MPRA

Munich Personal RePEc Archive

Asymptotic properties of weighted least squares estimation in weak parma models

Francq, Christian; Roy, Roch and Saidi, Abdessamad

CREST (CNRS) and University Lille 3 (EQUIPPE),
Université de Montréal, Bank Al-Maghrib

01. February 2011

Asymptotic Properties of Weighted Least Squares Estimation in Weak PARMA Models

Christian Francq,[*] Roch Roy[†] and Abdessamad Saidi[‡]

Abstract

The aim of this work is to investigate the asymptotic properties of weighted least squares (WLS) estimation for causal and invertible periodic autoregressive moving average (PARMA) models with uncorrelated but dependent errors. Under mild assumptions, it is shown that the WLS estimators of PARMA models are strongly consistent and asymptotically normal. It extends Theorem 3.1 of Basawa and Lund (2001) on least squares estimation of PARMA models with independent errors. It is seen that the asymptotic covariance matrix of the WLS estimators obtained under dependent errors is generally different from that obtained with independent errors. The impact can be dramatic on the standard inference methods based on independent errors when the latter are dependent. Examples and simulation results illustrate the practical relevance of our findings. An application to financial data is also presented.

Keywords: Weak periodic autoregressive moving average models, Seasonality, Weighted least squares, Asymptotic normality, Strong consistency, Weak periodic white noise, Strong mixing.

MSC 2010 subject classification: Primary 62M10; secondary 62M15.

First version: October 20, 2009
This revision: February 4, 2011

1   Introduction

Periodically correlated time series are common in many scientific fields where the observed phenomena may have significant periodic behavior in mean, variance and covariance structure, namely in hydrology, meteorology, finance and economy. An important class of stochastic models for describing such periodicity in mean and in covariances, are the periodic autoregressive moving average (PARMA) models. PARMA models are an extension of autoregressive moving average (ARMA) models in the

[*]CREST (CNRS) et Université Lille 3 (EQUIPPE), 59653 Villeneuve d'Ascq Cedex, France (e-mail: christian.francq@univ-lille3.fr).

[†]**Corresponding author**: Département de mathématiques et de statistique and Centre de recherches mathématiques, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, H3C 3J7, Canada (e-mail: royr@dms.umontreal.ca).

[‡]Département de la Recherche, Bank Al-Maghrib, Rabat, Maroc (e-mail: ab.saidi@bkam.ma). Most of the research was carried out while he was a postdoctoral fellow at Université de Montréal.

sense that they allow the model parameters to vary with respect to time. The literature on periodic time series models has abounded since the seventies. For prior works, see among others Gladyshev (1961) and Jones and Brelsford (1967). Tiao and Grupe (1980) illustrated the pitfalls of ignoring the periodic behavior in time series modelling. Empirical evidence supporting the usefulness of PARMA models was documented by many authors, see for example, Vecchia (1985a, 1985b), Salas and Obeysekera (1992), Lund (2006), Tesfaye *et al.* (2006) for applications to streamflow series, Bloomfield *et al.* (1994), Lund *et al.* (2006) to environmental data, Osborn and Smith (1989) to economic data and Gardner and Spooner (1994) for applications in signal processing.

Time series modelling usually involves three main steps: model identification, parameter estimation and diagnostic checking. There is a substantial literature on estimation of PARMA models. Pagano (1978) dealt with moment estimation of periodic autoregressive (PAR) models. He proved that those estimators are almost surely consistent and asymptotically efficient under Gaussianity. Salas *et al.* (1982) investigated moment estimation of low order PARMA models. They observed that the estimators of the periodic moving average parameters are often unsatisfactory and that the Yule-Walker equations become more complicated. Vecchia (1985a, 1985b) investigated Gaussian maximum likelihood estimation of PARMA models and established its superiority over moment estimation. Jimenez *et al.* (1989) presented an exact maximum likelihood procedure for estimating the parameters of a PARMA model using a state-space representation and a Kalman filtering algorithm. Basawa and Lund (2001) established the asymptotic properties of the least squares (LS) estimators of PARMA models with independent errors; they extended the results for periodic autoregression earlier derived by Pagano (1978) and Troutman (1979). Lund and Basawa (2000) developed an efficient algorithm for maximum likelihood estimation of Gaussian PARMA models. An extensive simulation study conducted by Smadi (2005) shows that LS estimation of PAR models with non-Gaussian errors is quite satisfactory even with heavy tails like in the Cauchy distribution.

The aforementioned estimation procedures for PARMA models were established under the assumption of independent errors (strong PARMA). Of course, this assumption is not satisfied for nonlinear processes that admit a weak PARMA representation (the errors are uncorrelated but dependent) such as the periodic generalized autoregressive conditional heteroskedastic (PGARCH) and periodic bilinear processes (PBL). Another argument in favor of considering the weak PARMA models comes from the fact that, in general, temporal aggregation or systematic sampling of a strong PARMA model yield a weak PARMA model, see Roy and Saidi (2008). Finally, note that many time series encountered in practice cannot be described by strong PARMA models. For instance, Wang *et al.* (2005, 2006) found evidences of the existence of autoregressive conditional heteroskedastic effects, a nonlinear

2

phenomenon in the variance behavior, in the residual series obtained from fitting conventional linear streamflow models to daily and monthly streamflow series of the upper Yellow River in China. In this type of situation, it is necessary to relax the independence assumption and to consider nonlinear models for describing such time series. All these examples have important practical meanings and emphasize the need for taking into account an eventual dependence of the errors when estimating a PARMA model.

In recent years, a large part of the time series and econometric literature was devoted to weaken the strong noise assumption. In particular, Romano and Thombs (1996) showed that the significance limits of the sample autocorrelations obtained under the strong ARMA assumption can be quite misleading if the underlying innovations are only uncorrelated rather than independent. Francq and Zakoïan (1998a) and Francq, Roy and Zakoïan (2005), considered least squares estimation and tests for lack of fit in weak ARMA models. They showed that the standard Box-Pierce and Ljung-Box portmanteau tests can perform poorly if the errors are only uncorrelated. Under mild assumptions, Francq and Zakoïan (2004) derived the strong consistency and asymptotic normality of the quasi-maximum likelihood estimator of pure GARCH models and of ARMA models with noise sequence driven by a GARCH model. Aknouche and Bibi (2009) extended this latter work to the case of pure PGARCH models and PARMA models with PGARCH noise.

The main goal of this paper is to study the asymptotic properties of least squares estimation for invertible and causal weak PARMA models. Four different LS estimators are considered: ordinary least squares (OLS), weighted least squares (WLS) for an arbitrary vector of weights, generalized least squares (GLS) in which the weights correspond to the theoretical seasonal variances and quasi-generalized least squares (QLS) where the weights are the estimated seasonal variances. It is seen that the GLS estimators are optimal in the class of WLS estimators when the noise sequence is in a particular class of martingale differences. The strong consistency and the asymptotic normality are established for each of them. Obviously, their asymptotic covariance matrices depend on the vector of weights. Our results extend Theorem 3.1 of Basawa and Lund (2001) for least squares estimation of PARMA with independent errors (strong PARMA). Furthermore, we retrieve results of Francq and Zakoïan (1998a), when the period is one, *i.e.*, the model is a weak stationary and invertible ARMA.

The paper is organized as follows. In Section 2, we provide examples of weak periodic noises and of nonlinear processes admitting a weak PARMA representation. The asymptotic results are described in Section 3. Since the proofs are rather long and technical, they are relegated to an Appendix. In Section 4, we present two examples of weak PARMA models for which the asymptotic covariance matrix of the least squares estimators is given in a close form and is compared to the corresponding matrix

3

under the assumption of a strong noise. Monte Carlo results are described in Section 5. In the first part of the experiment, we considered various white noises (strong or weak) to which we fitted a PAR model. The discrepancy between the empirical standard errors of the parameter estimators and their theoretical asymptotic standard errors under the assumption of a strong noise is examined, as well as the size distortion of a Wald test for the hypothesis that the model parameters are zero. In the second part, two different PARMA models with strong and weak noises were used to investigate the size and power of a Wald test based on a consistent estimator of the asymptotic covariance matrix, under the assumption of either a weak or strong noise. The rate of convergence of the estimated asymptotic standard errors is also analysed. Our results are exploited in Section 6 to address the question of day-of-the-week seasonality of four European stock market indices. Finally, some concluding remarks are presented in Section 7.

2 Weak and Strong PARMA models

A stochastic process $\{X_t\}$ is called periodically stationary if $\mu_t = \mathbb{E}[X_t]$ and $\gamma_t(h) = \mathbb{E}[X_t X_{t+h}]$, $h \in \mathbb{Z}$, are both periodic functions in time t with the same period T and $\mathbb{E}[X_t^2] < +\infty$ for all t. For convenience, the non-periodic notation X_t will be used interchangeably with the periodic notation $X_{nT+\nu}$ which refers to X_t during the season $\nu \in \{1, ..., T\}$ in the cycle n. By definition, a periodic process $\{X_t\}$ follows a periodic (with period T) autoregressive moving average model with the following parameters at season $\nu \in \{1, ..., T\}$: the mean μ_ν, the autoregressive order and coefficients p_ν, $\phi_1(\nu), ..., \phi_{p_\nu}(\nu)$, and the moving average order and coefficients $q_\nu, \theta_1(\nu), ..., \theta_{q_\nu}(\nu)$, denoted simply $\mathrm{PARMA}_T(p_1, ..., p_\nu, ..., p_T; q_1, ..., q_\nu, ..., q_T)$, if there exists a periodic white noise sequence $\{\epsilon_t\} = \{\epsilon_{nT+\nu}\}$, i.e. $\mathbb{E}[\epsilon_t] = 0$ for all t, $\mathbb{E}[\epsilon_t \epsilon_{t'}] = 0$ for all $t \neq t'$, and $\mathbb{E}[\epsilon_{nT+\nu}^2] = \sigma_\nu^2 > 0$, such that

$$(X_{nT+\nu} - \mu_\nu) - \sum_{k=1}^{p_\nu} \phi_k(\nu)(X_{nT+\nu-k} - \mu_{\nu-k}) = \epsilon_{nT+\nu} - \sum_{l=1}^{q_\nu} \theta_l(\nu)\epsilon_{nT+\nu-l}. \tag{2.1}$$

If the errors ϵ_t are uncorrelated but not necessarily independent, both periodic white noise or weak periodic white noise are used to qualify the error process $\{\epsilon_t\}$ and similarly the terminology PARMA or weak PARMA are used for the model (2.1). When the error terms $\sigma_\nu^{-1}\epsilon_{nT+\nu}$ are independent and identically distributed (iid) rather than only uncorrelated, the model (2.1) is called strong PARMA model and $\{\epsilon_t\}$ is a strong periodic white noise.

When the order of both the autoregressive and moving average components are not allowed to vary with season, i.e., when $p_1 = ... = p_T = p$ and $q_1 = ... = q_T = q$ we simply write $\mathrm{PARMA}_T(p; q)$ instead of $\mathrm{PARMA}_T(p, ..., p; q, ..., q)$. The terminology periodic autoregressive (PAR) model and periodic moving average (PMA) model are respectively used when the moving average orders are null, and when

4

the autoregressive orders are null. If $T = 1$ the process (2.1) is the usual stationary autoregressive moving average model (ARMA).

2.1 Examples of periodic weak white noises

In this section, we give examples of periodic white noises that are uncorrelated but dependent. We also present data-generating processes that can be compatible with weak PARMA representation.

2.1.1 Periodic weak white noise derived from a strong white noise

The following weak white noise example is inspired by examples given in Romano and Thombs (1996). Let $\{\xi_t\}$ be any sequence of iid random variables with $\mathbb{E}[\xi_t] = 0$, $\mathbb{E}[\xi_t^2] = 1$ and having finite fourth-order moment. For fixed $\nu \in \{1, ..., T\}$, let

$$\epsilon_{nT+\nu} = \sigma_\nu \prod_{j=0}^{m} \xi_{nT+\nu-j} \tag{2.2}$$

where $m > 0$ is a fixed integer and $\sigma_1, ..., \sigma_T$ are positive constants. The periodic process $\{\epsilon_{nT+\nu}\}$ is a weak white noise because $\mathbb{E}[\epsilon_t] = 0$ for all t, $\mathbb{E}[\epsilon_t \epsilon_{t'}] = 0$ for all $t \neq t'$, $\mathbb{E}[\epsilon_{nT+\nu}^2] = \sigma_\nu^2 > 0$. It is a m-dependent white noise since the variables ϵ_t and $\epsilon_{t'}$ are dependent if $|t - t'| \leq m$ but they are independent for $|t - t'| > m$.

2.1.2 Periodic weak white noise derived from nonlinear processes

Some models usually encountered in the nonlinear time series literature constitute important examples of periodic weak white noises. The simplest of these is the generalized autoregressive conditional heteroskedasticity (GARCH) model. Indeed, the stationary solution of a GARCH model is a martingale difference and therefore is a weak white noise. It must be noted that the variance here is constant. However, a periodic weak white noise of period T, is given by the following process

$$\epsilon_{nT+\nu} = \sigma_\nu \left(\eta_{nT+\nu} / \sqrt{\mathbb{E}[\eta_{nT+\nu}^2]} \right),$$

where $\{\eta_t\}$ represents the stationary solution of the following GARCH(P, Q) process

$$\begin{cases} \eta_t & = h_t \xi_t, \\ h_t^2 & = \omega + \sum_{i=1}^{Q} \alpha_i \eta_{t-i}^2 + \sum_{j=1}^{P} \beta_j h_{t-j}^2, \end{cases}$$

where $\{\xi_t\}$ is a sequence of iid centered variables with unit variance, the α_i and β_j are nonnegative constants, and ω is a positive constant. Under the assumption that $\sum_{i=1}^{Q} \alpha_i + \sum_{j=1}^{P} \beta_j < 1$, there exists an unique stationary and nonanticipative solution process $\{\eta_t\}$ that has a finite variance. We can easily

5

check that this is a periodic weak white noise and that the variance is not constant but periodic with period T. Under more restrictive conditions on the coefficients and $\mathbb{E}\left[\xi_t^4\right] < \infty$, then $\mathbb{E}\left[\epsilon_t^4\right] < \infty$ (see Ling and McAleer, 2002). The simple extension of GARCH models to the periodic case is discussed by Bollerslev and Ghysels (1996). To illustrate this case, consider the following periodic (with period $T = 2$) ARCH model

$$
\begin{cases}
\epsilon_{nT+\nu} = h_{nT+\nu}\xi_{nT+\nu}, \\
h_{nT+\nu}^2 = \alpha_{\nu,0} + \alpha_{\nu,1}\epsilon_{(n-1)T+\nu}^2,
\end{cases}
$$

with $\{\xi_t\}$ being a sequence of iid $\mathcal{N}(0,1)$ variables. It is easy to check that the periodic stationary solution is a periodic weak white noise. In a similar spirit, the class of bilinear processes and the class of periodic purely bilinear and strictly superdiagonal processes constitute important examples of periodic weak white noises. The last class is characterized by the following equation

$$
\epsilon_{nT+\nu} = \xi_{nT+\nu} + \sum_{i=2}^{P}\alpha_{\nu,i}\epsilon_{nT+\nu-i}\xi_{nT+\nu-1}
$$

with $P \geq 2$ and $\{\xi_t\}$ is any sequence of iid random variables with $\mathbb{E}\left[\xi_t\right] = \mathbb{E}\left[\xi_t^3\right] = 0$, $\mathbb{E}\left[\xi_t^2\right] = 1$ and having finite fourth-order moment. Bibi and Gautier (2006) give conditions ensuring the existence of a causal and invertible solution and show that the solution is a periodic weak white noise.

2.2 Data generating processes with weak PARMA representations

2.2.1 Temporal aggregation and systematic sampling

Temporal aggregation and systematic sampling of a stochastic process $\{X_t, t \in \mathbb{Z}\}$ over non-overlapping periods of length M are particular cases of the following linear transformation. If $\{Y_{\tilde{t}}, \tilde{t} \in \mathbb{Z}\}$ represents the resulting process at date $\tilde{t}$,

$$
Y_{\tilde{t}} = \sum_{i=1}^{M} c_i X_{M(\tilde{t}-1)+i}, \tag{2.3}
$$

where $c_1, c_2, ..., c_M$ are real constants. For temporal aggregation, $c_1 = c_2 = ... = c_M = 1$, and for systematic sampling, $c_1 = c_2 = ... = c_{M-1} = 0$, $c_M = 1$.

When the high frequency process $\{X_t\}$ is periodic of period T, we suppose that $M \leq T$ and that $T = M\tilde{T}$ for some $\tilde{T} \in \mathbb{N}$. For example, with monthly data aggregated in quarterly data, $T = 12$, $M = 3$ and $\tilde{T} = 4$. In such a situation, the low frequency process $\{Y_{\tilde{t}}\}$ is also periodically correlated of period $\tilde{T}$.

