i,m_1}^{t} \mathbf{c}_{j,m_2}^{t+s} \right\|_{2+\nu} \left\| \mathbf{c}_{i',m_1}^{t+h} \mathbf{c}_{j',m_2}^{t+h+s} \right\|_{2+\nu} \left\{ \alpha\left(\mathbf{c}_{i,m_1}^{t} \mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h} \mathbf{c}_{j',m_2}^{t+h+s} \right) \right\}^{\nu/(2+\nu)}, \end{aligned}$$

where $\alpha(U, V)$ denotes the strong mixing coefficient between the $\sigma-$field generated by the random variable U and that generated by V and where

$$\begin{aligned} \mathcal{C}_1 = \mathcal{C}_1(h) &= \left\{ (i, j, i', j') \in \{0, 1, \ldots, [h/2]\}^4 : i \geq j - s,\ j' \leq i' + s \right\}, \\ \mathcal{C}_2 = \mathcal{C}_2(h) &= \left\{ (i, j, i', j') \in \{0, 1, \ldots, [h/2]\}^4 : i \geq j - s,\ j' \geq i' + s \right\}, \\ \mathcal{C}_3 = \mathcal{C}_3(h) &= \left\{ (i, j, i', j') \in \{0, 1, \ldots, [h/2]\}^4 : i \leq j - s,\ j' \leq i' + s \right\}, \\ \mathcal{C}_4 = \mathcal{C}_4(h) &= \left\{ (i, j, i', j') \in \{0, 1, \ldots, [h/2]\}^4 : i \leq j - s,\ j' \geq i' + s \right\}. \end{aligned}$$

One checks easily that $\mathbf{c}_{i,m_1}^{t} \mathbf{c}_{j,m_2}^{t+s}$ and $\mathbf{c}_{i',m_1}^{t+h} \mathbf{c}_{j',m_2}^{t+h+s}$ are respectively measurable with respect to Δ_r, $r \in \{t - i, ..., t + s - 1\}$ and Δ_r, $r \in \{t - i' + h, ..., t + h + s - 1\}$. For $(i, j, i', j') \in \mathcal{C}_1$, we have $t - i \leq t + s - j$, $t + h - i' \leq t + h + s - j'$ and we thus deduce that

$$\begin{aligned} \left| \alpha\left(\mathbf{c}_{i,m_1}^{t} \mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h} \mathbf{c}_{j',m_2}^{t+h+s} \right) \right| &\leq \alpha_\Delta\left(h - i' - s + 1 \right), \quad \forall h \geq i' + s - 1, \\ \left| \alpha\left(\mathbf{c}_{i,m_1}^{t} \mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h} \mathbf{c}_{j',m_2}^{t+h+s} \right) \right| &\leq \alpha_\Delta\left(-i - h - s + 1 \right), \quad \forall h \leq -i - s + 1, \\ \left| \alpha\left(\mathbf{c}_{i,m_1}^{t} \mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h} \mathbf{c}_{j',m_2}^{t+h+s} \right) \right| &\leq \alpha_\Delta(0) \leq 1/4, \quad \forall h = -i - s + 1, \ldots, i' + s - 1. \end{aligned}$$

Note also that, by the Hölder inequality,

$$\left\| \mathbf{c}_{i,m_1}^{t} \mathbf{c}_{j,m_2}^{t+s} \right\|_{2+\nu} \leq \left\| \mathbf{c}_{i,m_1}^{t} \right\|_{4+2\nu} \left\| \mathbf{c}_{j,m_2}^{t+s} \right\|_{4+2\nu} \leq C \rho^{i+j}.$$

Therefore

$$\begin{aligned} &\sum_{h=0}^{\infty} \sum_{(i,j,i',j') \in \mathcal{C}_1} \left\| \mathbf{c}_{i,m_1}^{t} \mathbf{c}_{j,m_2}^{t+s} \right\|_{2+\nu} \left\| \mathbf{c}_{i',m_1}^{t+h} \mathbf{c}_{j',m_2}^{t+h+s} \right\|_{2+\nu} \left\{ \alpha\left(\mathbf{c}_{i,m_1}^{t} \mathbf{c}_{j,m_2}^{t+s}, \mathbf{c}_{i',m_1}^{t+h} \mathbf{c}_{j',m_2}^{t+h+s} \right) \right\}^{\nu/(2+\nu)}, \\ \leq\ & C^2 \sum_{i,j,i',j'=0}^{\infty} \rho^{i+j+i'+j'} \left(i' + 2s - 1 + i + \sum_{r=0}^{\infty} \alpha_\Delta^{\nu/(2+\nu)}(r) \right) < \infty. \end{aligned}$$

Continuing in this way, we obtain that $\sum_{h=0}^{\infty} v_5(h) < \infty$. It follows that

$$\begin{aligned} &\sum_{h=0}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \left| \operatorname{Cov}\left(c_{i,m_1}(\theta_0, \Delta_{t-1}, \ldots, \Delta_{t-i}) c_{j,m_2}(\theta_0, \Delta_{t+s-1}, \ldots, \Delta_{t+s-j}), \right.\right. \\ &\left.\left. c_{i',m_1}(\theta_0, \Delta_{t+h-1}, \ldots, \Delta_{t+h-i'}) c_{j',m_2}(\theta_0, \Delta_{t+s+h-1}, \ldots, \Delta_{t+s+h-j'}) \right) \right| \\ \leq\ & \sum_{h=0}^{\infty} \sum_{i=1}^{5} v_i(h) < \infty. \end{aligned} \tag{87}$$

The same bounds clearly holds for

$$\sum_{h=-\infty}^{0}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty}\Big|\mathrm{Cov}\left(c_{i,m_1}(\theta_0,\Delta_{t-1},\dots,\Delta_{t-i})c_{j,m_2}(\theta_0,\Delta_{t+s-1},\dots,\Delta_{t+s-j}),\right.$$
$$\left. c_{i',m_1}(\theta_0,\Delta_{t+h-1},\dots,\Delta_{t+h-i'})c_{j',m_2}(\theta_0,\Delta_{t+s+h-1},\dots,\Delta_{t+s+h-j'})\right)\Big|<\infty,$$

which shows that

$$\sum_{h=-\infty}^{\infty}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty}\Big|\mathrm{Cov}\left(c_{i,m_1}(\theta_0,\Delta_{t-1},\dots,\Delta_{t-i})c_{j,m_2}(\theta_0,\Delta_{t+s-1},\dots,\Delta_{t+s-j}),\right.$$
$$\left. c_{i',m_1}(\theta_0,\Delta_{t+h-1},\dots,\Delta_{t+h-i'})c_{j',m_2}(\theta_0,\Delta_{t+s+h-1},\dots,\Delta_{t+s+h-j'})\right)\Big|<\infty.$$