Roy and Saidi (2008) showed that the class of weak PARMA processes is closed under the aggregation transformation (2.3) but that property does not hold any more with the class of strong PARMA processes. Furthermore, a sufficient condition under which temporal aggregation of a strong PARMA

6

model yields a weak PARMA model is provided. Under that condition, the noise of the aggregated process is neither strong nor a martingale difference.

2.2.2 Nonlinear processes with weak PARMA representations

In general, it seems difficult to prove the existence of a PARMA representation for some models usually encountered in the periodic nonlinear time series literature. However, for some special cases like bilinear models, it is possible to show that they can admit a weak PARMA representation. For instance, consider the following periodic (with period T=2) bilinear model

$$\begin{cases} X_{nT+1} = \tilde{\epsilon}_{nT+1} + \alpha X_{nT}\tilde{\epsilon}_{nT-1}, \\ X_{nT+2} = \tilde{\epsilon}_{nT+2} + \beta X_{nT+1}\tilde{\epsilon}_{nT}, \end{cases}$$

where the $\tilde{\epsilon}_t$'s are independent with $\mathbb{E}[\tilde{\epsilon}_t] = 0$ for all t, and $\mathbb{E}\left[\tilde{\epsilon}_{nT+1}^2\right] = \tilde{\sigma}_1^2$, $\mathbb{E}\left[\tilde{\epsilon}_{nT+2}^2\right] = \tilde{\sigma}_2^2$. If $|\alpha\beta| < 1$, the process $\{X_t\}$ admits a periodic stationary solution and using Shao and Lund (2004) characterization of PMA models, we obtain that this solution admits a weak $\text{PMA}_2(3)$ representation.

2.2.3 Causal representations of noncausal PARMA models

Let us consider the following $\text{PAR}_2(2)$ model

$$\begin{cases} X_{nT+1} - \alpha X_{(n-1)T+1} = \tilde{\epsilon}_{nT+1}, \\ X_{nT+2} - \beta X_{(n-1)T+2} = \tilde{\epsilon}_{nT+2}, \end{cases}$$

where the $\tilde{\epsilon}_t$'s are independent with $\mathbb{E}[\tilde{\epsilon}_t] = 0$ for all t, and $\mathbb{E}\left[\tilde{\epsilon}_{nT+1}^2\right] = \tilde{\sigma}_1^2$, $\mathbb{E}\left[\tilde{\epsilon}_{nT+2}^2\right] = \tilde{\sigma}_2^2$. We also assume that $|\alpha| > 1$ and that $|\beta| > 1$. In that case, the process $\{X_t\}$ admits a noncausal representation of the form

$$\begin{cases} X_{nT+1} = -\sum_{i=1}^{\infty}\alpha^{-i}\tilde{\epsilon}_{(n+i)T+1}, \\ X_{nT+2} = -\sum_{j=1}^{\infty}\beta^{-j}\tilde{\epsilon}_{(n+j)T+2}. \end{cases}$$

Now, let

$$\begin{cases} \epsilon_{nT+1} = X_{nT+1} - \alpha^{-1}X_{(n-1)T+1}, \\ \epsilon_{nT+2} = X_{nT+2} - \beta^{-1}X_{(n-1)T+2}. \end{cases}$$

It is clear that $\mathbb{E}[\epsilon_t] = 0$ for all t, $\mathbb{E}[\epsilon_t\epsilon_{t'}] = 0$ for all $t \neq t'$ and $\mathbb{E}\left[\epsilon_{nT+\nu}^2\right] = \sigma_\nu^2 > 0$. Thus, $\{X_t\}$ admits a stationary $\text{PAR}_2(2)$ representation. Moreover, we can check that $\mathbb{E}\left[X_{(n-1)T+1}^3\right] = (1-\alpha^3)^{-1}\mathbb{E}\left[\tilde{\epsilon}_{nT+1}^3\right]$ and $\mathbb{E}\left[X_{nT+1}X_{(n-1)T+1}^2\right] = \alpha^{-2}(1-\alpha^3)^{-1}\mathbb{E}\left[\tilde{\epsilon}_{nT+1}^3\right]$. This implies that

$$\mathbb{E}\left[\epsilon_{nT+1}X_{(n-1)T+1}^2\right] = \mathbb{E}\left[X_{nT+1}X_{(n-1)T+1}^2\right] - \alpha^{-1}\mathbb{E}\left[X_{(n-1)T+1}^3\right] \neq 0,$$

whenever $\mathbb{E}\left[\tilde{\epsilon}_t^3\right] \neq 0$. Therefore, the periodic white noise $\{\epsilon_t\}$ is neither strong nor a martingale difference.

Furthermore, we can show, using Corollary 1 in Cheng (1999), that the noise $\{\epsilon_t\}$ is strong if and only if the process $\{X_t\}$ is Gaussian.

2.2.4 Approximation of the Wold decomposition for periodic processes

Weak PARMA processes can be viewed as an approximation of the Wold decomposition of periodically stationary processes. Indeed, any periodically stationary process $\{X_t\}$ of period T admits an infinite periodic moving average representation of the form

$$X_{nT+\nu} = \sum_{k=0}^{\infty} \psi_{\nu,k}\epsilon_{nT+\nu-k}, \tag{2.4}$$

where $\{\epsilon_t\}$ is the linear innovation process of $\{X_t\}$, $\psi_{\nu,0} = 1$ and $\sum_{k=0}^{\infty} \psi_{\nu,k}^2 < +\infty$. The process $\{X_{nT+\nu}\}$ can be approximated by the weak PMA$(q_1, ..., q_T)$ process

$$X_{nT+\nu}(q_\nu) = \sum_{k=0}^{q_\nu} \psi_{\nu,k}\epsilon_{nT+\nu-k}, \quad \nu = 1, ..., T,$$

because

$$\mathbb{E}\left[X_{nT+\nu}(q_\nu) - X_{nT+\nu}\right]^2 \leq (\max_{1 \leq \nu \leq T} \sigma_\nu^2) \sum_{k>q_\nu} \psi_{\nu,k}^2 \to 0$$

where $q_\nu \to \infty$. The linear model (2.4), which consists of the PARMA models and their limits is very general under the noise uncorrelatedness, but can be quite restrictive if the assumption of strong noise is made.

The previous examples demonstrate that weak PARMA models can arise from various situations. Making the assumption of a strong noise precludes most of these data generating processes (DGP), as well as many others.

3 Least squares estimation of weak PARMA models

In this section, we focus on the asymptotic properties of the least squares estimators of the autoregressive and moving average parameters of the PARMA$_T$ process (2.1). There is no loss of generality in assuming that $p_1 = ... = p_T = p$ and $q_1 = ... = q_T = q$ by adding coefficients equal to zero (Lund and Basawa, 2000). Furthermore, we suppose that the process is centered, that is $\mu_1 = ... = \mu_T = 0$. We make this assumption to lighten the presentation, but the results stated in this section extend directly for models with constants. Such models will be considered for the numerical illustrations. Thus, the PARMA$_T(p, q)$ process $\{X_{nT+\nu}\}$ satisfies the following difference equations

$$X_{nT+\nu} - \sum_{i=1}^{p} \phi_i(\nu)X_{nT+\nu-i} = \epsilon_{nT+\nu} - \sum_{j=1}^{q} \theta_j(\nu)\epsilon_{nT+\nu-j}, \tag{3.1}$$

$\nu = 1, ..., T$, where $\{\epsilon_{nT+\nu}\}$ is a periodic white noise (weak or strong) and we assume that $p + q > 0$.

The process $\{\epsilon_t = \epsilon_{nT+\nu}\}$ can be interpreted as the linear innovation of $\{X_t = X_{nT+\nu}\}$, i.e.

$$\epsilon_t = X_t - \mathbb{E}\left[X_t | \mathcal{H}_X(t-1)\right]$$

where $\mathcal{H}_X(t)$ is the Hilbert space spanned by $\{X_s, s \leq t\}$.

The difference equations (3.1) can be written in the T-dimensional vector form (Vecchia, 1985b)

$$\Phi_0 \mathbf{X}_n - \sum_{k=1}^{p^*} \Phi_k \mathbf{X}_{n-k} = \Theta_0 \boldsymbol{\epsilon}_n - \sum_{l=1}^{q^*} \Theta_l \boldsymbol{\epsilon}_{n-l}, \tag{3.2}$$

where

$$\mathbf{X}_n = (X_{nT+1}, ..., X_{nT+T})', \quad \boldsymbol{\epsilon}_n = (\epsilon_{nT+1}, ..., \epsilon_{nT+T})', \tag{3.3}$$

$p^* = [(p-1)/T] + 1$, $q^* = [(q-1)/T] + 1$, the matrix coefficients Φ_k, $k = 0, ..., p^*$ and Θ_l, $l = 0, ..., q^*$, are defined by

$$(\Phi_0)_{i,j} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i < j \\ -\phi_{i-j}(i) & \text{if} \quad i > j \end{cases}, \qquad (\Theta_0)_{i,j} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i < j \\ -\theta_{i-j}(i) & \text{if} \quad i > j \end{cases}, \tag{3.4}$$

and

$$(\Phi_k)_{i,j} = \phi_{kT+i-j}(i), \quad k = 1, ..., p^*, \qquad (\Theta_l)_{i,j} = \theta_{lT+i-j}(i), \quad l = 1, ..., q^*. \tag{3.5}$$

Here, it is implicit that $\phi_h(\nu) = 0$, $h \notin \{1, \ldots, p\}$ and $\theta_h(\nu) = 0$, $h \notin \{1, \ldots, q\}$. The covariance matrix of the T-dimensional white noise $\boldsymbol{\epsilon}_n$ is $\Sigma_\epsilon = \text{Diag}(\sigma_1^2, ..., \sigma_T^2) > 0$. Denote by B the lag operator such that $B^h \mathbf{X}_n = \mathbf{X}_{n-h}$. Equation (3.2) can be written as

$$\Phi(B) \mathbf{X}_n = \Theta(B) \boldsymbol{\epsilon}_n \tag{3.6}$$

where $\Phi(z) = \Phi_0 - \Phi_1 z - ... - \Phi_{p^*} z^{p^*}$ and $\Theta(z) = \Theta_0 - \Theta_1 z - ... - \Theta_{q^*} z^{q^*}$ are the matrix polynomials of the vectorial autoregressive moving average representation. It is important to note that the lag operator B operates on the cycle index n. When it acts on the time index $t = nT + \nu$ of the periodic process $\{Z_t\}$, it gives $B^k Z_{nT+\nu} = Z_{(n-k)T+\nu}$.

From (3.2), we can in principle deduce the properties of weak PARMA parameter estimation from existing results on parameter estimation of a vector ARMA model under general assumptions on the white noise process including dependence cases. In particular, Dunsmuir and Hannan (1976) and Dunsmuir (1979) assume a higher order martingale difference condition on the white noise while Hosoya and Taniguchi (1982) and, Taniguchi and Kakizawa (2000) impose what they called an assymptotically higher order martingale difference condition. Here, we have preferred to work in the univariate

PARMA setting for various reasons. First, results obtained directly in terms of the univariate PARMA representation are more directly usable because fewer parameters are involved and their estimation is easier. For example, in a $\text{PARMA}_T(1,1)$ model, there are $2T$ AR and MA parameters to estimate. The corresponding VARMA(1,1) representation contains $4T^2$ parameters among which $2T(2T-1)$ are constrained to 1 or 0, according to (3.4) and (3.5). The difference in the number of parameters is huge, specially with monthly or weekly data ($T = 12, 52$). Second, the VARMA representation (3.2) is not standard, the matrices Φ_0 and Θ_0 are not in general the identity matrix. Rescaling the vector noise in (3.2) via $\boldsymbol{\epsilon}_n^* = \Theta_0 \boldsymbol{\epsilon}_n$ and then multiplying both sides by Φ_0^{-1} leads to a standard VARMA representation. In doing so, the covariance matrix of $\{\boldsymbol{\epsilon}_n^*\}$ and the MA parameters would depend on both the AR parameters of the PARMA representation and the vector of variances $\vec{\sigma}^2$. Even in the strong case, a time series package that provides the inverse transformation in order to be able to make inference on the parameters of the VARMA representation (3.2) under the constraints (3.4) and (3.5) does not exist at our knowledge. With the PARMA representation, we have been able to derive in Theorem 3.2 a convenient form for the covariance matrix of the least squares estimators of the AR and MA parameters. A consistent estimator of that covariance matrix, as described in Remark 3.4, was implemented in the R-language and was used in the empirical part of the paper.

Hereafter, we impose a strong mixing condition on the process $\{\mathbf{X}_n\}$.

(A1) The PARMA process $\{X_{nT+\nu}\}$ is causal and invertible, in the sense that, the roots of $\det\{\Phi(z)\}$ and of $\det\{\Theta(z)\}$ are greater than one in modulus (Brockwell and Davis, 1991). Furthermore, we assume that the VARMA model (3.6) is identifiable (see Reinsel, 1997, Section 2.3.4 or Hannan and Deistler, 1988, Section 2.7).

For notation, let $\vec{\phi}(\nu) = (\phi_1(\nu), ..., \phi_p(\nu))'$ and $\vec{\theta}(\nu) = (\theta_1(\nu), ..., \theta_q(\nu))'$ respectively denote the vectors of autoregressive and moving average parameters for season ν. The $T(p+q)$-dimensional collection of all PARMA parameters is denoted by

$$\vec{\alpha} := \left(\vec{\phi}(1)', ..., \vec{\phi}(T)', \vec{\theta}(1)', ..., \vec{\theta}(T)' \right)'.$$

The white noise variances $\vec{\sigma}^2 = (\sigma_1^2, ..., \sigma_T^2)'$ will be treated as nuisance parameters.

Let $X_1, ..., X_{NT}$ be a data sample from the causal and invertible PARMA model (3.1) with the true parameter value $\vec{\alpha} = \vec{\alpha}_0$ and $\vec{\sigma}^2 = \vec{\sigma}_0^2$. The sample contains N full periods of data which are indexed from 0 to $N-1$. Indeed, when $0 \leq n \leq N-1$ and $1 \leq \nu \leq T$, $nT + \nu$ goes from 1 to NT. It is understood that $\vec{\alpha}_0$ belongs to the parameter space

$$\Omega = \left\{ \vec{\alpha} = \left(\vec{\phi}(1)', ..., \vec{\phi}(T)', \vec{\theta}(1)', ..., \vec{\theta}(T)' \right)' \in \mathbb{R}^{T(p+q)} \text{ such that (A1) is verified} \right\}.$$

For $\vec{\alpha} \in \Omega$, let $\epsilon_{nT+\nu}(\vec{\alpha})$ be the periodically second-order stationary solution of

$$\epsilon_{nT+\nu}(\vec{\alpha}) = X_{nT+\nu} - \sum_{i=1}^{p} \phi_i(\nu)X_{nT+\nu-i} + \sum_{j=1}^{q} \theta_j(\nu)\epsilon_{nT+\nu-j}(\vec{\alpha}). \qquad (3.7)$$

Note that, almost surely, $\epsilon_{nT+\nu}(\vec{\alpha}_0) = \epsilon_{nT+\nu}$ for all $n \in \mathbb{Z}$ and $\nu \in \{1, ..., T\}$. Moreover, $\epsilon_{nT+\nu}(\vec{\alpha})$ can be approximated by $e_{nT+\nu}(\vec{\alpha})$ which is also determined recursively in t via a truncated version of (3.7)

$$e_{nT+\nu}(\vec{\alpha}) = X_{nT+\nu} - \sum_{i=1}^{p} \phi_i(\nu)X_{nT+\nu-i} + \sum_{j=1}^{q} \theta_j(\nu)e_{nT+\nu-j}(\vec{\alpha}), \qquad (3.8)$$

where the unknown starting values are set to zero: $e_0(\vec{\alpha}) = ... = e_{1-q}(\vec{\alpha}) = X_0 = ... = X_{1-p} = 0$.

Let δ be a strictly positive constant chosen such that $\vec{\alpha}_0$ belongs to the interior of the compact set

$$\Omega_\delta = \left\{ \vec{\alpha} \in \mathbb{R}^{T(p+q)} | \text{the zeros of } \det\{\Phi(z)\} \text{ and those of } \det\{\Theta(z)\} \text{ have modulus} \geq 1 + \delta \right\}.$$

The random variable $\hat{\vec{\alpha}}_{\text{OLS}}$ is called the ordinary least squares (OLS) estimator of $\vec{\alpha}$ if it satisfies, almost surely,

$$S_N(\hat{\vec{\alpha}}_{\text{OLS}}) = \min_{\vec{\alpha} \in \Omega_\delta} S_N(\vec{\alpha}), \qquad (3.9)$$

where

$$S_N(\vec{\alpha}) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} e_{nT+\nu}^2(\vec{\alpha}). \qquad (3.10)$$

Because of the presence of heteroscedastic innovations, the OLS estimator might be inefficient. We will see that, for some vectors of weights $\vec{\omega}^2 = (\omega_1^2, ..., \omega_T^2)'$, the OLS estimator is asymptotically outperformed by the weighted least squares (WLS) estimator $\hat{\vec{\alpha}}_{\text{WLS}} = \hat{\vec{\alpha}}_{\text{WLS}}^{\vec{\omega}^2}$ defined by

$$Q_N^{\vec{\omega}^2}(\hat{\vec{\alpha}}_{\text{WLS}}) = \min_{\vec{\alpha} \in \Omega_\delta} Q_N^{\vec{\omega}^2}(\vec{\alpha}) \qquad (3.11)$$

where

$$Q_N^{\vec{\omega}^2}(\vec{\alpha}) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} \omega_\nu^{-2} e_{nT+\nu}^2(\vec{\alpha}). \qquad (3.12)$$

We will also see that an optimal WLS estimator is the generalized least squares (GLS) estimator

$$\hat{\vec{\alpha}}_{\text{GLS}} = \hat{\vec{\alpha}}_{\text{WLS}}^{\vec{\sigma}_0^2}. \qquad (3.13)$$

The GLS estimator assumes that $\vec{\sigma}_0^2$ is known. In practice, this parameter has also to be estimated. Given any consistent estimator $\hat{\vec{\sigma}}^2$ of $\vec{\sigma}_0^2$, a quasi-generalized least squares (QLS) estimator of $\vec{\alpha}_0$ is defined by

$$\hat{\vec{\alpha}}_{\text{QLS}} = \hat{\vec{\alpha}}_{\text{WLS}}^{\hat{\vec{\sigma}}^2}. \qquad (3.14)$$

One possible consistent estimator of σ_ν^2 is

$$\hat{\sigma}_\nu^2 = \frac{1}{N} \sum_{n=0}^{N-1} e_{nT+\nu}^2(\hat{\vec{\alpha}}_{\mathrm{OLS}}).$$

To establish the consistency of the least squares estimators, an additional assumption is needed.

(A2) The T-dimensional white noise $\left\{\boldsymbol{\epsilon}_n = (\epsilon_{nT+1}, ..., \epsilon_{nT+\nu}, ..., \epsilon_{nT+T})', n \in \mathbb{Z}\right\}$ in (3.6) is strictly stationary and ergodic.

Theorem 3.1 *Suppose that $\{X_{nT+\nu}\}$ is a $PARMA_T(p,q)$ process. Let $\hat{\vec{\alpha}}_{\mathrm{OLS}}$, $\hat{\vec{\alpha}}_{\mathrm{WLS}}$, $\hat{\vec{\alpha}}_{\mathrm{GLS}}$ and $\hat{\vec{\alpha}}_{\mathrm{QLS}}$ be the least squares estimators defined by (3.9), (3.11), (3.13) and (3.14). Then, under Assumptions (A1) and (A2), and for any $\vec{\omega}^2 = (\omega_1^2, \ldots, \omega_T^2) > 0$, where the inequality applies element-wise, we have*

$$\hat{\vec{\alpha}}_{\mathrm{OLS}} \to \vec{\alpha}_0, \quad \hat{\vec{\alpha}}_{\mathrm{WLS}}^{\vec{\omega}^2} \to \vec{\alpha}_0, \quad \hat{\vec{\alpha}}_{\mathrm{GLS}} \to \vec{\alpha}_0, \quad \hat{\vec{\alpha}}_{\mathrm{QLS}} \to \vec{\alpha}_0, \qquad \textit{almost surely as } N \to \infty.$$

Let $\mathcal{F}_{-\infty}^m$ and $\mathcal{F}_{m+h}^{+\infty}$ be the σ-fields generated by $\{\mathbf{X}_n, n \leq m\}$ and $\{\mathbf{X}_n, n \geq m+h\}$ respectively. The strong mixing coefficients of the T-variate stationary process $\{\mathbf{X}_n, n \in \mathbb{Z}\}$ are defined by

$$\underline{\alpha}_{\mathbf{X}}(h) = \sup_{A \in \mathcal{F}_{-\infty}^m, B \in \mathcal{F}_{m+h}^{+\infty}} |P(A \cap B) - P(A)P(B)|.$$

Let $||\mathbf{Z}||_r = [\mathbb{E}||\mathbf{Z}||^r]^{1/r}$ where $||.||$ stands for the Euclidean norm of a vector. In addition to Assumptions (A1) and (A2), we need the following assumption to establish the asymptotic normality of the least squares estimators previously introduced.

(A3) The T-variate stationary process $\{\mathbf{X}_n\}$ is such that for some $\tau > 0$, $||\mathbf{X}_n||_{4+2\tau} < \infty$ and $\sum_{h=0}^\infty [\underline{\alpha}_{\mathbf{X}}(h)]^{\frac{\tau}{2+\tau}} < \infty$.

Notice that Assumption (A3) does not require that the noise $\{\epsilon_t\}$ be strong or a martingale difference. The mixing condition is valid for large classes of processes. The moment condition is relatively mild given that the existence of $\mathbf{I}(\vec{\alpha}, \vec{\omega}^2)$ and $\mathbf{J}(\vec{\alpha}, \vec{\omega}^2)$ defined below requires $||\mathbf{X}_n||_4 < \infty$.

Theorem 3.2 *Under the assumptions of Theorem 3.1 and (A3), as $N \to \infty$,*

$$\sqrt{N}(\hat{\vec{\alpha}}_{\mathrm{LS}} - \vec{\alpha}_0) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{V}_{\mathrm{LS}})$$

where the subscript LS stands for OLS, WLS, GLS or QLS, and where

$$\mathbf{V}_{\mathrm{OLS}} = \mathbf{V}\left(\vec{\alpha}_0, (1, \ldots, 1)'\right), \quad \mathbf{V}_{\mathrm{WLS}} = \mathbf{V}(\vec{\alpha}_0, \vec{\omega}^2), \quad \mathbf{V}_{\mathrm{GLS}} = \mathbf{V}_{\mathrm{QLS}} = \mathbf{V}(\vec{\alpha}_0, \vec{\sigma}_0^2),$$

with

$$\mathbf{V}(\vec{\alpha}_0, \vec{\omega}^2) = \left(\mathbf{J}(\vec{\alpha}_0, \vec{\omega}^2)\right)^{-1} \mathbf{I}(\vec{\alpha}_0, \vec{\omega}^2) \left(\mathbf{J}(\vec{\alpha}_0, \vec{\omega}^2)\right)^{-1}, \tag{3.15}$$

$$\mathbf{I}(\vec{\alpha}_0, \vec{\omega}^2) = \sum_{\nu=1}^{T} \sum_{\nu'=1}^{T} \omega_\nu^{-2} \omega_{\nu'}^{-2} \sum_{k=-\infty}^{\infty} \mathbb{E}\left[\left(\epsilon_\nu(\vec{\alpha}_0) \left(\frac{\partial \epsilon_\nu(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \right) \left(\epsilon_{kT+\nu'}(\vec{\alpha}_0) \left(\frac{\partial \epsilon_{kT+\nu'}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \right)' \right]$$

and

$$\mathbf{J}(\vec{\alpha}_0, \vec{\omega}^2) = \sum_{\nu=1}^{T} \omega_\nu^{-2} \mathbb{E}\left[\left(\frac{\partial \epsilon_\nu(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \left(\frac{\partial \epsilon_\nu(\vec{\alpha})}{\partial \vec{\alpha}} \right)'_{\vec{\alpha}=\vec{\alpha}_0} \right].$$

Remark 3.1 In the periodic AR case, the OLS and WLS estimators coincide. Indeed we have in this particular case $\vec{\alpha} = (\vec{\phi}(1)', \dots, \vec{\phi}(T)')'$ and

$$Q_N^{\vec{\omega}^2}(\vec{\alpha}) = \sum_{\nu=1}^{T} \omega_\nu^{-2} \frac{1}{N} \sum_{n=0}^{N-1} e_{nT+\nu}^2(\vec{\phi}(\nu)).$$

Thus the WLS estimator does not depend on the vector of weights $\vec{\omega}^2$:

$$\hat{\vec{\alpha}} = (\hat{\vec{\phi}}(1)', \dots, \hat{\vec{\phi}}(T)')', \quad \text{where } \hat{\vec{\phi}}(\nu) = \arg\min_{\vec{\phi}(\nu)} \frac{1}{N} \sum_{n=0}^{N-1} e_{nT+\nu}^2(\vec{\phi}(\nu)).$$

Notice however that it does not hold when $\vec{\mu} = (\mu_1, \dots, \mu_T)' \neq \vec{0}$. In the general PARMA case, the WLS estimator varies with $\vec{\omega}^2$ because $e_{nT+\nu}^2$ depends on the entire parameter $\vec{\alpha}$, and not only on $\left(\vec{\phi}(\nu), \vec{\theta}(\nu) \right)$, when the MA term is present.

Remark 3.2 In the strong PARMA setting, *i.e.*, when $\{\sigma_{0\nu}^{-1} \epsilon_{nT+\nu}\}$ is an independent and identically distributed sequence, the asymptotic covariance matrix of the QLS estimators takes a simple form. Indeed, independence of the $\epsilon_{nT+\nu}$'s implies that only the terms for $k = 0$ and $\nu = \nu'$ are non zero. Therefore, we obtain that

$$\mathbf{I}(\vec{\alpha}_0, \vec{\sigma}_0^2) = \sum_{\nu=1}^{T} \sigma_{0\nu}^{-2} \mathbb{E}\left[\left(\frac{\partial \epsilon_\nu(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \left(\frac{\partial \epsilon_\nu(\vec{\alpha})}{\partial \vec{\alpha}} \right)'_{\vec{\alpha}=\vec{\alpha}_0} \right] = \mathbf{J}(\vec{\alpha}_0, \vec{\sigma}_0^2).$$

This implies that the asymptotic covariance matrix of the QLS estimators for a PARMA model with independent errors is

$$\mathbf{V}_{\text{QLS}} = \left(\mathbf{J}(\vec{\alpha}_0, \vec{\sigma}_0^2) \right)^{-1}. \tag{3.16}$$

This result was obtained by Basawa and Lund (2001). Moreover, applying Theorem 3.2 in the weak ARMA setting, *i.e.*, when the period T is equal to one $(T = 1)$, we retrieve the result obtained by Francq and Zakoïan (1998).

Remark 3.3 If $\{\epsilon_t\}$ is a martingale difference such that $\mathbb{E}\left[\epsilon_t^2 \mid \mathcal{F}_{t-1}\right] = \mathbb{E}\left[\epsilon_t^2\right]$ where $\mathcal{F}_t$ is the σ-field spanned by $\{\epsilon_s, s \leq t\}$, the QLS estimator is an optimal LS estimator in the sense that

$$\mathbf{V}_{\text{WLS}} - \mathbf{V}_{\text{QLS}} \text{ is a semi-definite positive matrix.}$$

Indeed, let the random vector

$$S_N^{\vec{\omega}^2} = \left\{\mathbf{J}(\vec{\alpha}_0, \vec{\omega}^2)\right\}^{-1} \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} \omega_\nu^{-2} \epsilon_{nT+\nu}(\vec{\alpha}_0) \left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}}\right)_{\vec{\alpha}=\vec{\alpha}_0}.$$

We have

$$\mathrm{Cov}\left(S_N^{\vec{\sigma}^2}, S_N^{\vec{\omega}_0^2}\right)$$

$$= \left\{\mathbf{J}(\vec{\alpha}_0, \vec{\omega}^2)\right\}^{-1} \frac{1}{N} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} \omega_\nu^{-2} \sigma_{0\nu}^{-2} \mathbb{E}\epsilon_{nT+\nu}^2 \mathbb{E}\left[\left(\frac{\partial \epsilon_\nu(\vec{\alpha})}{\partial \vec{\alpha}}\right)_{\vec{\alpha}=\vec{\alpha}_0} \left(\frac{\partial \epsilon_\nu(\vec{\alpha})}{\partial \vec{\alpha}}\right)'_{\vec{\alpha}=\vec{\alpha}_0}\right] \left\{\mathbf{J}(\vec{\alpha}_0, \vec{\sigma}_0^2)\right\}^{-1}$$

$$\rightarrow \left\{\mathbf{J}(\vec{\alpha}_0, \vec{\sigma}_0^2)\right\}^{-1}.$$

Therefore

$$\lim_{N\to\infty} \mathrm{Var}\left(S_N^{\vec{\omega}^2} - S_N^{\vec{\sigma}_0^2}\right) = \left\{\mathbf{J}(\vec{\alpha}_0, \vec{\omega}^2)\right\}^{-1} - \left\{\mathbf{J}(\vec{\alpha}_0, \vec{\sigma}_0^2)\right\}^{-1} = \mathbf{V}_{\mathrm{WLS}} - \mathbf{V}_{\mathrm{QLS}}$$

and the conclusion follows.

Remark 3.4 It can be shown that $\mathbf{J}(\vec{\alpha}_0, \vec{\omega}^2)$ is consistently estimated by the empirical mean

$$\hat{\mathbf{J}}(\vec{\alpha}_0, \vec{\omega}^2) = \sum_{\nu=1}^{T} \omega_\nu^{-2} \frac{1}{N} \sum_{n=0}^{N-1} \left[\left(\frac{\partial e_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}}\right)_{\vec{\alpha}=\hat{\vec{\alpha}}_{\mathrm{LS}}} \left(\frac{\partial e_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}}\right)'_{\vec{\alpha}=\hat{\vec{\alpha}}_{\mathrm{LS}}}\right].$$

Note that the matrix $(2\pi)^{-1}\mathbf{I}(\vec{\alpha}_0, \vec{\omega}^2)$ is the spectral density at frequency zero of the process

$$\Upsilon_n = \sum_{\nu=1}^{T} \omega_\nu^{-2} \epsilon_{nT+\nu}(\vec{\alpha}_0) \left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}}\right)_{\vec{\alpha}=\vec{\alpha}_0}.$$

Estimators of such long-run variances are available in the literature (see *e.g.* den Haan and Levin (1997) for a general reference). For the numerical illustrations presented in this paper, we used a VAR spectral estimator consisting in: i) fitting VAR(p) models for $p = 0, \ldots, p_{\max}$ to the series $\hat{\Upsilon}_n$, $n = 0, \ldots, N-1$, where $\hat{\Upsilon}_n$ is obtained by replacing $\epsilon_{nT+\nu}(\vec{\alpha}_0)$) and its derivatives by $e_{nT+\nu}(\hat{\vec{\alpha}}_{\mathrm{LS}})$ and its derivatives in Υ_n; ii) selecting the order p which minimizes an information criteria and approximating $\mathbf{I}(\vec{\alpha}_0, \vec{\omega}^2)$ by (2π) times the spectral density at frequency zero of the estimated VAR(p) model. Hereafter, we used the AIC model selection criterion with $p_{\max} = 25$.

Remark 3.5 Most of the descriptive techniques for identifying a strong PAR model, as described in McLeod (1994) and Hipel and McLeod (1994), remain valid for a weak PAR model. We have seen in Remark 3.1 that the LS estimator of $\vec{\phi}(\nu)$ in a weak PAR model with means zero only involves the sequence $\{e_{nT+\nu}, n = 0, \ldots, N-1\}$. Therefore, the sample ACF and PACF of the ν-th season of the original series can be used to identify the AR order p_ν. Valid significance limits for ACF and

PACF in the weak case were obtained by Romano and Thombs (1996). See also Berlinet and Francq (1997) and Francq and Zakoïan (2009). The popular AIC and BIC model selection criteria can also be applied to each season. Francq and Zakoïan (1998b) showed that asymptotically, the orders of a weak ARMA model are not underestimated when these criteria are employed. As in the strong case, note however that these techniques do not work for a general PARMA model.

Also with PAR models, the significance limits for residual ACF and the modified Ljung-box test described in Francq *et al.* (2005) can be applied at each season either for testing the hypothesis of weak white noise of the observed periodic series or for checking the validity of the estimated model. A global goodness-of-fit test would be welcomed but it is beyond the scope of this paper.

4   Examples of covariance matrix calculations

The asymptotic covariance matrix of the QLS estimators obtained under independent errors is generally different from the one obtained under uncorrelated but dependent errors. Here, we give explicit expressions for the asymptotic covariance of the QLS estimator of a weak $\mathrm{PAR}_2(1)$ model for two different weak white noises. In both cases, it is seen that the difference with the asymptotic covariance matrix under the assumption of a strong noise can be huge.

4.1   Example 1

Consider the weak periodic white noise of section 2.1.1 with $T = 2$:

$$\epsilon_{nT+\nu} = \sigma_\nu \prod_{j=0}^{m} \xi_{nT+\nu-j}, \qquad \nu = 1, 2, \tag{4.1}$$

and assume that the iid sequence $\{\xi_t\}$ has a finite fourth-order moment $\kappa = \mathbb{E}\left[\xi_t^4\right]$.