A slight extension of Corollary A.3 in [24] shows that

$$\sum_{h=-\infty}^{\infty}\left|\mathrm{Cov}\left(Y_{1,i}Y_{1+s,j},Y_{1+h,i'}Y_{1+s+h,j'}\right)\right|<\infty. \tag{88}$$

Because, by Cauchy Schwarz inequality

$$\left|\mathbb{E}\left[\epsilon_t\epsilon_{t-i}\epsilon_{t+s}\epsilon_{t+s-j}\right]\right|\leq\mathbb{E}\left|\epsilon_t\right|^4<\infty$$

by the assumption that $\mathbb{E}\left|\epsilon_t\right|^{8+4\nu}<\infty$ and in view of (86) it follows that

$$\begin{aligned}
&\sum_{h=-\infty}^{\infty}\left|\mathrm{Cov}\left\{\Upsilon_1(m_1)\Upsilon_{1+s}(m_2),\Upsilon_{1+h}(m_1)\Upsilon_{1+s+h}(m_2)\right\}\right|\\
\leq\;&\kappa\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty}\rho^{i+j+i'+j'}\sum_{h=-\infty}^{\infty}\left|\mathrm{Cov}\left(Y_{1,i}Y_{1+s,j},Y_{1+h,i'}Y_{1+s+h,j'}\right)\right|\\
&+\kappa'\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty}\sum_{h=-\infty}^{\infty}\Big|\mathrm{Cov}\left(c_{i,m_1}(\theta_0,\Delta_{t-1},\dots,\Delta_{t-i})c_{j,m_2}(\theta_0,\Delta_{t+s-1},\dots,\Delta_{t+s-j}),\right.\\
&\left. c_{i',m_1}(\theta_0,\Delta_{t+h-1},\dots,\Delta_{t+h-i'})c_{j',m_2}(\theta_0,\Delta_{t+s+h-1},\dots,\Delta_{t+s+h-j'})\right)\Big|
\end{aligned}$$

The conclusion follows from (87) and (88). □

Let $\hat{\Sigma}_{\Upsilon}$ be the matrix obtained by replacing $\hat{\Upsilon}_t$ by Υ_t in $\hat{\Sigma}_{\hat{\Upsilon}}$.

Lemma A.9. *Under the assumptions of Theorem 3.10,* $\sqrt{r}\|\hat{\Sigma}_{\underline{\Upsilon}_r}-\Sigma_{\underline{\Upsilon}_r}\|$, $\sqrt{r}\|\hat{\Sigma}_{\Upsilon}-\Sigma_{\Upsilon}\|$, *and* $\sqrt{r}\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r}-\Sigma_{\Upsilon,\underline{\Upsilon}_r}\|$ *tend to zero in probability as* $n\to\infty$ *when* $r=\mathrm{o}(n^{1/3})$.

Proof. For $1\leq m_1,m_2\leq K(p+q)$ and $1\leq r_1,r_2\leq r$, the element of the $\{(r_1-1)(p+q)K+m_1\}$-th row and $\{(r_2-1)(p+q)K+m_2\}$-th column of $\hat{\Sigma}_{\underline{\Upsilon}_r}$ is of the form $n^{-1}\sum_{t=1}^{n}Z_t$ where $Z_t:=Z_{t,r_1,r_2}(m_1,m_2)=\Upsilon_{t-r_1}(m_1)\Upsilon_{t-r_2}(m_2)$. By stationarity of (Z_t), we have

$$\mathrm{Var}\left(\frac{1}{n}\sum_{t=1}^{n}Z_t\right)=\frac{1}{n^2}\sum_{h=-n+1}^{n-1}(n-|h|)\,\mathrm{Cov}\left(Z_t,Z_{t-h}\right)\leq\frac{C_1}{n}, \tag{89}$$

where, by Lemma A.8, C_1 is a constant independent of r_1, r_2, m_1, m_2 and r, n. Now using the Tchebychev inequality, we have

$$\forall \beta > 0, \quad \mathbb{P}\left\{\sqrt{r}\|\hat{\Sigma}_{\underline{\Upsilon}_r} - \Sigma_{\underline{\Upsilon}_r}\| > \beta\right\} \leq \frac{1}{\beta^2}\mathbb{E}\left\{r\|\hat{\Sigma}_{\underline{\Upsilon}_r} - \Sigma_{\underline{\Upsilon}_r}\|^2\right\}.$$

In view of (79) and (89) we have

$$\begin{aligned}
&\mathbb{E}\left\{r\|\hat{\Sigma}_{\Upsilon} - \Sigma_{\Upsilon}\|^2\right\} \leq \mathbb{E}\left\{r\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r} - \Sigma_{\Upsilon,\underline{\Upsilon}_r}\|^2\right\} \\
&\leq \mathbb{E}\left\{r\|\hat{\Sigma}_{\underline{\Upsilon}_r} - \Sigma_{\underline{\Upsilon}_r}\|^2\right\} \leq r\sum_{m_1,m_2=1}^{K(p+q)r} \mathrm{Var}\left(\frac{1}{n}\sum_{t=1}^{n} Z_t\right) \leq \frac{C_1 K^2(p+q)^2 r^3}{n} = \mathrm{o}(1)
\end{aligned}$$

as $n \to \infty$ when $r = \mathrm{o}(n^{1/3})$. Hence, when $r = \mathrm{o}(n^{1/3})$

$$\begin{aligned}
\sqrt{r}\|\hat{\Sigma}_{\underline{\Upsilon}_r} - \Sigma_{\underline{\Upsilon}_r}\| &= \mathrm{o}_{\mathbb{P}}(1), \\
\sqrt{r}\|\hat{\Sigma}_{\Upsilon} - \Sigma_{\Upsilon}\| &= \mathrm{o}_{\mathbb{P}}(1) \text{ and } \sqrt{r}\|\hat{\Sigma}_{\Upsilon,\underline{\Upsilon}_r} - \Sigma_{\Upsilon,\underline{\Upsilon}_r}\| = \mathrm{o}_{\mathbb{P}}(1).
\end{aligned}$$

The proof is complete. □

We now show that the previous lemma applies when Υ_t is replaced by $\hat{\Upsilon}_t$.