¿From a realization $X_t = \epsilon_t, t = 1, ..., NT$, of that weak white noise, suppose that a statistician fits the following $\mathrm{PAR}_2(1)$ model:

$$\begin{cases} X_{nT+1} - \phi_1 X_{nT} = \epsilon_{nT+1}, \\ X_{nT+2} - \phi_2 X_{nT+1} = \epsilon_{nT+2}. \end{cases} \tag{4.2}$$

The true parameter values are $\phi_1 = \phi_2 = 0$ and $\vec{\sigma}^2 = \vec{\sigma}_0^2$. According to Theorem 3.2, some moment calculations show that the asymptotic covariance matrix of the QLS estimator of $\sqrt{N}\left(\hat{\phi}_1, \hat{\phi}_2\right)'$ is given by

$$\mathbf{V}_{QLS}^{(w)} = \left(\mathbf{J}(\vec{\alpha}_0, \vec{\sigma}_0^2)\right)^{-1} \mathbf{I}(\vec{\alpha}_0, \vec{\sigma}_0^2) \left(\mathbf{J}(\vec{\alpha}_0, \vec{\sigma}_0^2)\right)^{-1} = \kappa^m \begin{pmatrix} \frac{\sigma_{01}^2}{\sigma_{02}^2} & 0 \\ 0 & \frac{\sigma_{02}^2}{\sigma_{01}^2} \end{pmatrix}. \tag{4.3}$$

On the other hand from (3.16), the corresponding asymptotic covariance matrix under the assumption of a strong noise is equal to

$$\mathbf{V}_{QLS}^{(s)} = \left(\mathbf{J}(\vec{\alpha}_0, \vec{\sigma}_0^2)\right)^{-1} = \begin{pmatrix} \frac{\sigma_{01}^2}{\sigma_{02}^2} & 0 \\ 0 & \frac{\sigma_{02}^2}{\sigma_{01}^2} \end{pmatrix}. \tag{4.4}$$

It is clear that $\mathbf{V}_{QLS}^{(w)}$ and $\mathbf{V}_{QLS}^{(s)}$ can be very different. For example, if the iid sequence $\{\xi_t\}$ is $\mathcal{N}(0,1)$, $\mathbb{E}\left[\xi_t^4\right] = 3$, and the discrepancy between the two matrices is important even for small m. It may lead the statistician to wrongly reject the hypothesis that $\phi_1 = \phi_2 = 0$ if he does not take into account the dependence of the errors ϵ_t.

4.2 Example 2

To illustrate the influence of the kind of dependence on the asymptotic covariance matrix of the LS estimators, let us go back to the weak periodic white noise of section 2.1.1 with $T = 2$:

$$\epsilon_{nT+\nu} = \sigma_\nu \left(\eta_{nT+\nu} / \sqrt{\mathbb{E}\left[\eta_{nT+\nu}^2\right]}\right), \tag{4.5}$$

where $\{\eta_t\}$ is the causal solution of the following ARCH(1) model

$$\eta_t = \sqrt{1 + \alpha\eta_{t-1}^2}\,\xi_t, \qquad t \in \mathbb{Z},$$

with $0 < \alpha < 1$ and $\{\xi_t\}$ is a sequence of symmetric and centered iid random variables with unit variance and having finite fourth-order moment. Let $\kappa = \mathbb{E}\left[\xi_t^4\right]$ and assume that $\kappa > 1$ and $0 < \kappa\alpha^2 < 1$.

Given a realization $X_t = \epsilon_t, t = 1, ..., NT$, from (4.5), let us estimate the PAR$_2$(1) model (4.2) for that series. Once again, the true parameter values are $\phi_1 = \phi_2 = 0$ and $\vec{\sigma}^2 = \vec{\sigma}_0^2$. Direct computation of the matrices in (3.15) lead to the following asymptotic covariance matrix of the QLS estimators of $\sqrt{N}\left(\hat{\phi}_1, \hat{\phi}_2\right)'$:

$$\mathbf{V}_{QLS}^{(w)} = \left(\mathbf{J}(\vec{\alpha}_0, \vec{\sigma}_0^2)\right)^{-1} \mathbf{I}(\vec{\alpha}_0, \vec{\sigma}_0^2) \left(\mathbf{J}(\vec{\alpha}_0, \vec{\sigma}_0^2)\right)^{-1} = \left\{1 + \frac{(\kappa - 1)\alpha}{1 - \kappa\alpha^2}\right\} \begin{pmatrix} \frac{\sigma_{01}^2}{\sigma_{02}^2} & 0 \\ 0 & \frac{\sigma_{02}^2}{\sigma_{01}^2} \end{pmatrix}. \tag{4.6}$$

It is obvious that (4.6) can be quite different from the asymptotic covariance matrix (4.4) corresponding to a strong white noise. For example, if $\{\xi_t\}$ is an iid $\mathcal{N}(0,1)$ sequence, and $\alpha = 0.5, \mathbf{V}_{QLS}^{(w)} = 5\mathbf{V}_{QLS}^{(s)}$.

With a series of 2000 observations ($N = 1000$) and if $\sigma_{01}^2 = \sigma_{02}^2$, the standard error of the QLS estimators of ϕ_1 and ϕ_2 is 0.0707 for that weak white noise whilst it is equal to 0.0316 under the assumption of a strong noise.

5 Some Monte Carlo results

The aim of this Monte Carlo study is to underline that the standard inference procedures developed for strong PARMA models can be quite misleading when analyzing data generated by weak PARMA models. In the first part of this simulation experiment, we considered various white noises (strong or weak) to which we fitted a $PAR_2(1)$ model. The discrepancy between the empirical standard errors of the parameter estimators and their asymptotic standard errors under the assumption of a strong noise is examined. The size distortion of a Wald test for the hypothesis that the model coefficients are zero, based on the asymptotic covariance matrix under the assumption of a strong noise, is also investigated. In the second part, two different $PARMA_2(1,1)$ models with strong and weak noises were used to investigate the size and power of a Wald test based on a consistent estimator of the asymptotic covariance matrix under the assumption of either a weak or strong noise. The rate of convergence of the estimated asymptotic standard errors is also analysed.

5.1 Using the theoretical covariance structure

To make the presentation easier, we again consider the $PAR_2(1)$ model (4.2) with three different types of periodic white noises $\{\epsilon_{nT+\nu}\}$.

1. *Type 1:* The periodic white noise $\{\epsilon_t\}$ is assumed to be a sequence of independent random variables. More precisely, the random variables $\sigma_\nu^{-1}\epsilon_{nT+\nu}$ are independent and identically distributed with mean zero and variance one. We consider the following distributions for $\sigma_\nu^{-1}\epsilon_{nT+\nu}$: standard Normal ($\mathcal{N}$), Student with 3 degrees freedom (t), Lognormal with parameters (4,1) (LN), Chi-square with 1 degree freedom (χ^2), Exponential with parameter one (Exp), and Gamma with parameters (5,1) (Gam). When necessary, each of these six distributions have been normalized to obtain distributions with mean equal to zero and variance equal to one.

2. *Type 2:* Here $\{\epsilon_t\}$ is the periodic weak white noise defined by (2.2) and we considered two particular cases: $m = 1$ (WN1) and $m = 2$ (WN2).

3. *Type 3:* This is the periodic ARCH white noise defined by (4.5). For this type of weak white noise, we considered three values for $\alpha : 0.3, 0.4$, and 0.5.

For each of these eleven different periodic white noises, 1000 replications of length $(N + 200) \times 2$ were generated. These sequences were plugged into the $PAR_2(1)$ model (4.2), yielding 1000 replications of the periodic process $\{X_t\}$ of length $(N + 200) \times 2$. Initial values were set to zero and in order to achieve periodic stationarity, the first 400 observations were dropped. For each replication of length

$NT = N \times 2$, the $PAR_2(1)$ model (4.2) was estimated by ordinary least squares. As pointed out in Remark 3.1, the OLS and WLS estimators coincide in the periodic AR case. The OLS estimators of ϕ_1 and ϕ_2 are denoted $\hat{\phi}_1$ and $\hat{\phi}_2$. In the experiment, we considered the three values of N: 50, 100, 1000.

We carried out simulations for different set of values of ϕ_1, ϕ_2, σ_1 and σ_2. However, for sake of brevity, we only present the results for $\phi_1 = \phi_2 = 0$ and $\sigma_1 = \sigma_2 = 1.0$. For each value of N and for each type of periodic white noise, we report, in Table 1, the empirical standard errors $S_{\hat{\phi}_1}$ and $S_{\hat{\phi}_2}$ of $\hat{\phi}_1$ and $\hat{\phi}_2$ respectively, based on 1000 replications. The mean values of the estimates $\bar{\hat{\phi}}_1$ and $\bar{\hat{\phi}}_2$ are not reported since they are always very close to the true values $\phi_1 = \phi_2 = 0$. The empirical standard errors should be compared to the corresponding asymptotic standard errors $\sqrt{V(\hat{\phi}_1)}$, for either a strong noise as given by (3.16) or for a weak noise as provided by (3.15). For the $PAR_2(1)$ model chosen, the formulas of the asymptotic variances of the LS estimators corresponding to the three types of white noises considered are provided respectively by (4.4), (4.3) and (4.6). Moreover, for each type of periodic white noise, and for each replication we test the following null hypotheses $H_0^{(1)} : \phi_1 = 0$, $H_0^{(2)} : \phi_2 = 0$, and $H_0 : \phi_1 = \phi_2 = 0$ using the Wald test computed under the assumption of a strong $PAR_2(1)$ model. The rejection frequencies in 1000 replications at significance level 5% are reported in Table 1.

Inspection of Table 1 reveals that the Asymptotic Standard Errors (ASE) are reasonably close to the Empirical Standard Errors (ESE) even for short series ($N = 50, 100$) except for the 2-dependent noise WN2 and the ARCH(1) noise with $\alpha = 0.5$. For long series ($N = 1000$), ASE and ESE are very close for all the white noises considered. Also, comparison of the ESE's for types 2 and 3 noises with the ASE's for a strong noise (ASE_s) shows that the ESE's are considerably larger than ASE_s. Even for short series, the dependence of the errors considerably inflate the true standard error. For instance, with WN2, the ESE is more than twice the corresponding ASE_s and with the ARCH(1) noise with $\alpha = 0.5$, the ratio is greater than 1.5. This remark also hold for $N = 1000$.

Perusal of the rejection frequencies of the standard Wald test for the three series lengths shows that the size of the test is considerably affected by the dependence of the errors. Based on 1000 replications, the standard error of the rejection frequencies at the nominal level 0.05 is 0.0069. With the strong noises, the size of the Wald test is reasonably well controlled. For N=50, we observe a slight tendency to overreject, but all frequencies except 2 are within 3 standard errors from the nominal level 0.05. At N=100, 1000, all the rejection frequencies, except one for each series length, are within 2 standard errors from 0.05. The standard Wald test clearly overreject with the weak noises considered. The smallest rejection frequency 0.12 is obtained with the ARCH(1) noise when $\alpha = 0.3$. For the

Type 2 noises and the Type 3 with $\alpha = 0.5$, all the rejection frequencies are greater than 0.2.

This small simulation experiment clearly shows that the dependence of the errors invalidates the standard inference procedures developed for PARMA models with independent errors.

	Type 1						Type 2		Type 3		
	$\mathcal{N}$	t	LN	χ^2	Exp	Gam	WN1	WN2	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$
$NT = 100$											
Empirical standard errors											
$S_{\hat{\phi}_1}$	.1487	.1671	.1462	.1504	.1548	.1526	.2355	.3434	.1783	.1984	.2265
$S_{\hat{\phi}_2}$	.1406	.1683	.1447	.1376	.1399	.1504	.2362	.3249	.1863	.1984	.2192
Asymptotic standard errors ($i = 1, 2$)											
$\sqrt{V(\hat{\phi}_i)}$	.1414	.1414	.1414	.1414	.1414	.1414	.2449	.4243	.1909	.2253	.3162
Rejection frequencies in 1000 replications of the null hypothesis											
$H_0^{(1)}$	.0640	.0590	.0670	.0680	.0740	.0740	.2580	.4100	.1280	.1670	.2200
$H_0^{(2)}$	.0640	.0700	.0610	.0460	.0520	.0590	.2610	.3940	.1590	.1680	.2070
H_0	.0640	.0660	.0690	.0580	.0680	.0680	.3830	.5650	.1820	.2310	.3060
$NT = 200$											
Empirical standard errors											
$S_{\hat{\phi}_1}$	.1013	.1130	.1025	.0998	.0992	.0991	.1770	.2531	.1294	.1380	.1616
$S_{\hat{\phi}_2}$	.0994	.1126	.1004	.1016	.1002	.1001	.1676	.2506	.1271	.1466	.1570
Asymptotic standard errors ($i = 1, 2$)											
$\sqrt{V(\hat{\phi}_i)}$	.1000	.1000	.1000	.1000	.1000	.1000	.1732	.3000	.1350	.1593	.2236
Rejection frequencies in 1000 replications of the null hypothesis											
$H_0^{(1)}$	.0570	.0540	.0620	.0520	.0480	.0490	.2830	.4110	.1370	.1560	.2190
$H_0^{(2)}$	.0560	.0630	.0530	.0590	.0500	.0630	.2340	.4290	.1270	.1770	.2000
H_0	.0580	.0650	.0550	.0520	.0560	.0510	.3810	.6140	.1670	.2260	.2960
$NT = 2000$											
Empirical standard errors											
$S_{\hat{\phi}_1}$	.0311	.0320	.0309	.0330	.0324	.0314	.0529	.0918	.0430	.0498	.0612
$S_{\hat{\phi}_2}$	.0322	.0321	.0320	.0314	.0319	.0310	.0542	.0925	.0411	.0500	.0609
Asymptotic standard errors ($i = 1, 2$)											
$\sqrt{V(\hat{\phi}_i)}$	.0316	.0316	.0316	.0316	.0316	.0316	.0548	.0948	.0427	.0504	.0707
Rejection frequencies in 1000 replications of the null hypothesis											
$H_0^{(1)}$	.0400	.0510	.0450	.0540	.0580	.0550	.2460	.4840	.1510	.2020	.2750
$H_0^{(2)}$	.0620	.0460	.0510	.0470	.0630	.0360	.2530	.4920	.1200	.2120	.2820
H_0	.0510	.0450	.0480	.0520	.0530	.0460	.3560	.6870	.1730	.2840	.4040

Table 1. Empirical and asymptotic standard errors of the OLS estimators of ϕ_1 and ϕ_2 in the PAR$_2$(1) model (4.2) and the rejection frequencies at the nominal level 0.05 of the standard Wald test computed under independent errors for three types of white noises and three series lengths NT. The number of replications is 1000.

5.2 Using an estimated covariance structure

In this second experiment, we used the $\text{PARMA}_2(1,1)$ model defined by

$$X_{nT+\nu} = \mu_\nu + \phi_\nu X_{nT+\nu-1} + \epsilon_{nT+\nu} - \theta_\nu \epsilon_{nT+\nu-1}, \quad \nu = 1,2, \tag{5.1}$$

with various sets of parameter values. Including means, the vector of parameters becomes $\vec{\alpha} = (\mu_1, \mu_2, \phi_1, \phi_2, \theta_1, \theta_2)'$. For each realisation, we computed the QLS estimator of $\vec{\alpha}$ and its estimated asymptotic covariance matrix using the method described in Remark (3.4) under both the assumptions of a strong noise (denoted by S) and a weak noise (denoted by W). The strong noise is an iid $N(0,1)$ sequence while the weak noise in a 3-dependent process as defined by Equation (2.2) with $m = 3$, the ξ_t are iid $N(0,1)$ random variables and $\sigma_1 = \sigma_2 = 1$. For each of these two noises, the PARMA realisations were obtained by plugging in (5.1) the noise series. In all cases, the initial values were set to zero and for each realisation of length NT, $NT + 400$ data were generated and the first 400 were discarded. For each of the three series lengths (500, 1000 and 2000), the QLS estimator of $\vec{\alpha}$ was obtained as well as the estimated standard errors under both the assumptions of a strong noise and a weak noise. From the 1000 independent realisations, various statistics for studying the variability and the rate of convergence are reported in Table 2, namely the biais, the Empirical Standard Error (ESE), the Mean Asymptotic Standard Errors under the assumption of a strong noise ($\text{MASE}^{(s)}$) and a weak noise ($\text{MASE}^{(w)}$). These results correspond to the parameter values $\vec{\alpha} = (0.05, -0.05, 0.8, 0.75, -0.5, -0.45)'$.

Here are some comments about these numerical results. The bias is negligible for all parameters and the three series lengths. In the case of a strong noise, $\text{MASE}^{(s)}$ and $\text{MASE}^{(w)}$ are close to each other and also reasonably close to the corresponding ESE for all parameters but μ_2. For this latter parameter, both asymptotic standard errors underestimate the true standard error even with 2000 observations. With a weak noise, $\text{MASE}^{(s)}$ is quite far from the corresponding ESE for all parameters except μ_1. $\text{MASE}^{(w)}$ provides a rather poor approximation of the true standard error when $NT = 500$. As expected, the approximation improves as the series length increases and it is rather satisfactory with 2000 observations. For all parameters but μ_1, $\text{MASE}^{(w)}$ is much closer to ESE than $\text{MASE}^{(s)}$.