Lemma A.10. *Under the assumptions of Theorem 3.10,* $\sqrt{r}\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r} - \Sigma_{\underline{\Upsilon}_r}\|$, $\sqrt{r}\|\hat{\Sigma}_{\hat{\Upsilon}} - \Sigma_{\Upsilon}\|$, *and* $\sqrt{r}\|\hat{\Sigma}_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} - \Sigma_{\Upsilon,\underline{\Upsilon}_r}\|$ *tend to zero in probability as* $n \to \infty$ *when* $r = \mathrm{o}(n^{1/3})$.

Proof. We first show that the replacement of the unknown initial values $\{X_u,\ u \leq 0\}$ by zero is asymptotically unimportant. Let $\hat{\Sigma}_{\underline{\Upsilon}_{r,n}}$ be the matrix obtained by replacing $e_t(\hat{\theta}_n)$ by $\epsilon_t(\hat{\theta}_n)$ in $\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r}$. We start by evaluating $\mathbb{E}\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r} - \hat{\Sigma}_{\underline{\Upsilon}_{r,n}}\|^2$. We first note that

$$\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r} - \hat{\Sigma}_{\underline{\Upsilon}_{r,n}} = \left[\frac{1}{n}\sum_{t=1}^{n} a_{t-i,t-i',m_1,m_2}(\hat{\theta}_n)\right]$$

for $i, i' = 1, \dots, r$ and $m_1, m_2 = 1, \dots, K(p+q)$ and where

$$a_{t-i,t-i',m_1,m_2}(\hat{\theta}_n) = e_{t-i}(\hat{\theta}_n)e_{t-i'}(\hat{\theta}_n)\frac{\partial e_{t-i}(\hat{\theta}_n)}{\partial \theta_{m_1}}\frac{\partial e_{t-i'}(\hat{\theta}_n)}{\partial \theta_{m_2}} - \epsilon_{t-i}(\hat{\theta}_n)\epsilon_{t-i'}(\hat{\theta}_n)\frac{\partial \epsilon_{t-i}(\hat{\theta}_n)}{\partial \theta_{m_1}}\frac{\partial \epsilon_{t-i'}(\hat{\theta}_n)}{\partial \theta_{m_2}}.$$

Using (79), we have

$$\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r} - \hat{\Sigma}_{\underline{\Upsilon}_{r,n}}\|^2 \leq \sum_{i,i'=1}^{r}\sum_{m_1,m_2=1}^{K(p+q)}\left[\frac{1}{n}\sum_{t=1}^{n} a_{t-i,t-i',m_1,m_2}(\hat{\theta}_n)\right]^2.$$

We thus deduce the following L^2 estimate:

$$\begin{aligned}
\mathbb{E}\|\hat{\Sigma}_{\underline{\hat{\Upsilon}}_r} - \hat{\Sigma}_{\underline{\Upsilon}_{r,n}}\|^2 &\leq \sum_{i,i'=1}^{r}\sum_{m_1,m_2=1}^{K(p+q)}\left\|\frac{1}{n}\sum_{t=1}^{n} a_{t-i,t-i',m_1,m_2}(\hat{\theta}_n)\right\|_2^2 \\
&\leq \sum_{i,i'=1}^{r}\sum_{m_1,m_2=1}^{K(p+q)}\frac{1}{n}\sum_{t=1}^{n}\left\|a_{t-i,t-i',m_1,m_2}(\hat{\theta}_n)\right\|_2^2,
\end{aligned}$$

by Minkowski's inequality. Thanks to Hölder's inequality:

$$\left\|a_{t-i,t-i',m_1,m_2}(\hat{\theta}_n)\right\|_2 \leq \sum_{j=1}^{4} \mathcal{A}^j_{t-i,t-i',m_1,m_2}, \text{ with}$$

$$\begin{aligned}
\mathcal{A}^1_{t-i,t-i',m_1,m_2} &= \left\|\sup_{\theta\in\Theta}|e_{t-i}(\theta)-\epsilon_{t-i}(\theta)|\right\|_4 \sup_{t\geq 0}\left\|\sup_{\theta\in\Theta}|e_t(\theta)|\right\|_{12}\left(\sup_{t\geq 0}\left\|\sup_{\theta\in\Theta}\left\|\frac{\partial e_t(\theta)}{\partial\theta}\right\|\right\|_{12}\right)^2\\
\mathcal{A}^2_{t-i,t-i',m_1,m_2} &= \left\|\sup_{\theta\in\Theta}|\epsilon_t(\theta)|\right\|_{12}\left\|\sup_{\theta\in\Theta}|e_{t-i'}(\theta)-\epsilon_{t-i'}(\theta)|\right\|_4\left(\sup_{t\geq 0}\left\|\sup_{\theta\in\Theta}\left\|\frac{\partial e_t(\theta)}{\partial\theta}\right\|\right\|_{12}\right)^2\\
\mathcal{A}^3_{t-i,t-i',m_1,m_2} &= \left(\left\|\sup_{\theta\in\Theta}|\epsilon_t(\theta)|\right\|_{16}\right)^2\left\|\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}\left(e_{t-i}(\theta)-\epsilon_{t-i}(\theta)\right)\right\|\right\|_{16/5}\sup_{t\geq 0}\left\|\sup_{\theta\in\Theta}\left\|\frac{\partial e_t(\theta)}{\partial\theta}\right\|\right\|_{16}\\
\mathcal{A}^4_{t-i,t-i',m_1,m_2} &= \left(\left\|\sup_{\theta\in\Theta}|\epsilon_t(\theta)|\right\|_{16}\right)^2\left\|\sup_{\theta\in\Theta}\left\|\frac{\partial\epsilon_t(\theta)}{\partial\theta}\right\|\right\|_{16}\left\|\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}\left(e_{t-i'}(\theta)-\epsilon_{t-i'}(\theta)\right)\right\|\right\|_{16/5}.
\end{aligned}$$

We deal with $\mathcal{A}^1_{t-i,t-i',m_1,m_2}$ and $\mathcal{A}^2_{t-i,t-i',m_1,m_2}$, as $\mathcal{A}^3_{t-i,t-i',m_1,m_2}$ and $\mathcal{A}^4_{t-i,t-i',m_1,m_2}$ are dealt with similarly. In view of Lemma A.7, one has

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathcal{A}^1_{t-i,t-i',m_1,m_2} &\leq \kappa_1\frac{1}{n}\sum_{t=1}^{n}\left\|\sup_{\theta\in\Theta}|e_{t-i}(\theta)-\epsilon_{t-i}(\theta)|\right\|_4\\
&\leq \frac{\kappa_1}{n}\left(\sum_{t=1}^{n-r}\left\|\sup_{\theta\in\Theta}|e_t(\theta)-\epsilon_t(\theta)|\right\|_4 + r\left\|\sup_{\theta\in\Theta}|\epsilon_0(\theta)|\right\|_4\right) = \mathrm{O}\left(\frac{1}{n}+\frac{r}{n}\right) = \mathrm{O}\left(\frac{r}{n}\right),
\end{aligned}$$

independent from i, i', m_1 and m_2. Similarly, one has

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\mathcal{A}^3_{t-i,t-i',m_1,m_2} &\leq \kappa_3\frac{1}{n}\sum_{t=1}^{n}\left\|\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}\left(e_{t-i}(\theta)-\epsilon_{t-i}(\theta)\right)\right\|\right\|_{16/5}\\
&\leq \kappa_3\frac{1}{n}\left(\sum_{t=1}^{n-r}\left\|\sup_{\theta\in\Theta}\left\|\frac{\partial}{\partial\theta}\left(e_t(\theta)-\epsilon_t(\theta)\right)\right\|\right\|_{16/5}\right.\\
&\qquad\left.+r\left\|\sup_{\theta\in\Theta}\left\|\frac{\partial\epsilon_0(\theta)}{\partial\theta}\right\|\right\|_{16/5}\right) = \mathrm{O}\left(\frac{1}{n}+\frac{r}{n}\right) = \mathrm{O}\left(\frac{r}{n}\right),
\end{aligned}$$

because $\sum_{t=1}^{\infty}\left\|\sup_{\theta\in\Theta}\left\|\partial\left(e_t(\theta)-\epsilon_t(\theta)\right)/\partial\theta\right\|\right\|_{16/5}<\infty$ and $\left\|\sup_{\theta\in\Theta}\left\|\partial\epsilon_0(\theta)/\partial\theta\right\|\right\|_{16/5}<\infty$ (see Lemma A.7, Point 4). Gathering $\mathcal{A}^1_{t-i,t-i',m_1,m_2}$, $\mathcal{A}^2_{t-i,t-i',m_1,m_2}$, $\mathcal{A}^3_{t-i,t-i',m_1,m_2}$ and $\mathcal{A}^4_{t-i,t-i',m_1,m_2}$, we arrive at

$$\mathbb{E}\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}-\hat{\Sigma}_{\underline{\Upsilon}_{r,n}}\|^2 \leq \sum_{i,i'=1}^{r}\sum_{m_1,m_2=1}^{K(p+q)}\left(\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{4}\mathcal{A}^j_{t-i,t-i',m_1,m_2}\right)^2 = \mathrm{O}\left(r^2\left\{\frac{r}{n}\right\}^2\right) = \mathrm{O}\left(\frac{r^4}{n^2}\right).$$

We thus deduce that

$$\sqrt{r}\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}-\hat{\Sigma}_{\underline{\Upsilon}_{r,n}}\| = \mathrm{o}_{\mathbb{P}}(1), \text{ when } r=r(n)=\mathrm{o}\left(n^{2/5}\right). \tag{90}$$

We now prove that

$$\sqrt{r}\|\hat{\Sigma}_{\underline{\Upsilon}_{r,n}}-\hat{\Sigma}_{\underline{\Upsilon}_r}\|=\mathrm{o}_{\mathbb{P}}(1),\ \text{ when } r=r(n)=\mathrm{o}\left(n^{1/3}\right).$$

Taylor expansions around θ_0 yield

$$\left|\epsilon_t(\hat{\theta}_n)-\epsilon_t(\theta_0)\right|\leq r_t\left\|\hat{\theta}_n-\theta_0\right\|,\qquad\left|\frac{\partial\epsilon_t(\hat{\theta}_n)}{\partial\theta_m}-\frac{\partial\epsilon_t(\theta_0)}{\partial\theta_m}\right|\leq s_t(m)\left\|\hat{\theta}_n-\theta_0\right\| \tag{91}$$

with $r_t=\sup_{\theta\in\Theta}\|\partial\epsilon_t(\theta)/\partial\theta\|$, $s_t(m)=\sup_{\theta\in\Theta}\left\|\partial^2\epsilon_t(\theta)/\partial\theta\partial\theta_m\right\|$ where $m=m_1=m_2$. Define Z_t as in the proof of Lemma A.9, and let $Z_{t,n}$ be obtained by replacing $\Upsilon_t(m)$ by $\Upsilon_{t,n}(m)=\epsilon_t(\hat{\theta}_n)\partial\epsilon_t(\hat{\theta}_n)/\partial\theta_m$ in Z_t. Using (91), for $i,i'=1,\dots,r$ and $m_1,m_2=1,\dots,K(p+q)$, we have

$$\left|\epsilon_{t-i}(\hat{\theta}_n)\epsilon_{t-i'}(\hat{\theta}_n)\frac{\partial\epsilon_{t-i}(\hat{\theta}_n)}{\partial\theta_{m_1}}\frac{\partial\epsilon_{t-i'}(\hat{\theta}_n)}{\partial\theta_{m_2}}-\epsilon_{t-i}(\theta_0)\epsilon_{t-i'}(\theta_0)\frac{\partial\epsilon_{t-i}(\theta_0)}{\partial\theta_{m_1}}\frac{\partial\epsilon_{t-i'}(\theta_0)}{\partial\theta_{m_2}}\right|\leq\sum_{j=1}^{4}\mathcal{B}^j_{t-i,t-i',m_1,m_2}, \tag{92}$$