Table 3 gives the rejection frequencies of the Wald statistic for testing the null hypothesis H_0 of absence of seasonality in a $\text{PARMA}_2(1,1)$ model. It is equivalent to testing that $(\mu_1, \phi_1, \theta_1)' = (\mu_2, \phi_2, \theta_2)'$. For the null hypothesis, we used the parameter values $\vec{\alpha} = (0, 0, 0.8, 0.8, -0.5, -0.5)'$. Under the alternative hypothesis H_1, we employed the same parameter values as in the first part of this experiment, that is $\vec{\alpha} = (0.05, -0.05, 0.8, 0.75, -0.5, -0.45)'$. We also considered alternatives further away from the null but the results are reported because too many rejection frequencies were

equal to 100%. The Wald test under the assumption of a strong noise is denoted WS while WW represents the Wald test under a weak noise. Based on 1000 realisations, the standard error of the rejection frequencies at the nominal levels 1, 5 and 10% are respectively 0.31, 0.69 and 0.95%.

NT	Noise	Statistics	μ_1	μ_2	ϕ_1	ϕ_2	θ_1	θ_2
500	S	Bias	.007	.007	-.007	-.006	-.013	-.003
		ESE	.258	.235	.040	.031	.103	.042
		MASE$^{(s)}$	.281	.137	.044	.023	.120	.035
		MASE$^{(w)}$	.307	.149	.043	.023	.119	.035
	W	Bias	-.018	-.015	-.008	-.005	-.022	.001
		ESE	1.218	1.141	.072	.053	.269	.110
		MASE$^{(s)}$	.240	.119	.037	.020	.101	.030
		MASE$^{(w)}$	.245	.121	.040	.026	.270	.070
1000	S	Bias	.000	.000	-.005	-.004	-.010	-.002
		ESE	.173	.158	.026	.021	.072	.030
		MASE$^{(s)}$	.204	.099	.031	.016	.085	.025
		MASE$^{(w)}$	.216	.105	.031	.016	.084	.025
	W	Bias	.003	.004	-.008	-.005	-.015	.004
		ESE	.157	.144	.055	.041	.222	.089
		MASE$^{(s)}$	.184	.091	.028	.015	.077	.023
		MASE$^{(w)}$	.190	.094	.052	.023	.233	.061
2000	S	Bias	.002	.002	-.003	-.002	.000	-.001
		ESE	.116	.107	.019	.015	.049	.021
		MASE$^{(s)}$	.146	.071	.022	.011	.060	.018
		MASE$^{(w)}$	.151	.073	.021	.011	.060	.018
	W	Bias	-.006	-.005	-.004	-.002	-.016	.004
		ESE	.113	.104	.044	.032	.170	.072
		MASE$^{(s)}$	.139	.068	.021	.011	.057	.017
		MASE$^{(w)}$	.143	.070	.042	.018	.192	.050

Table 2. Bias, Empirical Standard Error (ESE), Mean Asymptotic Standard Error under the assumption of a strong noise (MASE$^{(s)}$) and a weak noise (MASE$^{(w)}$) of the QLS estimators of the PARMA$_2(1,1)$ model (5.1) with $\vec{\alpha} = (0.05, -0.05, 0.8, 0.75, -0.5, -0.45)'$, based on 1000 independent realisations for each series length NT.

For strong PARMA series, all the rejection frequencies with WS and WW are within two standard errors of the corresponding nominal levels except one which is between 2 and 3 standard errors. Therefore, the levels of both tests are well controlled, even with series of 500 observations. Also, the power of both tests are almost identical. At least for this model, there is nothing to loose in applying WW if we have doubts about the nature of the noise (strong or not).

When the noise is weak, the size of WS is out of control and considerably increases with NT. For example, at the 5% level, it varies from 42.9% to 59.1% as NT increases. In contrast, the size of WW decreases with NT and seems to converge to the true level. However, there is also a tendency to over

21

reject even with $NT = 2000$. For example, at 5%, the empirical level is 9.9%. Under H_1, we cannot say which test is more powerful since the empirical levels of WS and WW are too far apart. However, WW is clearly more powerful when the noise is strong than in the weak case.

This experiment illustrates the usefulness of the proposed estimator of the covariance structure of the QLS estimators of PARMA parameters in presence of possibly dependent errors.

Hypothesis	NT	Test Statistic	*Strong noise*			*Weak noise*		
			1	5	10	1	5	10
H_0	500	WS	1.2	4.1	7.8	37.1	47.9	54.2
		WW	1.5	5.0	8.4	9.7	18.9	25.3
	1000	WS	.6	4.3	9.5	41.7	51.9	57.8
		WW	.7	4.5	9.6	5.6	13.2	19.6
	2000	WS	.9	5.0	9.0	48.9	59.1	64.8
		WW	1.1	5.0	8.8	3.1	9.9	16.2
H_1	500	WS	6.9	21.9	34.4	53.0	66.2	73.9
		WW	8.1	23.9	37.2	34.6	49.2	57.0
	1000	WS	26.4	54.5	69.7	68.7	82.2	86.6
		WW	26.8	54.0	69.8	34.3	51.8	62.4
	2000	WS	78.3	95.4	98.5	87.3	93.8	95.9
		WW	78.4	95.4	98.1	44.7	65.3	75.4

Table 3. Rejection frquencies (%) at the nominal levels 1, 5 and 10% of the Wald test for the null hypothesis of non periodicity (H_0) and for a fixed alternative (H_1) under the assumption of a strong (WS) or a weak noise (WW) based on 1000 independent realisations for each series length NT of PARMA$_2(1,1)$ model (5.1). Under H_0, $\vec{\alpha} = (0, 0, 0.8, 0.8, -0.5, -0.5)'$ and under H_1, $\vec{\alpha} = (0.05, -0.05, 0.8, 0.75, -0.5, -0.45)'$.

6 Application to real data

In this section, we consider the daily returns of four European stock market indices: CAC 40 (Paris), DAX (Frankfurt), FTSE 120 (London), for the period from January 7, 1991 to July 3, 2009, and the index SMI (Swisserland), from November 12, 1990 to July 3, 2009. The number of observations varies between 4674 and 4692. The data were obtained from Yahoo Finance. Standard models for such financial series are weak white noises of the form $r_t = \sigma_t \eta_t$ where r_t is the log-return, $\{\eta_t\}$ is an iid noise with variance equal to 1, and σ_t^2 is the so-called volatility. For the GARCH-type models, σ_t is a measurable function of $\{r_s, s < t\}$.

In recent decades, many researchers addressed the question of day-of-the-week seasonality in stock markets, see among other Franses and Paap (2000, 2004), Balaban *et al.* (2001), Bollerslev and Ghysels (1996), Peiro (1994). Most of these studies focus on the description of day-of-the-week seasonality in

returns and volatility. In particular, it was observed that in many stock markets, the Monday returns are often lower than those of other days. In the finance literature, it is referred to as the Monday effect.

In order to analyse the seasonality of these four European indices, we fitted the following simple $PAR_5(1)$ model to each series:

$$r_{nT+\nu} = \mu_\nu + \phi_\nu r_{nT+\nu-1} + \epsilon_{nT+\nu}, \quad \nu = 1, \ldots, T = 5, \tag{6.1}$$

where $r_t = 100 ln(I_t/I_{t-1})$ is the log-return multiplied by 100 and I_t is the value of the index at time t. Because of the legal holidays, many weeks comprise less than five observations. However, we cannot talk of missing values because these variables do not exist on those days. For that reason, we preferred removing the entire weeks when there was less than five data available rather than estimating the "pseudo-missing" observations by an ad hoc method. The effective number of observations used in the analysis is given in Table 5. For each index, Model (6.1) was used to test the hypothesis of white noise (H_{01}) and the hypothesis of non seasonality (H_{02}). In terms of the parameters in (6.1), these hypotheses correspond to

H_{01}: $\mu_1 = \ldots = \mu_5$, $\phi_1 = \ldots = \phi_5 = 0$, and H_{02}: $\mu_1 = \ldots = \mu_5$, $\phi_1 = \ldots = \phi_5$.

In Table 4, we present the p-values of the Wald test for H_{01} and H_{02} under both the assumptions of a strong noise (WS) and a weak noise (WW). At the 5% significance level, the hypothesis of strong noise is rejected for all indices except FTSE 120. On the other hand, the hypothesis of a weak noise is rejected for the four indices. Since the class of strong noises is a subset of the class of weak noises, these results show that the standard inference based on the assumption of a strong noise can be misleading. For the hypothesis H_{02} of non seasonality, we get similar results. At the 5% level, WS rejects except with FTSE 120 whilst WW rejects with the four indices.

The hypothesis of weak white noise being rejected, we present in Table 5 the estimated parameters and their estimated standard error under the assumption of a weak noise. Unsurprisingly, all the estimates are rather small and very few are significant, even at the 10% level. With CAC 40, $\hat{\phi}_5$ is significant at 5%, with DAX, $\hat{\mu}_1$, $\hat{\phi}_3$ and $\hat{\phi}_5$ are significant at 10%, with SMI, $\hat{\phi}_5$ is significant at 1%. With FTSE 120, the smallest p-value 0.129 corresponds to $\hat{\phi}_3$.

With respect to the means $\hat{\mu}_\nu$, even if they are all positive on Monday, we cannot really talk of a global Monday effect since only one value (DAX) is significant. Wednesday seems a particularly bad day with negative returns for the four indices.

Index	H_{01}		H_{02}	
	WS	WW	WS	WW
CAC 40	.026	.005	.022	.013
DAX	.013	.038	.009	.035
FTSE 120	.165	.001	.180	.014
SMI	.001	.005	.002	.008

Table 4. p-values of the Wald test for the hypotheses H_{01} of white noise and H_{02} of non seasonality on the daily returns of the four indices, under the assumption of strong noise (WS) or a weak noise (WW), based on Model (6.1).

Index	CAC 40			DAX			FTSE 120			SMI		
NT	4165			4200			4165			4185		
Day	$\hat{\mu}_\nu$	$\hat{\phi}_\nu$	$\hat{\sigma}_\nu$	$\hat{\mu}_\nu$	$\hat{\phi}_\nu$	$\hat{\sigma}_\nu$	$\hat{\mu}_\nu$	$\hat{\phi}_\nu$	$\hat{\sigma}_\nu$	$\hat{\mu}_\nu$	$\hat{\phi}_\nu$	$\hat{\sigma}_\nu$
Monday	.045 (.056)	−.081 (.124)	1.91	$.132^c$ (.072)	−.060 (.104)	2.06	.061 (.046)	−.042 (.072)	1.51	.038 (.050)	−.008 (.120)	1.66
Tuesday	−.012 (.040)	−.045 (.029)	1.35	.013 (.039)	−.031 (.042)	1.41	.003 (.041)	−.056 (.037)	1.09	.000 (.037)	−.016 (.042)	1.13
Wednesday	−.055 (.058)	−.055 (.036)	1.41	−.090 (.058)	$−.066^c$ (.038)	1.43	−.071 (.050)	−.015 (.031)	1.12	−.038 (.044)	−.008 (.052)	1.14
Thursday	.001 (.067)	−.008 (.067)	1.46	−.044 (.062)	−.008 (.072)	1.47	.004 (.043)	.002 (.088)	1.15	−.011 (.055)	.061 (.071)	1.24
Friday	.021 (.042)	$.092^b$ (.042)	1.35	−.010 (.043)	$.086^c$ (.048)	1.38	.004 (.028)	.052 (.047)	1.18	.010 (.036)	$.159^a$ (.040)	1.15

Table 5. QLS estimates and their estimated standard errors under the assumption of a weak white noise (in parentheses) of Model (6.1) fitted to the daily returns of the four European stock market indices. a, b and c respectively mean significant at the 1%, 5% and 10% levels.

The autoregressive coefficients $\hat{\phi}_\nu$ which also represent the correlation between today returns and those of yesterday, are all negative on Monday, Tuesday and Wednesday but they are all positive on Friday. Furthermore, three of them are significant (CAC, DAX and SMI) on Friday. With these three indices and the period considered, it is probably more appropriate to talk of a Friday effect rather than a Monday effect.

Finally, perusal of the estimated noise standard deviations shows that for the four indices, the volatility is considerably greater on Monday and for the other days, it is smaller and almost constant. A comparison of the four indices indicates that CAC and DAX are systematically more volatile over the five days of the week than the other two.

7 Conclusion

In this work, we have established under mild assumptions, the almost sure consistency and the asymptotic normality of the weighted least squares estimators for invertible and causal PARMA models with dependent but uncorrelated errors. Our results extend Theorem 3.1 of Basawa and Lund (2001) for PARMA models with independent errors. The asymptotic covariance matrix of WLS estimators obtained under independent errors is generally different from the one under dependent errors and the difference may be huge with some types of dependence. The standard procedures of estimation and inference in PARMA models under the assumption of independent errors can be quite misleading when analysing data from PARMA models with dependent errors.

The empirical results of Sections 5.2 and 6 illustrate the applicability of our theoretical results using a consistent estimator of the covariance matrix of the QLS estimators of weak PARMA parameters. In the model building process for weak PAR models, a global diagnostic checking procedure along the lines of Francq *et al.* (2005) would be useful but it is beyond the scope of this paper. In contrast with the strong PAR case, as described in McLeod (1994), the asymptotic covariance matrix of the QLS estimators of a weak PAR model is no longer block diagonal with respect to seasons and depends on the fourth-order moments of the noise process. The usual model selection criteria (AIC, BIC, ...) also need to be studied thoroughly in the context of periodic models.

APPENDIX

The proofs of Theorems 3.1 and 3.2 are splitted in a series of lemmas. The strong consistency of the LS estimators follows from Lemmas 1 to 10. The asymptotic normality is deduced from Lemma 11 to 14. In this appendix, the letters K, Δ and M stand for generic positive constants that may change from one place to another.

It will be shown in the sequel that working with the true errors $\epsilon_t(\vec{\alpha})$ rather than the truncated ones $(e_t(\vec{\alpha}))$ does not alter the asymptotic results and we will use the criterion

$$O_N^{\vec{\omega}^2}(\vec{\alpha}) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} \omega_\nu^{-2} \epsilon_{nT+\nu}^2(\vec{\alpha}). \tag{7.1}$$

instead of $Q_N^{\vec{\omega}^2}(\vec{\alpha})$ defined by (3.12).

Lemma 7.1 *For any $\vec{\alpha} \in \Omega$, let $(C_i(\vec{\alpha}))_{i\in\mathbb{N}}$ be the sequence satisfying*

$$\boldsymbol{\epsilon}_n(\vec{\alpha}) = \sum_{i=0}^{\infty} C_i(\vec{\alpha})\mathbf{X}_{n-i}.$$

Then, there exists a constant K such that for all $i \in \mathbb{N}$,

$$\sup_{\vec{\alpha} \in \Omega_\delta} \|C_i(\vec{\alpha})\| \leq K i^{Tq^*} \left(\frac{1}{1+\delta} \right)^i.$$

Proof. Consider the case where $q^* = 0$ and $p^* > 0$. Let $L = \sup_{i=0,\dots,p^*} \sup_{\vec{\alpha} \in \Omega_\delta} \|C_i(\vec{\alpha})\|$ and put $K = L(1+\delta)^{p^*}$. It is not difficult to show that

$$\sup_{\vec{\alpha} \in \Omega_\delta} \|C_i(\vec{\alpha})\| \leq K \left(\frac{1}{1+\delta} \right)^i.$$

Now, consider the case $q^* > 0$. Define the $q^*T \times 1$ vectors

$$\tilde{\mathbf{X}}_n = \begin{pmatrix} \mathbf{X}_n \\ 0_{T \times 1} \\ \vdots \\ 0_{T \times 1} \end{pmatrix}, \qquad \underline{\boldsymbol{\epsilon}}_n(\vec{\alpha}) = \begin{pmatrix} \boldsymbol{\epsilon}_n(\vec{\alpha}) \\ \boldsymbol{\epsilon}_{n-1}(\vec{\alpha}) \\ \vdots \\ \boldsymbol{\epsilon}_{n-q^*+1}(\vec{\alpha}) \end{pmatrix}, \qquad \underline{\mathbf{e}}_n(\vec{\alpha}) = \begin{pmatrix} \mathbf{e}_n(\vec{\alpha}) \\ \mathbf{e}_{n-1}(\vec{\alpha}) \\ \vdots \\ \boldsymbol{\epsilon}_{n-q^*+1}(\vec{\alpha}) \end{pmatrix}, \qquad (7.2)$$

and the $q^*T \times q^*T$ companion matrices

$$A_i = \begin{pmatrix} \Theta_0^{-1}\Phi_i & 0_{T \times T} & \dots & 0_{T \times T} \\ 0_{T \times T} & 0_{T \times T} & \dots & 0_{T \times T} \\ \vdots & & \ddots & 0_{T \times T} \\ 0_{T \times T} & 0_{T \times T} & \dots & 0_{T \times T} \end{pmatrix}, \qquad i = 1, \dots, p^*,$$

$$D = \begin{pmatrix} -\Theta_0^{-1}\Theta_1 & -\Theta_0^{-1}\Theta_2 & \cdots & -\Theta_0^{-1}\Theta_{q^*-1} & -\Theta_0^{-1}\Theta_{q^*} \\ \mathbf{I}_{T \times T} & 0_{T \times T} & \cdots & 0_{T \times T} & 0_{T \times T} \\ 0_{T \times T} & \mathbf{I}_{T \times T} & \ddots & \vdots & 0_{T \times T} \\ \vdots & \ddots & \ddots & 0_{T \times T} & 0_{T \times T} \\ 0_{T \times T} & \cdots & 0_{T \times T} & \mathbf{I}_{T \times T} & 0_{T \times T} \end{pmatrix}. \qquad (7.3)$$

It is easy to check that

$$A_0 \tilde{\mathbf{X}}_n + \sum_{i=1}^{p^*} A_i \tilde{\mathbf{X}}_{n-i} = \underline{\boldsymbol{\epsilon}}_n(\vec{\alpha}) - D\underline{\boldsymbol{\epsilon}}_{n-1}(\vec{\alpha}).$$

This implies that

$$\underline{\boldsymbol{\epsilon}}_n(\vec{\alpha}) = \sum_{j=0}^{\infty} D^j \left(A_0 \tilde{\mathbf{X}}_{n-j} + A_1 \tilde{\mathbf{X}}_{n-j-1} + \dots + A_{p^*} \tilde{\mathbf{X}}_{n-j-p^*} \right) = \sum_{i=0}^{\infty} \tilde{C}_i(\vec{\alpha}) \tilde{\mathbf{X}}_{n-i},$$

where

$$\tilde{C}_i(\vec{\alpha}) = \sum_{j=0}^{min(i,p^*)} D^{i-j} A_j. \qquad (7.4)$$

Again, using a multiplicative norm, it can be shown that there exists a constant K_1 independent of $\vec{\alpha}$ such that

$$\|\tilde{C}_i(\vec{\alpha})\| \leq \sum_{j=0}^{p^*} \|D^{i-j}\| \|A_j\| \leq K_1 i^{Tq^*} \left(\frac{1}{1+\delta} \right)^i.$$

The conclusion follow from the fact that $C_i(\vec{\alpha})$ is the matrix in the first row and the first column of the block matrix $\tilde{C}_i(\vec{\alpha})$.

26

Lemma 7.2 *We have*

$$\mathbb{E}\left[\sup_{\vec{\alpha}\in\Omega_\delta}\epsilon^2_{nT+\nu}(\vec{\alpha})\right]<\infty.$$

Proof. From (3.3) and Lemma 7.1, there exists a constant K such that

$$\sup_{\vec{\alpha}\in\Omega_\delta}|\epsilon_{nT+\nu}(\vec{\alpha})|\le K\sum_{i=0}^{\infty}i^{Tq^*}\left(\frac{1}{1+\delta}\right)^i|X_{nT+\nu-i}|.$$

Using the Cauchy criterion, it can be shown that the series $\sum_{i=0}^{\infty}i^{Tq^*}\left(\frac{1}{1+\delta}\right)^i|X_{nT+\nu-i}|$ converges in mean squares. The result follows.

Lemma 7.3 *Let $\boldsymbol{\epsilon}_n(\vec{\alpha})$ as defined by (3.3). For any $\vec{\alpha}\in\Omega$,*

$$\boldsymbol{\epsilon}_n(\vec{\alpha})=\boldsymbol{\epsilon}_n(\vec{\alpha}_0)\quad a.s.\quad\Rightarrow\vec{\alpha}=\vec{\alpha}_0.$$

Proof. Since the covariance matrix $\Sigma_{0\epsilon}=\mathrm{Diag}(\sigma^2_{01},...,\sigma^2_{0T})$ is supposed to be strictly positive definite, for any sequence of $T\times T$ real matrices $(\Psi_i)_{i\in\mathbb{N}}$, we have that

$$\sum_{i=0}^{\infty}\Psi_i\mathbf{X}_{n-i}=0\quad a.s.\quad\Rightarrow\Psi_i=0,i\ge 0.$$

Let $\vec{\alpha}\in\Omega$, we have (see Lemma 7.1)

$$\boldsymbol{\epsilon}_n(\vec{\alpha})=\sum_{i=0}^{\infty}C_i(\vec{\alpha})\mathbf{X}_{n-i}\qquad\text{and}\qquad\boldsymbol{\epsilon}_n(\vec{\alpha}_0)=\sum_{i=0}^{\infty}C_i(\vec{\alpha}_0)\mathbf{X}_{n-i}.$$

Then,

$$\boldsymbol{\epsilon}_n(\vec{\alpha})-\boldsymbol{\epsilon}_n(\vec{\alpha}_0)=0\quad a.s.\quad\Rightarrow\quad\sum_{i=0}^{\infty}\left(C_i(\vec{\alpha})-C_i(\vec{\alpha}_0)\right)\mathbf{X}_{n-i}=0\quad a.s.$$
$$\Rightarrow\quad C_i(\vec{\alpha})=C_i(\vec{\alpha}_0),i\ge 0.$$

The conclusion hold by invoking the identifiability assumption.

Lemma 7.4 *For any $\vec{\alpha}\in\Omega$ and any $\vec{\omega}^2=(\omega^2_1,\ldots,\omega^2_T)>0$, let*

$$O^{\vec{\omega}^2}_{\infty}(\vec{\alpha})=\sum_{\nu=1}^{T}\omega^{-2}_\nu\mathbb{E}\left[\epsilon^2_\nu(\vec{\alpha})\right].$$

Then, for any $\vec{\alpha}\neq\vec{\alpha}_0$, we have

$$O^{\vec{\omega}^2}_{\infty}(\vec{\alpha}_0)=\sum_{\nu=1}^{T}\frac{\sigma^2_{0\nu}}{\omega^2_\nu}<O^{\vec{\omega}^2}_{\infty}(\vec{\alpha}).$$

Proof. To ease the writing of the proof, we suppose without loss of generality that $n = 0$ in $t = nT + \nu$. It is clear that $\epsilon_\nu(\vec{\alpha}) - \epsilon_\nu(\vec{\alpha}_0)$ belongs to the Hilbert space $\mathcal{H}_X(\nu - 1)$. Therefore the linear innovation $\epsilon_\nu(\vec{\alpha}_0)$ is not correlated with $\epsilon_\nu(\vec{\alpha}) - \epsilon_\nu(\vec{\alpha}_0)$. Thus

$$
\begin{aligned}
\mathbb{E}\left[\epsilon_\nu^2(\vec{\alpha})\right] &= \mathbb{E}\left[(\epsilon_\nu(\vec{\alpha}) - \epsilon_\nu(\vec{\alpha}_0) + \epsilon_\nu(\vec{\alpha}_0))^2\right] \\
&= \mathbb{E}\left[\epsilon_\nu^2(\vec{\alpha}_0)\right] + \mathbb{E}\left[(\epsilon_\nu(\vec{\alpha}) - \epsilon_\nu(\vec{\alpha}_0))^2\right] + 2\mathrm{Cov}\left(\epsilon_\nu(\vec{\alpha}_0), \epsilon_\nu(\vec{\alpha}) - \epsilon_\nu(\vec{\alpha}_0)\right) \\
&= \sigma_{0\nu}^2 + \mathbb{E}\left[(\epsilon_\nu(\vec{\alpha}) - \epsilon_\nu(\vec{\alpha}_0))^2\right].
\end{aligned}
$$

If $\vec{\alpha} \neq \vec{\alpha}_0$, Lemma 7.3 implies that the second term in the right hand side of the last equality is strictly positive for at least one $\nu \in \{1, ..., T\}$. Therefore,

$$
O_\infty^{\vec{\omega}^2}(\vec{\alpha}) = \sum_{\nu=1}^T \frac{1}{\omega_\nu^2} \mathbb{E}\left[\epsilon_\nu^2(\vec{\alpha})\right] > \sum_{\nu=1}^T \frac{\sigma_{0\nu}^2}{\omega_\nu^2}
$$

if $\vec{\alpha} \neq \vec{\alpha}_0$.

Lemma 7.5 *For any $\vec{\alpha}^\star \in \Omega$, $\vec{\alpha}^\star \neq \vec{\alpha}_0$, and any $\vec{\omega}^2 > 0$, there exists a neighbourhood $V(\vec{\alpha}^\star)$ of $\vec{\alpha}^\star$ such that $V(\vec{\alpha}^\star) \subset \Omega$ and*

$$
\liminf_{N \to \infty} \inf_{\vec{\alpha} \in V(\vec{\alpha}^\star)} O_N^{\vec{\omega}^2}(\vec{\alpha}) > \lim_{N \to \infty} O_N^{\vec{\omega}^2}(\vec{\alpha}_0) = \sum_{\nu=1}^T \frac{\sigma_{0\nu}^2}{\omega_\nu^2} \quad \text{a.s.}
$$

where $O_N^{\vec{\omega}^2}(\vec{\alpha})$ is given by (7.1).

Proof. Let $V_m(\vec{\alpha}^\star)$ be the open sphere with centre $\vec{\alpha}^\star$ and radius $\frac{1}{m}$. Let

$$
S_m(n) = \inf_{\vec{\alpha} \in V_m(\vec{\alpha}^\star) \cap \Omega} \sum_{\nu=1}^T \omega_\nu^{-2} \epsilon_{nT+\nu}^2(\vec{\alpha}).
$$

The variable $S_m(n)$ is measurable because it can be written as the inf over a dense countable subset. Moreover, $S_m(n)$ belongs to L^1. The ergodic theorem applied to the stationary process $\{S_m(n) : n \in \mathbb{Z}\}$ shows that almost surely

$$
\inf_{\vec{\alpha} \in V_m(\vec{\alpha}^\star) \cap \Omega} O_N^{\vec{\omega}^2}(\vec{\alpha}) = \inf_{\vec{\alpha} \in V_m(\vec{\alpha}^\star) \cap \Omega} \left\{ \frac{1}{N} \sum_{\nu=1}^T \omega_\nu^{-2} \sum_{n=0}^{N-1} \epsilon_{nT+\nu}^2(\vec{\alpha}) \right\} \geq \frac{1}{N} \sum_{n=0}^{N-1} S_m(n) \xrightarrow[N \to \infty]{} \mathbb{E}\left[S_m(0)\right].
$$

Since $S_m(0)$ increases to $\sum_{\nu=1}^T \omega_\nu^{-2} \epsilon_\nu^2(\vec{\alpha}^\star)$ as m tends to infinity, by Lemma 7.4 and the monotone convergence theorem, we obtain that

$$
\lim_{m \to \infty} \mathbb{E}\left[S_m(0)\right] = \sum_{\nu=1}^T \omega_\nu^{-2} \mathbb{E}\left[\epsilon_\nu^2(\vec{\alpha}^\star)\right] = O_\infty^{\vec{\omega}^2}(\vec{\alpha}^\star) > \sum_{\nu=1}^T \frac{\sigma_{0\nu}^2}{\omega_\nu^2}.
$$

This implies that

$$\liminf_{m\to\infty} \liminf_{N\to\infty} \inf_{\vec{\alpha}\in V_m(\vec{\alpha}^\star)} O_N^{\vec{\omega}^2}(\vec{\alpha}) \ge O_\infty^{\vec{\omega}^2}(\vec{\alpha}^\star) > \sum_{\nu=1}^{T} \frac{\sigma_{0\nu}^2}{\omega_\nu^2},$$

and the result follows.

Lemma 7.6 *Let $\vec{\omega}^2 = (\omega_1^2, \ldots, \omega_T^2)'$ be a vector of strictly positive constants, and let $\hat{\vec{\omega}}^2$ be a sequence of random vectors such that $\hat{\vec{\omega}}^2 \to \vec{\omega}^2$ almost surely as $N \to \infty$. For any $\vec{\alpha}^\star \in \Omega$, $\vec{\alpha}^\star \ne \vec{\alpha}_0$, there exists a neighbourhood $V(\vec{\alpha}^\star)$ of $\vec{\alpha}^\star$ such that $V(\vec{\alpha}^\star) \subset \Omega$ and*

$$\liminf_{N\to\infty} \inf_{\vec{\alpha}\in V(\vec{\alpha}^\star)} O_N^{\hat{\vec{\omega}}^2}(\vec{\alpha}) > \lim_{N\to\infty} O_N^{\vec{\omega}^2}(\vec{\alpha}_0) = \sum_{\nu=1}^{T} \frac{\sigma_{0\nu}^2}{\omega_\nu^2} \quad \text{a.s.}$$

Proof. For any $\varepsilon > 0$, we have almost surely

$$\max_{\nu=1,\ldots,T} \left| \frac{1}{\hat{\omega}_\nu^2} - \frac{1}{\omega_\nu^2} \right| < \varepsilon$$

for N large enough. By the ergodic theorem and Lemma 7.2, we thus have

$$
\begin{aligned}
\limsup_{N\to\infty} \sup_{\vec{\alpha}\in\Omega} \left| O_N^{\hat{\vec{\omega}}^2}(\vec{\alpha}) - O_N^{\vec{\omega}^2}(\vec{\alpha}) \right| &\le \limsup_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} \left| \frac{1}{\hat{\omega}_\nu^2} - \frac{1}{\omega_\nu^2} \right| \sup_{\vec{\alpha}\in\Omega} \left| \epsilon_{nT+\nu}^2(\vec{\alpha}) \right| \\
&\le \varepsilon \sum_{\nu=1}^{T} E \sup_{\vec{\alpha}\in\Omega} \left| \epsilon_\nu^2(\vec{\alpha}) \right|.
\end{aligned}
$$

Since the inequality holds for any $\varepsilon > 0$, we have

$$\lim_{N\to\infty} \sup_{\vec{\alpha}\in\Omega} \left| O_N^{\hat{\vec{\omega}}^2}(\vec{\alpha}) - O_N^{\vec{\omega}^2}(\vec{\alpha}) \right| = 0 \quad \text{a.s..}$$

In view of Lemma 7.5, the conclusion follows.

Lemma 7.7 *Let $\epsilon_{nT+\nu}(\vec{\alpha})$ given by (3.7) and $e_{nT+\nu}(\vec{\alpha})$ given by (3.8). Almost surely, there exist $K > 0$ and $\rho \in (0,1)$ such that*

$$\sup_{\vec{\alpha}\in\Omega_\delta} |\epsilon_{nT+\nu}(\vec{\alpha}) - e_{nT+\nu}(\vec{\alpha})| \le K\rho^n$$

for all $\nu \in \{1, 2, \ldots, T\}$ and all $n \ge 1$.

Proof. For $q^* = 0$, the result is obvious. Otherwise, consider and the $q^*T \times 1$ vectors $\underline{\boldsymbol{\epsilon}}_n(\vec{\alpha})$, $\underline{\mathbf{e}}_n(\vec{\alpha})$ defined by (7.2) and the $q^*T \times q^*T$ companion matrix D given by (7.3). We have

$$\underline{\boldsymbol{\epsilon}}_n(\vec{\alpha}) - \underline{\mathbf{e}}_n(\vec{\alpha}) = D^{n-p^*} \left(\underline{\boldsymbol{\epsilon}}_p(\vec{\alpha}) - \underline{\mathbf{e}}_p(\vec{\alpha}) \right) \qquad \forall n > p^*. \tag{7.5}$$

Consider the Jordan decomposition $D = P\Lambda P^{-1}$, where the matrix Λ takes the form

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0 & ... & 0 \\ 0 & \Lambda_2 & ... & 0 \\ \vdots & & \ddots & \\ 0 & 0 & ... & \Lambda_s \end{pmatrix} \quad \text{with} \quad \Lambda_h = \begin{pmatrix} \lambda_h & 1 & 0 & ... & 0 \\ 0 & \lambda_h & 1 & & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & & & \lambda_h & 1 \\ 0 & 0 & ... & 0 & \lambda_h \end{pmatrix}.$$

Using this decomposition, it can be shown that $D^t = P\Lambda^t P^{-1}$ where $\Lambda^t = \text{Diag}\left(\Lambda_1^t, ..., \Lambda_s^t\right)$ with

$$\Lambda_h^t = \begin{pmatrix} \lambda_h^t & \begin{pmatrix} t \\ 1 \end{pmatrix}\lambda_h^{t-1} & ... & \begin{pmatrix} t \\ r_h - 1 \end{pmatrix}\lambda_h^{t-r_h+1} \\ 0 & \lambda_h^t & ... & \begin{pmatrix} t \\ r_h - 2 \end{pmatrix}\lambda_i^{t-r_h+2} \\ \vdots & & \ddots & \\ 0 & 0 & ... & \lambda_h^t \end{pmatrix}.$$

Since the nonzero eigenvalues λ_i of D are equal to the inverse of the zeros of $\det\{\Theta(z)\}$, we have $\max_h |\lambda_h| \leq \frac{1}{1+\delta}$. Therefore, there exists a positive constant K independent of $\vec{\alpha}$ such that

$$\|D^t\| = \|P\Lambda^t P^{-1}\| \leq \|P\|\|P^{-1}\|\|\Lambda^t\| \leq Kt^{Tq^*}\left(\frac{1}{\delta+1}\right)^t \tag{7.6}$$

where $\|.\|$ stands for any multiplicative matrix norm. The conclusion follows from (7.5) and (7.6).