with

$$\begin{aligned}
\mathcal{B}^1_{t-i,t-i',m_1,m_2}&=r_{t-i}\left\|\hat{\theta}_n-\theta_0\right\|\sup_{\theta\in\Theta}|\epsilon_{t-i'}(\theta)|\sup_{\theta\in\Theta}\left|\frac{\partial\epsilon_{t-i}(\theta)}{\partial\theta_{m_1}}\right|\sup_{\theta\in\Theta}\left|\frac{\partial\epsilon_{t-i'}(\theta)}{\partial\theta_{m_2}}\right|\\
\mathcal{B}^2_{t-i,t-i',m_1,m_2}&=r_{t-i'}\left\|\hat{\theta}_n-\theta_0\right\|\sup_{\theta\in\Theta}|\epsilon_{t-i}(\theta)|\sup_{\theta\in\Theta}\left|\frac{\partial\epsilon_{t-i}(\theta)}{\partial\theta_{m_1}}\right|\sup_{\theta\in\Theta}\left|\frac{\partial\epsilon_{t-i'}(\theta)}{\partial\theta_{m_2}}\right|\\
\mathcal{B}^3_{t-i,t-i',m_1,m_2}&=s_{t-i}(m_1)\left\|\hat{\theta}_n-\theta_0\right\|\sup_{\theta\in\Theta}|\epsilon_{t-i}(\theta)|\sup_{\theta\in\Theta}|\epsilon_{t-i'}(\theta)|\sup_{\theta\in\Theta}\left|\frac{\partial\epsilon_{t-i'}(\theta)}{\partial\theta_{m_2}}\right|\\
\mathcal{B}^4_{t-i,t-i',m_1,m_2}&=s_{t-i'}(m_2)\left\|\hat{\theta}_n-\theta_0\right\|\sup_{\theta\in\Theta}|\epsilon_{t-i}(\theta)|\sup_{\theta\in\Theta}|\epsilon_{t-i'}(\theta)|\sup_{\theta\in\Theta}\left|\frac{\partial\epsilon_{t-i}(\theta)}{\partial\theta_{m_1}}\right|.
\end{aligned}$$

We deal with $\mathcal{B}^1_{t-i,t-i',m_1,m_2}$ and $\mathcal{B}^2_{t-i,t-i',m_1,m_2}$, as $\mathcal{B}^3_{t-i,t-i',m_1,m_2}$ and $\mathcal{B}^4_{t-i,t-i',m_1,m_2}$ are dealt with similarly. We note first that, for all $i=1,\dots,r$,

$$\begin{aligned}
\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|\epsilon_{t-i}(\theta)|^4&=\frac{1}{n}\sum_{t=1-i}^{n-i}\sup_{\theta\in\Theta}|\epsilon_t(\theta)|^4=\frac{1}{n}\sum_{t=1-i}^{0}\sup_{\theta\in\Theta}|\epsilon_t(\theta)|^4+\frac{1}{n}\sum_{t=1}^{n-i}\sup_{\theta\in\Theta}|\epsilon_t(\theta)|^4\\
&\leq\frac{r}{n}\frac{1}{r}\sum_{t=1-r}^{0}\sup_{\theta\in\Theta}|\epsilon_t(\theta)|^4+\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}|\epsilon_t(\theta)|^4\\
&=\left(\frac{r}{n}+1\right)\left(\left\|\sup_{\theta\in\Theta}|\epsilon_0(\theta)|\right\|_4^4+\mathrm{o}_{p.s.}(1)\right),
\end{aligned} \tag{93}$$

by ergodic Theorem. Similarly to (93), one has

$$\frac{1}{n}\sum_{t=1}^{n}\sup_{\theta\in\Theta}\left|\frac{\partial\epsilon_{t-i}(\theta)}{\partial\theta_m}\right|^4\leq\left(\frac{r}{n}+1\right)\left(\left\|\sup_{\theta\in\Theta}\left|\frac{\partial\epsilon_0(\theta)}{\partial\theta_m}\right|\right\|_4^4+\mathrm{o}_{p.s.}(1)\right). \tag{94}$$

By the Cauchy-Schwartz inequality and using (93) and (94), we have

$$\begin{aligned}
\sum_{i,i'=1}^{r}\sum_{m_1,m_2=1}^{K(p+q)}\frac{1}{n}\sum_{t=1}^{n}\mathcal{B}^1_{t-i,t-i',m_1,m_2}&\leq r^2\left\|\hat{\theta}_n-\theta_0\right\|\left(\frac{r}{n}+1\right)^3(\kappa_1+\mathrm{o}_{p.s.}(1))\\
&=r^2\left\|\hat{\theta}_n-\theta_0\right\|\mathrm{O}(1)\left(\kappa_1+\mathrm{o}_{p.s.}(1)\right),
\end{aligned}$$

when $r = \mathrm{o}\left(n^{1/3}\right)$ and for some constant $\kappa_1 > 0$. Similar inequalities hold for $\mathcal{B}^{j}_{t-i,t-i',m_1,m_2}$, for $j = 2, 3, 4$. We thus deduce from (79) and (92) that

$$r\|\hat{\Sigma}_{\underline{\Upsilon}_{r,n}} - \hat{\Sigma}_{\underline{\Upsilon}_r}\|^2 \quad \leq \quad r^3 \left\|\hat{\theta}_n - \theta_0\right\|^2 \mathrm{O}_{\mathbb{P}}(1). \tag{95}$$

Since $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ converges in distribution, a tightness argument yields $\left\|\hat{\theta}_n - \theta_0\right\| = \mathrm{O}_{\mathbb{P}}\left(n^{-1/2}\right)$ and hence from (95), we obtain for $r = \mathrm{o}(n^{1/3})$

$$\sqrt{r}\|\hat{\Sigma}_{\underline{\Upsilon}_{r,n}} - \hat{\Sigma}_{\underline{\Upsilon}_r}\| = \mathrm{o}_{\mathbb{P}}(1). \tag{96}$$

By Lemma A.9 , (90) and (96) show that $\sqrt{r}\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r} - \Sigma_{\underline{\Upsilon}_r}\| = \mathrm{o}_{\mathbb{P}}(1)$. The other results are obtained similarly. $\square$

Write $\underline{\mathbf{\Phi}}^*_r = (\Phi_1 \cdots \Phi_r)$ where the Φ_i's are defined by (22).

Lemma A.11. *Under the assumptions of Theorem 3.10,*

$$\sqrt{r}\,\|\underline{\mathbf{\Phi}}^*_r - \underline{\mathbf{\Phi}}_r\| \to 0,$$

as $r \to \infty$.

Proof. Recall that by (22) and (77)

$$\Upsilon_t \quad = \quad \underline{\mathbf{\Phi}}_r \underline{\Upsilon}_{r,t} + u_{r,t} = \underline{\mathbf{\Phi}}^*_r \underline{\Upsilon}_{r,t} + \sum_{i=r+1}^{\infty} \Phi_i \Upsilon_{t-i} + u_t := \underline{\mathbf{\Phi}}^*_r \underline{\Upsilon}_{r,t} + u^*_{r,t}.$$

Hence, using the orthogonality conditions in (22) and (77)

$$\underline{\mathbf{\Phi}}^*_r - \underline{\mathbf{\Phi}}_r \quad = \quad -\Sigma_{u^*_r,\underline{\Upsilon}_r} \Sigma^{-1}_{\underline{\Upsilon}_r} \tag{97}$$

where $\Sigma_{u^*_r,\underline{\Upsilon}_r} = \mathbb{E} u^*_{r,t} \underline{\Upsilon}'_{r,t}$. Using arguments and notations of the proof of Lemma A.8, there exists a constant C_2 independent of s and m_1, m_2 such that

$$\mathbb{E}\left|\Upsilon_1(m_1)\Upsilon_{1+s}(m_2)\right| \leq C_1 \sum_{h_1,h_2=0}^{\infty} \rho^{h_1+h_2} \|\epsilon_1\|_4^4 \leq C_2.$$

By the Cauchy-Schwarz inequality and (79), we then have

$$\left\|\mathrm{Cov}\left(\Upsilon_{t-r-h}, \underline{\Upsilon}_{r,t}\right)\right\| \leq C_2 r^{1/2} K(p+q).$$

Thus,

$$\begin{aligned}
\|\Sigma_{u^*_r,\underline{\Upsilon}_r}\| \quad &= \quad \|\sum_{i=r+1}^{\infty} \Phi_i \mathbb{E} \Upsilon_{t-i} \underline{\Upsilon}'_{r,t}\| \leq \sum_{h=1}^{\infty} \|\Phi_{r+h}\| \left\|\mathrm{Cov}\left(\Upsilon_{t-r-h}, \underline{\Upsilon}_{r,t}\right)\right\| \\
&= \quad \mathrm{O}(1) r^{1/2} \sum_{h=1}^{\infty} \|\Phi_{r+h}\|. \qquad (98)
\end{aligned}$$

Note that the assumption $\|\Phi_i\| = \mathrm{o}\left(i^{-2}\right)$ entails $r \sum_{h=1}^{\infty} \|\Phi_{r+h}\| = \mathrm{o}(1)$ as $r \to \infty$. The lemma therefore follows from (97), (98) and Lemma A.6. $\square$

The following lemma is similar to Lemma 3 in [6].

Lemma A.12. *Under the assumptions of Theorem 3.10,*

$$\sqrt{r}\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}-\Sigma_{\underline{\Upsilon}_r}^{-1}\| \quad = \quad \mathrm{o}_{\mathbb{P}}(1)$$

as $n\to\infty$ when $r=\mathrm{o}(n^{1/3})$ and $r\to\infty$.

Proof. We have

$$\begin{aligned}\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}-\Sigma_{\underline{\Upsilon}_r}^{-1}\right\| &= \left\|\left\{\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}-\Sigma_{\underline{\Upsilon}_r}^{-1}+\Sigma_{\underline{\Upsilon}_r}^{-1}\right\}\left\{\Sigma_{\underline{\Upsilon}_r}-\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}\right\}\Sigma_{\underline{\Upsilon}_r}^{-1}\right\| \\ &\le \left(\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}-\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|+\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|\right)\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}-\Sigma_{\underline{\Upsilon}_r}\right\|\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|.\end{aligned}$$

Iterating this inequality, we obtain

$$\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}-\Sigma_{\underline{\Upsilon}_r}^{-1}\right\| \quad \le \quad \left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|\sum_{i=1}^{\infty}\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}-\Sigma_{\underline{\Upsilon}_r}\right\|^i\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|^i.$$

Thus, for every $\varepsilon>0$,

$$\begin{aligned}&\mathbb{P}\left(\sqrt{r}\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}-\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|>\varepsilon\right)\\ \le\ &\mathbb{P}\left(\sqrt{r}\frac{\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|^2\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}-\Sigma_{\underline{\Upsilon}_r}\right\|}{1-\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}-\Sigma_{\underline{\Upsilon}_r}\right\|\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|}>\varepsilon \text{ and } \left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}-\Sigma_{\underline{\Upsilon}_r}\right\|\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|<1\right)\\ &+\mathbb{P}\left(\sqrt{r}\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}-\Sigma_{\underline{\Upsilon}_r}\right\|\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|\ge 1\right)\\ \le\ &\mathbb{P}\left(\sqrt{r}\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}-\Sigma_{\underline{\Upsilon}_r}\right\|>\frac{\varepsilon}{\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|^2+\varepsilon r^{-1/2}\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|}\right)\\ &+\mathbb{P}\left(\sqrt{r}\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}-\Sigma_{\underline{\Upsilon}_r}\right\|\ge\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|^{-1}\right)=\mathrm{o}(1)\end{aligned}$$

by Lemmas A.9 and A.6. This establishes Lemma A.12. □

Lemma A.13. *Under the assumptions of Theorem 3.10,*

$$\sqrt{r}\left\|\hat{\underline{\mathbf{\Phi}}}_r-\underline{\mathbf{\Phi}}_r\right\|=\mathrm{o}_{\mathbb{P}}(1)$$

as $r\to\infty$ and $r=\mathrm{o}(n^{1/3})$.