Lemma 7.8 Let $Q_N^{\vec{\omega}^2}(\vec{\alpha})$ given by (3.12) and $O_N^{\vec{\omega}^2}(\vec{\alpha})$ given by (7.1). For any constant $\underline{\omega} > 0$, we have almost surely

$$\sup_{\vec{\omega}^2 > \underline{\omega}} \sup_{\vec{\alpha} \in \Omega_\delta} |O_N^{\vec{\omega}^2}(\vec{\alpha}) - Q_N^{\vec{\omega}^2}(\vec{\alpha})| = O(N^{-1}) \qquad as \qquad N \to \infty.$$

Proof. Using Lemma 7.7, we obtain that

$$N \sup_{\vec{\alpha} \in \Omega_\delta,\, \vec{\omega}^2 > \underline{\omega}} |O_N^{\vec{\omega}^2}(\vec{\alpha}) - Q_N^{\vec{\omega}^2}(\vec{\alpha})|$$

$$= \sup_{\vec{\alpha} \in \Omega_\delta,\, \vec{\omega}^2 > \underline{\omega}} \left|\sum_{\nu=1}^{T} \omega_\nu^{-2} \sum_{n=0}^{N-1} \left(e_{nT+\nu}^2(\vec{\alpha}) - \epsilon_{nT+\nu}^2(\vec{\alpha})\right)\right|$$

$$\leq \sum_{\nu=1}^{T} \underline{\omega}^{-2} \sup_{\vec{\alpha} \in \Omega_\delta} \left|\sum_{n=0}^{N-1} \left(e_{nT+\nu}^2(\vec{\alpha}) - \epsilon_{nT+\nu}^2(\vec{\alpha})\right)\right|$$

$$= \sum_{\nu=1}^{T} \underline{\omega}^{-2} \sup_{\vec{\alpha} \in \Omega_\delta} \left|\sum_{n=0}^{N-1} (e_{nT+\nu}(\vec{\alpha}) - \epsilon_{nT+\nu}(\vec{\alpha}))^2 + 2\epsilon_{nT+\nu}(\vec{\alpha})\left(e_{nT+\nu}(\vec{\alpha}) - \epsilon_{nT+\nu}(\vec{\alpha})\right)\right|$$

$$\leq K\sum_{\nu=1}^{T}\sum_{n=0}^{N-1} \rho^n \left(1 + \sup_{\vec{\alpha} \in \Omega_\delta} |\epsilon_{nT+\nu}(\vec{\alpha})|\right)$$

for $K > 0$ and $\rho \in (0, 1)$. By Lemma 7.2 and the Beppo-Levy theorem, we have

$$E \sum_{n=0}^{\infty} \rho^n \left(1 + \sup_{\vec{\alpha} \in \Omega_\delta} |\epsilon_{nT+\nu}(\vec{\alpha})| \right) < \infty.$$

Therefore, the series $\sum_{n=0}^{\infty} \rho^n \left(1 + \sup_{\vec{\alpha} \in \Omega_\delta} |\epsilon_{nT+\nu}(\vec{\alpha})| \right)$ converges almost surely, which completes the proof.

Lemma 7.9 *For any $\vec{\alpha}^\star \in \Omega$, $\vec{\alpha}^\star \neq \vec{\alpha}_0$, and any $\vec{\omega}^2 > 0$, there exists a neighbourhood $V(\vec{\alpha}^\star)$ of $\vec{\alpha}^\star$ such that $V(\vec{\alpha}^\star) \subset \Omega$ and*

$$\liminf_{N \to \infty} \inf_{\vec{\alpha} \in V(\vec{\alpha}^\star)} Q_N^{\vec{\omega}^2}(\vec{\alpha}) > \sum_{\nu=1}^{T} \frac{\sigma_{0\nu}^2}{\omega_\nu^2} \quad \text{a.s.}$$

For any neighbourhood $V(\vec{\alpha}_0)$ of $\vec{\alpha}_0$,

$$\limsup_{N \to \infty} \inf_{\vec{\alpha} \in V(\vec{\alpha}_0)} Q_N^{\vec{\omega}^2}(\vec{\alpha}) \leq \sum_{\nu=1}^{T} \frac{\sigma_{0\nu}^2}{\omega_\nu^2} \quad \text{a.s.}$$

Proof. Since $V(\vec{\alpha}^\star)$ can be chosen as being included in Ω_δ for some $\delta > 0$. We have

$$\inf_{\vec{\alpha} \in V(\vec{\alpha}^\star)} Q_N^{\vec{\omega}^2}(\vec{\alpha}) \geq \inf_{\vec{\alpha} \in V(\vec{\alpha}^\star)} O_N^{\vec{\omega}^2}(\vec{\alpha}) - \sup_{\vec{\alpha} \in \Omega_\delta} |O_N^{\vec{\omega}^2}(\vec{\alpha}) - Q_N^{\vec{\omega}^2}(\vec{\alpha})|$$

and

$$\inf_{\vec{\alpha} \in V(\vec{\alpha}_0)} Q_N^{\vec{\omega}^2}(\vec{\alpha}) \leq O_N^{\vec{\omega}^2}(\vec{\alpha}_0) + \sup_{\vec{\alpha} \in \Omega_\delta} |O_N^{\vec{\omega}^2}(\vec{\alpha}) - Q_N^{\vec{\omega}^2}(\vec{\alpha})|$$

The conclusion follows from Lemmas 7.5 and 7.8.

Lemma 7.10 *Let $\vec{\omega}^2 = (\omega_1^2, \ldots, \omega_T^2)'$ be a vector of strictly positive constants, and let $\hat{\vec{\omega}}^2$ be a sequence of random vectors such that $\hat{\vec{\omega}}^2 \to \vec{\omega}^2$ almost surely as $N \to \infty$. For any $\vec{\alpha}^\star \in \Omega$, $\vec{\alpha}^\star \neq \vec{\alpha}_0$, there exists a neighbourhood $V(\vec{\alpha}^\star)$ of $\vec{\alpha}^\star$ such that $V(\vec{\alpha}^\star) \subset \Omega$ and*

$$\liminf_{N \to \infty} \inf_{\vec{\alpha} \in V(\vec{\alpha}^\star)} Q_N^{\hat{\vec{\omega}}^2}(\vec{\alpha}) > \sum_{\nu=1}^{T} \frac{\sigma_{0\nu}^2}{\omega_\nu^2} \quad \text{a.s.}$$

For any neighbourhood $V(\vec{\alpha}_0)$ of $\vec{\alpha}_0$

$$\limsup_{N \to \infty} \inf_{\vec{\alpha} \in V(\vec{\alpha}_0)} Q_N^{\hat{\vec{\omega}}^2}(\vec{\alpha}) \leq \sum_{\nu=1}^{T} \frac{\sigma_{0\nu}^2}{\omega_\nu^2} \quad \text{a.s.}$$

Proof. The proof is the same as that of Lemma 7.9, referring to Lemma 7.6 instead of Lemma 7.5.

Proof of Theorem 3.1. Let $V(\vec{\alpha}_0)$ be a neighbourhood of $\vec{\alpha}_0$. Clearly Ω_δ is covered by $V(\vec{\alpha}_0)$ and the union of all the $V(\vec{\alpha}^\star)$, $\vec{\alpha}^\star \in \Omega_\delta - V(\vec{\alpha}_0)$ where $V(\vec{\alpha}^\star)$ is defined in Lemma 7.9. By the compactness

31

of Ω_δ, there exist $\vec{\alpha}_1, ..., \vec{\alpha}_k$ such that Ω_δ is covered by $V(\vec{\alpha}_0), V(\vec{\alpha}_1), ..., V(\vec{\alpha}_k)$. Lemma 7.9 shows that, almost surely,

$$\inf_{\vec{\alpha} \in \Omega_\delta} Q_N^{\vec{\omega}^2}(\vec{\alpha}) = \min_{i=0,1,...,k} \inf_{\vec{\alpha} \in V(\vec{\alpha}_i) \cap \Omega_\delta} Q_N^{\vec{\omega}^2}(\vec{\alpha}) = \inf_{\vec{\alpha} \in V(\vec{\alpha}_0) \cap \Omega_\delta} Q_N^{\vec{\omega}^2}(\vec{\alpha})$$

for N large enough. Therefore, $\hat{\vec{\alpha}}_{\text{WLS}}^{\vec{\omega}^2}$ almost surely belongs to $V(\vec{\alpha}_0)$ for N large enough. Since $V(\vec{\alpha}_0)$ can be arbitrarily small, $\hat{\vec{\alpha}}_{\text{WLS}}^{\vec{\omega}^2} \to \vec{\alpha}_0$ almost surely as $N \to \infty$. The OLS and GLS estimators correspond to WLS estimators, for particular choices of the weights vector $\vec{\omega}^2$. Thus the first three consistencies of Theorem 3.1 are shown. Using Lemma 7.10, the previous arguments show that $\hat{\vec{\alpha}}_{\text{WLS}}^{\hat{\vec{\sigma}}^2} \to \vec{\alpha}_0$ almost surely whenever $\hat{\vec{\sigma}}^2 \to \vec{\sigma}_0^2$ as $N \to \infty$. It follows that the QLS estimator is also consistent. $\square$

Theorem 3.2 can be established using the following lemmas.

Lemma 7.11 *For any $\vec{\alpha}^\star \in \Omega$, and any $m \in \{1, 2, ..., (p+q)T\}$, there exist absolutely summable sequences $(C_i(\vec{\alpha}))_{i \in \mathbb{N}}$ and $(C_{m,i}(\vec{\alpha}))_{i \in \mathbb{N}}$ such that*

$$\boldsymbol{\epsilon}_n(\vec{\alpha}) = \sum_{i=0}^{\infty} C_i(\vec{\alpha}) \mathbf{X}_{n-i} \qquad \text{and} \qquad \frac{\partial \boldsymbol{\epsilon}_n(\vec{\alpha})}{\partial \alpha_m} = \sum_{i=0}^{\infty} C_{m,i}(\vec{\alpha}) \mathbf{X}_{n-i}. \tag{7.7}$$

Moreover, there exist $\rho \in (0, 1)$ and $K \in [0, \infty)$ such that, for all $i \geq 0$,

$$\sup_{\vec{\alpha} \in \Omega_\delta} \|C_i(\vec{\alpha})\| \leq K\rho^i \qquad \text{and} \qquad \sup_{\vec{\alpha} \in \Omega_\delta} \|C_{m,i}(\vec{\alpha})\| \leq K\rho^i. \tag{7.8}$$

Proof. For $q^* = 0$, the result is obvious. Otherwise, we use arguments similar to those used to prove Lemma 7.1. Indeed, as shown in Lemma 7.1, $\underline{\boldsymbol{\epsilon}}_n(\vec{\alpha}) = \sum_{i=0}^{\infty} \tilde{C}_i(\vec{\alpha}) \tilde{\mathbf{X}}_{n-i}$, where $\tilde{\mathbf{X}}_n$ and $\underline{\boldsymbol{\epsilon}}_n(\vec{\alpha})$ are given by (7.2) and $\tilde{C}_i(\vec{\alpha}) = \sum_{j=0}^{\min(i,p^*)} B^{i-j} A_j$ as in (7.4). Then,

$$\frac{\partial \underline{\boldsymbol{\epsilon}}_n(\vec{\alpha})}{\partial \alpha_m} = \sum_{i=0}^{\infty} \frac{\partial \tilde{C}_i(\vec{\alpha})}{\partial \alpha_m} \tilde{\mathbf{X}}_{n-i},$$

and since,

$$\boldsymbol{\epsilon}_n(\vec{\alpha}) = \sum_{i=0}^{\infty} L_1' \tilde{C}_i(\vec{\alpha}) L_1 \mathbf{X}_{n-i} = \sum_{i=0}^{\infty} C_i(\vec{\alpha}) \mathbf{X}_{n-i},$$

we have

$$\frac{\partial \boldsymbol{\epsilon}_n(\vec{\alpha})}{\partial \alpha_m} = \sum_{i=0}^{\infty} L_1' \frac{\partial \tilde{C}_i(\vec{\alpha})}{\partial \alpha_m} L_1 \mathbf{X}_{n-i} = \sum_{i=0}^{\infty} C_{m,i}(\vec{\alpha}) \mathbf{X}_{n-i},$$

where $C_i(\vec{\alpha}) = L_1' \tilde{C}_i(\vec{\alpha}) L_1$ and $C_{i,m}(\vec{\alpha}) = L_1' \frac{\partial \tilde{C}_i(\vec{\alpha})}{\partial \alpha_m} L_1$ with $L_1 = (I_{T \times T}, 0_{T \times T}, ..., 0_{T \times T})'$.

Also, from Lemma 7.1 there exists a positive constant K_1 independent of $\vec{\alpha}$ such that

$$
\begin{aligned}
\sup_{\vec{\alpha} \in \Omega_\delta} \|C_i(\vec{\alpha})\| &\leq K_1 i^{Tq^*} \left(\frac{1}{1+\delta}\right)^i \\
&\leq K_1 i^{Tq^*} \left(\frac{1}{1+(\delta/2)}\right)^i \left(\frac{1+(\delta/2)}{1+\delta}\right)^i \\
&\leq K \rho^i.
\end{aligned}
$$

where $\rho = \dfrac{1+(\delta/2)}{1+\delta} \in (0,1)$ and $K = K_1 \max_i \left\{ i^{Tq^*} \left(\dfrac{1}{1+(\delta/2)}\right)^i \right\} \leq K_1 e^{-Tq^*} \left[\dfrac{q^* T}{log(1+(\delta/2))}\right]^{Tq^*}$.

Now, consider the second inequality in (7.8). We have

$$
\frac{\partial \tilde{C}_i(\vec{\alpha})}{\partial \alpha_m} = \sum_{j=0}^{\min(i,p^*)} \left\{ \frac{\partial D^{i-j}}{\partial \alpha_m} A_j + D^{i-j} \frac{\partial A_j}{\partial \alpha_m} \right\}
$$

and

$$
\sup_{\vec{\alpha} \in \Omega_\delta} \|\frac{\partial \tilde{C}_i(\vec{\alpha})}{\partial \alpha_m}\| \leq \sum_{j=0}^{\min(i,p^*)} \left\{ \|\frac{\partial D^{i-j}}{\partial \alpha_m}\| \|A_j\| + \|D^{i-j}\| \|\frac{\partial A_j}{\partial \alpha_m}\| \right\}. \tag{7.9}
$$

Using the Jordan decomposition $D = P\Lambda P^{-1}$, we have

$$
\frac{\partial D^t}{\partial \alpha_m} = \frac{\partial P}{\partial \alpha_m} \Lambda^t P^{-1} + P \frac{\partial \Lambda^t}{\partial \alpha_m} P^{-1} + P\Lambda^t \frac{\partial P^{-1}}{\partial \alpha_m}.
$$

This implies that

$$
\|\frac{\partial D^t}{\partial \alpha_m}\| \leq \|\frac{\partial P}{\partial \alpha_m}\| \|\Lambda^t\| \|P^{-1}\| + \|P\| \|\frac{\partial \Lambda^t}{\partial \alpha_m}\| \|P^{-1}\| + \|P\| \|\Lambda^t\| \|\frac{\partial P^{-1}}{\partial \alpha_m}\|
$$

with $\dfrac{\partial \Lambda^t}{\partial \alpha_m} = \text{Diag}\left(\dfrac{\partial \Lambda_1^t}{\partial \alpha_m}, ..., \dfrac{\partial \Lambda_s^t}{\partial \alpha_m}\right)$ where

$$
\frac{\partial \Lambda_h^t}{\partial \alpha_m} = \frac{\partial \lambda_h}{\partial \alpha_m}
\begin{pmatrix}
n\lambda_h^{t-1} & \binom{t}{1}(t-1)\lambda_h^{t-2} & ... & \binom{t}{r_h-1}(t-r_h+1)\lambda_l^{t-r_h} \\
0 & n\lambda_h^{t-1} & ... & \binom{t}{r_h-2}(t-r_h+2)\lambda_h^{t-r_h+1} \\
\vdots & & \ddots & \\
0 & 0 & ... & t\lambda_h^{t-1}
\end{pmatrix}.
$$

Since $\max_h |\lambda_h| \leq \dfrac{1}{1+\delta}$, there exists a positive constant K_2 independent of $\vec{\alpha}$ such that

$$
\|\frac{\partial D^t}{\partial \alpha_m}\| \leq K_2 t^{Tq^*} \left(\frac{1}{\delta+1}\right)^t. \tag{7.10}
$$

¿From (7.6), (7.9) and (7.10), we obtain that

$$
\begin{aligned}
\sup_{\vec{\alpha} \in \Omega_\delta} \|C_{m,i}(\vec{\alpha})\| &\leq K_3 i^{q^*} \left(\frac{1}{1+\delta}\right)^i \\
&\leq K_3 i^{Tq^*} \left(\frac{1}{1+(\delta/2)}\right)^i \left(\frac{1+(\delta/2)}{1+\delta}\right)^i \\
&\leq K \rho^i
\end{aligned}
$$

33

where $\rho = \dfrac{1 + (\delta/2)}{1 + \delta} \in (0,1)$ and $K = K_3 \max_i \left\{ i^{Tq^*} \left(\dfrac{1}{1 + (\delta/2)} \right)^i \right\} \leq K_3 e^{-Tq^*} \left[\dfrac{q^* T}{log(1 + (\delta/2))} \right]^{Tq^*}$.