Proof. By the triangle inequality and Lemmas A.6 and A.12, we have

$$\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}\right\|\le\left\|\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}-\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|+\left\|\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|=\mathrm{O}_{\mathbb{P}}(1). \tag{99}$$

Note that the orthogonality conditions in (77) entail that $\underline{\mathbf{\Phi}}_r=\Sigma_{\Upsilon,\underline{\Upsilon}_r}\Sigma_{\underline{\Upsilon}_r}^{-1}$. By Lemmas A.6, A.9, A.12, and (99), we then have

$$\begin{aligned}\sqrt{r}\left\|\hat{\underline{\mathbf{\Phi}}}_r-\underline{\mathbf{\Phi}}_r\right\| &= \sqrt{r}\left\|\hat{\Sigma}_{\hat{\Upsilon},\hat{\underline{\Upsilon}}_r}\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}-\Sigma_{\Upsilon,\underline{\Upsilon}_r}\Sigma_{\underline{\Upsilon}_r}^{-1}\right\|\\ &= \sqrt{r}\left\|\left(\hat{\Sigma}_{\hat{\Upsilon},\hat{\underline{\Upsilon}}_r}-\Sigma_{\Upsilon,\underline{\Upsilon}_r}\right)\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}+\Sigma_{\Upsilon,\underline{\Upsilon}_r}\left(\hat{\Sigma}_{\hat{\underline{\Upsilon}}_r}^{-1}-\Sigma_{\underline{\Upsilon}_r}^{-1}\right)\right\|=\mathrm{o}_{\mathbb{P}}(1).\end{aligned}$$

□

Proof of Theorem 3.10. In view of (21), it suffices to show that $\hat{\underline{\mathbf{\Phi}}}_r(1) \to \underline{\mathbf{\Phi}}(1)$ and $\hat{\Sigma}_{u_r} \to \Sigma_u$ in probability. Let the $r \times 1$ vector $\mathbf{1}_r = (1, \dots, 1)'$ and the $r(p+q)K \times (p+q)K$ matrix $\mathbf{E}_r = \mathbb{I}_{(p+q)K} \otimes \mathbf{1}_r$, where $\otimes$ denotes the matrix Kronecker product and $\mathbb{I}_d$ the $d \times d$ identity matrix. Using (79), and Lemmas A.11, A.13, we obtain

$$\begin{aligned}
\left\|\hat{\underline{\mathbf{\Phi}}}_r(1) - \underline{\mathbf{\Phi}}(1)\right\| &\leq \left\|\sum_{i=1}^{r}\left(\hat{\Phi}_{r,i} - \Phi_{r,i}\right)\right\| + \left\|\sum_{i=1}^{r}\left(\Phi_{r,i} - \Phi_i\right)\right\| + \left\|\sum_{i=r+1}^{\infty}\Phi_i\right\| \\
&= \left\|\left(\hat{\underline{\mathbf{\Phi}}}_r - \underline{\mathbf{\Phi}}_r\right)\mathbf{E}_r\right\| + \left\|\left(\underline{\mathbf{\Phi}}_r^* - \underline{\mathbf{\Phi}}_r\right)\mathbf{E}_r\right\| + \left\|\sum_{i=r+1}^{\infty}\Phi_i\right\| \\
&\leq \sqrt{(p+q)K}\sqrt{r}\left\{\left\|\hat{\underline{\mathbf{\Phi}}}_r - \underline{\mathbf{\Phi}}_r\right\| + \left\|\underline{\mathbf{\Phi}}_r^* - \underline{\mathbf{\Phi}}_r\right\|\right\} + \left\|\sum_{i=r+1}^{\infty}\Phi_i\right\| \\
&= \mathrm{o}_{\mathbb{P}}(1).
\end{aligned}$$

Now note that

$$\hat{\Sigma}_{u_r} = \hat{\Sigma}_{\hat{\Upsilon}} - \hat{\underline{\mathbf{\Phi}}}_r\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r}$$

and, by (22)

$$\begin{aligned}
\Sigma_u &= \mathbb{E}u_t u_t' = \mathbb{E}u_t\Upsilon_t' = \mathbb{E}\left\{\left(\Upsilon_t - \sum_{i=1}^{\infty}\Phi_i\Upsilon_{t-i}\right)\Upsilon_t'\right\} \\
&= \Sigma_\Upsilon - \sum_{i=1}^{\infty}\Phi_i\mathbb{E}\Upsilon_{t-i}\Upsilon_t' = \Sigma_\Upsilon - \underline{\mathbf{\Phi}}_r^*\Sigma'_{\Upsilon,\underline{\Upsilon}_r} - \sum_{i=r+1}^{\infty}\Phi_i\mathbb{E}\Upsilon_{t-i}\Upsilon_t'.
\end{aligned}$$

Thus,

$$\begin{aligned}
\left\|\hat{\Sigma}_{u_r} - \Sigma_u\right\| &= \Big\|\hat{\Sigma}_{\hat{\Upsilon}} - \Sigma_\Upsilon - \left(\hat{\underline{\mathbf{\Phi}}}_r - \underline{\mathbf{\Phi}}_r^*\right)\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} \\
&\qquad -\underline{\mathbf{\Phi}}_r^*\left(\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} - \Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right) + \sum_{i=r+1}^{\infty}\Phi_i\mathbb{E}\Upsilon_{t-i}\Upsilon_t'\Big\| \\
&\leq \left\|\hat{\Sigma}_{\hat{\Upsilon}} - \Sigma_\Upsilon\right\| + \left\|\left(\hat{\underline{\mathbf{\Phi}}}_r - \underline{\mathbf{\Phi}}_r^*\right)\left(\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} - \Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right)\right\| \\
&\qquad + \left\|\left(\hat{\underline{\mathbf{\Phi}}}_r - \underline{\mathbf{\Phi}}_r^*\right)\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right\| + \left\|\underline{\mathbf{\Phi}}_r^*\left(\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} - \Sigma'_{\Upsilon,\underline{\Upsilon}_r}\right)\right\| \\
&\qquad + \left\|\sum_{i=r+1}^{\infty}\Phi_i\mathbb{E}\Upsilon_{t-i}\Upsilon_t'\right\|. \qquad (100)
\end{aligned}$$

In the right-hand side of this inequality, the first norm is $\mathrm{o}_{\mathbb{P}}(1)$ by Lemma A.9. By Lemmas A.11 and A.13, we have $\|\hat{\underline{\mathbf{\Phi}}}_r - \underline{\mathbf{\Phi}}_r^*\| = \mathrm{o}_{\mathbb{P}}(r^{-1/2}) = \mathrm{o}_{\mathbb{P}}(1)$, and by Lemma A.9, $\|\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} - \Sigma'_{\Upsilon,\underline{\Upsilon}_r}\| = \mathrm{o}_{\mathbb{P}}(r^{-1/2}) = \mathrm{o}_{\mathbb{P}}(1)$. Therefore the second norm in the right-hand side of (100) tends to zero in probability. The third norm tends to zero in probability because $\|\hat{\underline{\mathbf{\Phi}}}_r - \underline{\mathbf{\Phi}}_r^*\| = \mathrm{o}_{\mathbb{P}}(1)$ and, by Lemma A.6, $\|\Sigma'_{\Upsilon,\underline{\Upsilon}_r}\| = \mathrm{O}(1)$. The fourth norm tends to zero in probability because, in view of Lemma A.9, $\|\hat{\Sigma}'_{\hat{\Upsilon},\underline{\hat{\Upsilon}}_r} - \Sigma'_{\Upsilon,\underline{\Upsilon}_r}\| = \mathrm{o}_{\mathbb{P}}(1)$, and, in view of (79), $\|\underline{\mathbf{\Phi}}_r^*\|^2 \leq \sum_{i=1}^{\infty}\mathrm{Tr}(\Phi_i\Phi_i') < \infty$. Clearly, the last norm tends to zero, which completes the proof. □

Acknowledgements. The authors wish to acknowledge the support from the " Séries temporelles et valeurs extrêmes : théorie et applications en modélisation et estimation des risques" Projet Région grant No OPE-2017-0068.

References

[1] Anderson, P. L. and Meerschaert, M. M. (1997). Periodic moving averages of random variables with regularly varying tails. *Ann. Statist.* **25**, 2, 771–785.

[2] Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 3, 817–858.

[3] Azrak, R. and Mélard, G. (1998). The exact quasi-likelihood of time-dependent ARMA models. *J. Statist. Plann. Inference* **68**, 1, 31–45.

[4] Azrak, R. and Mélard, G. (2006). Asymptotic properties of quasi-maximum likelihood estimators for ARMA models with time-dependent coefficients. *Stat. Inference Stoch. Process.* **9**, 3, 279–330.

[5] Basawa, I. V. and Lund, R. (2001). Large sample properties of parameter estimates for periodic ARMA models. *J. Time Ser. Anal.* **22**, 6, 651–663.

[6] Berk, K. N. (1974). Consistent autoregressive spectral estimates. *Ann. Statist.* *2*, 489–502. Collection of articles dedicated to Jerzy Neyman on his 80th birthday.

[7] Bibi, A. and Francq, C. (2003). Consistent and asymptotically normal estimators for cyclically time-dependent linear models. *Ann. Inst. Statist. Math.* **55**, 1, 41–68.

[8] Billio, M., Monfort, A., and Robert, C. P. (1999). Bayesian estimation of switching ARMA models. *J. Econometrics* **93**, 2, 229–255.

[9] Boubacar Mainassara, Y., Carbon, M., and Francq, C. (2012). Computing and estimating information matrices of weak ARMA models. *Comput. Statist. Data Anal.* **56**, 2, 345–361.

[10] Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. in Appl. Probab.* **18**, 1, 211–220.

[11] Brockwell, P. J. and Davis, R. A. (1991). *Time series: theory and methods*, Second ed. Springer Series in Statistics. Springer-Verlag, New York.

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

[13] Davidson, J. (1994). *Stochastic limit theory.* Advanced Texts in Econometrics. The Clarendon Press, Oxford University Press, New York. An introduction for econometricians.

[14] Davydov, J. A. (1968). Convergence of distributions generated by stationary stochastic processes. *Theor. of Proba. and Appli.* **13**, 2, 691–696.

[15] den Haan, W. J. and Levin, A. T. (1997). A practitioner's guide to robust covariance matrix estimation. In *Robust inference.* Handbook of Statist., Vol. **15**. North-Holland, Amsterdam, 299–342.

[16] Francq, C. and Gautier, A. (2004a). Estimation of time-varying ARMA models with Markovian changes in regime. *Statist. Probab. Lett.* **70**, 4, 243–251.

[17] Francq, C. and Gautier, A. (2004b). Large sample properties of parameter least squares estimates for time-varying ARMA models. *J. Time Ser. Anal.* **25**, 5, 765–783.

[18] Francq, C. and Roussignol, M. (1997). On white noises driven by hidden Markov chains. *J. Time Ser. Anal.* **18**, 6, 553–578.

[19] Francq, C. and Roussignol, M. (1998). Ergodicity of autoregressive processes with Markov-switching and consistency of the maximum-likelihood estimator. *Statistics* **32**, 2,

151–173.

[20] Francq, C. and Zakoïan, J.-M. (1998). Estimating linear representations of nonlinear processes. *J. Statist. Plann. Inference* **68**, 1, 145–165.

[21] Francq, C. and Zakoïan, J.-M. (2001). Stationarityof multivariate markov-switching arma models. *J. Econometrics* **102**, 2, 339–364.

[22] Francq, C. and Zakoïan, J.-M. (2002). Autocovariance structure of powers of switching-regime ARMA processes. *ESAIM Probab. Statist. 6*, 259–270. New directions in time series analysis (Luminy, 2001).

[23] Francq, C. and Zakoïan, J.-M. (2007). HAC estimation and strong linearity testing in weak ARMA models. *J. Multivariate Anal.* **98**, 1, 114–144.

[24] Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications.* Wiley.

[25] Grenander, U. and Szegö, G. (1958). *Toeplitz forms and their applications.* California Monographs in Mathematical Sciences. University of California Press, Berkeley-Los Angeles.

[26] Hamilton, J. D. (1988). Rational-expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. *J. Econom. Dynam. Control* **12**, 2-3, 385–423. Economic time series with random walk and other nonstationary components.

[27] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**, 2, 357–384.

[28] Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *J. Econometrics* **45**, 1-2, 39–70.

[29] Hamilton, J. D. (1994). *Time series analysis.* Princeton University Press, Princeton, NJ.

[30] Hamilton, J. D. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of econometrics* **64**, 1, 307–333.

[31] Herrndorf, N. (1984). A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Probab.* **12**, 1, 141–153.

[32] Jones, G. L. (2004). On the markov chain central limit theorem. *Probab. Surv. 1*, 299–320.

[33] Kim, C.-J. and Kim, J. (2015). Bayesian inference in regime-switching ARMA models with absorbing states: the dynamics of the ex-ante real interest rate under regime shifts. *J. Bus. Econom. Statist.* **33**, 4, 566–578.

[34] Newey, W. K. and West, K. D. (1987). A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 3, 703–708.

[35] Nicholls, D. F. and Quinn, B. G. (1982). *Random coefficient autoregressive models: an introduction.* Lecture Notes in Statistics, Vol. **11**. Springer-Verlag, New York-Berlin. Lecture Notes in Physics, 151.

[36] Romano, J. P. and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *J. Amer. Statist. Assoc.* **91**, 434, 590–600.