Which completes the proof of Lemma 7.11.

Lemma 7.12 *Let the assumptions of Theorem 2 be satisfied. For all $\vec{\alpha} \in \Omega_\delta$ and all $\vec{\omega}^2 > 0$, the matrix*

$$\mathbf{I}(\vec{\alpha}, \vec{\omega}^2) = \frac{1}{4} \lim_{N \to \infty} \operatorname{Var} \left(\sqrt{N} \frac{\partial O_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)$$

exists.

Proof. Let the diagonal matrix $\Sigma_\omega = \operatorname{Diag}\left(\vec{\omega}^2\right)$ and

$$\mathbf{Y}_n = \frac{\partial \boldsymbol{\epsilon}_n'(\vec{\alpha})}{\partial \vec{\alpha}} \Sigma_\omega^{-1} \boldsymbol{\epsilon}_n(\vec{\alpha}) = \sum_{\nu=1}^{T} \omega_\nu^{-2} \epsilon_{nT+\nu}(\vec{\alpha}) \frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}}.$$

The process $\{\mathbf{Y}_n\}$ is strictly stationary and ergodic. Moreover,

$$\frac{1}{4} \operatorname{Var} \left(\sqrt{N} \frac{\partial O_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{n'=0}^{N-1} \operatorname{Cov}\left(\mathbf{Y}_n, \mathbf{Y}_{n'}\right).$$

Denote by

$$\mathbf{I}_N(l, m) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{n'=0}^{N-1} \operatorname{Cov}\left(Y_n(l), Y_{n'}(m)\right) = \frac{1}{N} \sum_{k=1-N}^{N-1} (N - |k|) c(k)$$

where

$$c(k) = \operatorname{Cov}\left(Y_n(l), Y_{n-k}(m)\right)$$

with

$$Y_n(l) = \frac{\partial \boldsymbol{\epsilon}_n'(\vec{\alpha})}{\partial \alpha_l} \Sigma_\omega^{-1} \boldsymbol{\epsilon}_n(\vec{\alpha}) \quad \text{and} \quad Y_{n-k}(m) = \frac{\partial \boldsymbol{\epsilon}_{n-k}'(\vec{\alpha})}{\partial \alpha_m} \Sigma_\omega^{-1} \boldsymbol{\epsilon}_{n-k}(\vec{\alpha}).$$

Suppose that $k \geq 0$. From Lemma 7.11,

$$
\begin{aligned}
|c(k)| &= \left| \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \operatorname{Cov}\left(\mathbf{X}_{n-j}' C_{l,j}'(\vec{\alpha}) \Sigma_\omega^{-1} C_i(\vec{\alpha}) \mathbf{X}_{n-i}, \mathbf{X}_{n-k-j'}' C_{m,j'}'(\vec{\alpha}) \Sigma_\omega^{-1} C_{i'}(\vec{\alpha}) \mathbf{X}_{n-k-i'} \right) \right| \\
&\leq \sum_{i,j,i',j'=0}^{\infty} \left| \operatorname{Cov}\left(\mathbf{X}_{n-j}' C_{l,j}'(\vec{\alpha}) \Sigma_\omega^{-1} C_i(\vec{\alpha}) \mathbf{X}_{n-i}, \mathbf{X}_{n-k-j'}' C_{m,j'}'(\vec{\alpha}) \Sigma_\omega^{-1} C_{i'}(\vec{\alpha}) \mathbf{X}_{n-k-i'} \right) \right| \\
&\leq g_1 + g_2 + g_3 + g_4 + g_5
\end{aligned}
$$

where

$$g_1 = \sum_{i>k/2}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty}\left|\text{Cov}\left(\mathbf{X}'_{n-j}C'_{l,j}(\vec{\alpha})\Sigma_\omega^{-1}C_i(\vec{\alpha})\mathbf{X}_{n-i}, \mathbf{X}'_{n-k-j'}C'_{m,j'}(\vec{\alpha})\Sigma_\omega^{-1}C_{i'}(\vec{\alpha})\mathbf{X}_{n-k-i'}\right)\right|,$$

$$g_2 = \sum_{i'>k/2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{j'=0}^{\infty}\left|\text{Cov}\left(\mathbf{X}'_{n-j}C'_{l,j}(\vec{\alpha})\Sigma_\omega^{-1}C_i(\vec{\alpha})\mathbf{X}_{n-i}, \mathbf{X}'_{n-k-j'}C'_{m,j'}(\vec{\alpha})\Sigma_\omega^{-1}C_{i'}(\vec{\alpha})\mathbf{X}_{n-k-i'}\right)\right|,$$

$$g_3 = \sum_{j>k/2}\sum_{i=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty}\left|\text{Cov}\left(\mathbf{X}'_{n-j}C'_{l,j}(\vec{\alpha})\Sigma_\omega^{-1}C_i(\vec{\alpha})\mathbf{X}_{n-i}, \mathbf{X}'_{n-k-j'}C'_{m,j'}(\vec{\alpha})\Sigma_\omega^{-1}C_{i'}(\vec{\alpha})\mathbf{X}_{n-k-i'}\right)\right|,$$

$$g_4 = \sum_{j'>k/2}\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\left|\text{Cov}\left(\mathbf{X}'_{n-j}C'_{l,j}(\vec{\alpha})\Sigma_\omega^{-1}C_i(\vec{\alpha})\mathbf{X}_{n-i}, \mathbf{X}'_{n-k-j'}C'_{m,j'}(\vec{\alpha})\Sigma_\omega^{-1}C_{i'}(\vec{\alpha})\mathbf{X}_{n-k-i'}\right)\right|,$$

$$g_5 = \sum_{i\leq k/2}\sum_{j\leq k/2}\sum_{i'\leq k/2}\sum_{j'\leq k/2}\left|\text{Cov}\left(\mathbf{X}'_{n-j}C'_{l,j}(\vec{\alpha})\Sigma_\omega^{-1}C_i(\vec{\alpha})\mathbf{X}_{n-i}, \mathbf{X}'_{n-k-j'}C'_{m,j'}(\vec{\alpha})\Sigma_\omega^{-1}C_{i'}(\vec{\alpha})\mathbf{X}_{n-k-i'}\right)\right|.$$

By the Cauchy-Schwarz inequality, we obtain

$$g_1 \leq \sum_{i>k/2}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty}\left\{\mathbb{E}\left[\left(\mathbf{X}'_{n-j}C'_{l,j}(\vec{\alpha})\Sigma_\omega^{-1}C_i(\vec{\alpha})\mathbf{X}_{n-i}\right)^2\right]\mathbb{E}\left[\left(\mathbf{X}'_{n-k-j'}C'_{m,j'}(\vec{\alpha})\Sigma_\omega^{-1}C_{i'}(\vec{\alpha})\mathbf{X}_{n-k-i'}\right)^2\right]\right\}^{1/2}$$

$$\leq \sum_{i>k/2}\sum_{j=0}^{\infty}\sum_{i'=0}^{\infty}\sum_{j'=0}^{\infty}\left\{\mathbb{E}\left[\|\mathbf{X}'_{n-j}\|^2\|C'_{l,j}(\vec{\alpha})\|^2\|\Sigma_\omega^{-1}\|^2\|C_i(\vec{\alpha})\|^2\|\mathbf{X}_{n-i}\|^2\right]\right.$$

$$\left.\mathbb{E}\left[\|\mathbf{X}'_{n-k-j'}\|^2\|C'_{m,j'}(\vec{\alpha})\|^2\|\Sigma_\omega^{-1}\|^2\|C_{i'}(\vec{\alpha})\|^2\|\mathbf{X}_{n-k-i'}\|^2\right]\right\}^{1/2}$$

$$\leq \mathbb{E}\left[\|\mathbf{X}_n\|^4\right]\|\Sigma_\omega^{-1}\|^2\sum_{i>k/2}\|C_i(\vec{\alpha})\|\sum_{j=0}^{\infty}\|C_{l,j}(\vec{\alpha})\|\sum_{i'=0}^{\infty}\|C_{i'}(\vec{\alpha})\|\sum_{j'=0}^{\infty}\|C_{m,j'}(\vec{\alpha})\|$$

$$\leq M_1\sum_{i>k/2}\|C_i(\vec{\alpha})\| \leq \Delta_1\rho^{k/2}$$

for some positive constants M_1 and Δ_1. Using the same arguments we obtain that g_i $(i=2,3,4)$ is bounded by $\Delta_i\rho^{k/2}$.

In the other hand, Davydov inequality (Davydov, 1968) implies that, there exists a positive constant M_5 such that

$$g_5 \leq \sum_{i\leq k/2}\sum_{j\leq k/2}\sum_{i'\leq k/2}\sum_{j'\leq k/2}M_5\|\mathbf{X}'_{n-j}C'_{l,j}(\vec{\alpha})\Sigma_\omega^{-1}C_i(\vec{\alpha})\mathbf{X}_{n-i}\|_{2+\tau}\|\mathbf{X}'_{n-k-j'}C'_{m,j'}(\vec{\alpha})\Sigma_\omega^{-1}C_{i'}(\vec{\alpha})\mathbf{X}_{n-k-i'}\|_{2+\tau}$$

$$\left[\underline{\alpha}_\mathbf{X}(\min\{k+i'-j, k+j'-j, k+i'-i, k+j'-i\})\right]^{\frac{\tau}{2+\tau}}$$

$$\leq M\left[\underline{\alpha}_\mathbf{X}(k/2)\right]^{\frac{\tau}{2+\tau}}$$

for some positive constant M. Therefore, for $k\geq 0$, there exist positive constants M and Δ such that

$$|c(k)| \leq \Delta\rho^{|k|/2} + M\left[\underline{\alpha}_\mathbf{X}(k/2)\right]^{\frac{\tau}{2+\tau}}.$$

A similar inequality hold for $k < 0$. Therefore

$$\sum_{k=-\infty}^{\infty} |c(k)| < \infty.$$

Then, the dominated convergence theorem gives

$$\mathbf{I}_N(l, m) = \frac{1}{N} \sum_{k=1-N}^{N-1} (N - |k|)c(k) \underset{N \to \infty}{\longrightarrow} \sum_{k=-\infty}^{\infty} |c(k)|.$$

This implies that

$$\mathbf{I}(\vec{\alpha}, \vec{\omega}^2) = \sum_{\nu=1}^{T} \sum_{\nu'=1}^{T} \omega_\nu^{-2} \omega_{\nu'}^{-2} \sum_{k=-\infty}^{\infty} \mathbb{E}\left[\left(\epsilon_{nT+\nu}(\vec{\alpha}) \left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}} \right) \right) \left(\epsilon_{(n-k)T+\nu'}(\vec{\alpha}) \left(\frac{\partial \epsilon_{(n-k)T+\nu'}(\vec{\alpha})}{\partial \vec{\alpha}} \right) \right)' \right]$$

exists. Which completes the proof of Lemma 7.12.

Lemma 7.13 *Under the assumptions of Theorem 2, the random vector*

$$\sqrt{N} \left(\frac{\partial Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha} = \vec{\alpha}_0} \overset{\mathcal{L}}{\to} \mathcal{N}(\mathbf{0}, 4\mathbf{I}(\vec{\alpha}_0, \vec{\omega}^2)) \qquad as \ N \to \infty.$$

Proof. In the PAR case, the proof is simple and follows from the central limit theorem for mixing processes. The situation is more complicated in the PARMA case. First, note that we can show that

$$\left(\frac{\partial Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha} = \vec{\alpha}_0} - \left(\frac{\partial O_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha} = \vec{\alpha}_0} = o_p(1).$$

Thus, $\left(\frac{\partial Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha} = \vec{\alpha}_0}$ and $\left(\frac{\partial O_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha} = \vec{\alpha}_0}$ have the same asymptotic distribution. Therefore, it remains to show that $\sqrt{N} \left(\frac{\partial O_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha} = \vec{\alpha}_0} \overset{\mathcal{L}}{\to} \mathcal{N}(\mathbf{0}, 4\mathbf{I}(\vec{\alpha}_0, \vec{\omega}^2))$ as $N \to \infty$. Moreover, note that

$$\sqrt{N} \left(\frac{\partial O_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right) = \frac{2}{\sqrt{N}} \sum_{n=0}^{N-1} \mathbf{Y}_n = \frac{2}{\sqrt{N}} \sum_{n=0}^{N-1} \frac{\partial \boldsymbol{\epsilon}_n'(\vec{\alpha})}{\partial \vec{\alpha}} \Sigma_\omega^{-1} \boldsymbol{\epsilon}_n(\vec{\alpha})$$

$$= \frac{2}{\sqrt{N}} \sum_{n=0}^{N-1} \sum_{j=0}^{\infty} \left(I_d \times \mathbf{X}_{n-j}' \right) E_j(\vec{\alpha}) \Sigma_\omega^{-1} \sum_{i=0}^{\infty} C_i(\vec{\alpha}) \mathbf{X}_{n-i}$$

where $d = T(p+q)$ and $E_j(\vec{\alpha}) = (C_{1,j}(\vec{\alpha}), ..., C_{l,j}(\vec{\alpha}), ..., C_{d,j}(\vec{\alpha}))'$. Since $\left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha} = \vec{\alpha}_0}$ belongs to the Hilbert space $\mathcal{H}_X(nT + \nu - 1)$, the random variables $\epsilon_{nT+\nu}(\vec{\alpha}_0)$ and $\left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha} = \vec{\alpha}_0}$ are orthogonal and it is easy to verify that $\mathbb{E}\left[\sqrt{N} \left(\frac{\partial O_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha} = \vec{\alpha}_0} \right] = 0$.

Now, for any positive integer r, we have

$$\sqrt{N}\left(\frac{\partial O_N^{\vec{\alpha}2}(\vec{\alpha})}{\partial \vec{\alpha}}\right)_{\vec{\alpha}=\vec{\alpha}_0} = \frac{2}{\sqrt{N}}\sum_{n=0}^{N-1}(\mathbf{Y}_{n,r} - \mathbb{E}[\mathbf{Y}_{n,r}]) + \frac{2}{\sqrt{N}}\sum_{n=0}^{N-1}(\mathbf{Z}_{n,r} - \mathbb{E}[\mathbf{Z}_{n,r}])$$

where

$$
\begin{aligned}
\mathbf{Z}_{n,r} &= \mathbf{U}_{n,r} + \mathbf{V}_{n,r} \\
\mathbf{U}_{n,r} &= \sum_{i=r+1}^{\infty}\sum_{j=0}^{\infty}\left(I_d \times \mathbf{X}'_{n-j}\right)E_j(\vec{\alpha}_0)\Sigma_{\omega}^{-1}C_i(\vec{\alpha}_0)\mathbf{X}_{n-i} \\
\mathbf{V}_{n,r} &= \sum_{i=0}^{r}\sum_{j=r+1}^{\infty}\left(I_d \times \mathbf{X}'_{n-j}\right)E_j(\vec{\alpha}_0)\Sigma_{\omega}^{-1}C_i(\vec{\alpha}_0)\mathbf{X}_{n-i}
\end{aligned}
$$

and

$$\mathbf{Y}_{n,r} = \sum_{i=0}^{r}\sum_{j=0}^{r}\left(I_d \times \mathbf{X}'_{n-j}\right)E_j(\vec{\alpha}_0)\Sigma_{\omega}^{-1}C_i(\vec{\alpha}_0)\mathbf{X}_{n-i}.$$

Note that $\mathbf{Y}_{n,r}$ is function of a finite number of values of the process $\{\mathbf{X}_n\}$. Therefore, the process $(\mathbf{Y}_{n,r})_{n\in\mathbb{Z}}$ satisfies the strong mixing condition in Assumption (A3). The central limit theorem for strong mixing processes (Ibragimov, 1962) implies that

$$\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}(\mathbf{Y}_{n,r} - \mathbb{E}[\mathbf{Y}_{n,r}]) \xrightarrow[N\to\infty]{} \mathcal{N}(\mathbf{0}, \tilde{\mathbf{I}}_r).$$

Moreover, $\tilde{\mathbf{I}}_r = \lim_{N\to\infty}\text{Var}\left(\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}(\mathbf{Y}_{n,r} - \mathbb{E}[\mathbf{Y}_{n,r}])\right) \xrightarrow[r\to\infty]{} \mathbf{I}(\vec{\alpha}_0, \vec{\omega}^2))$. The result follows from a straightforward adaptation of Corollary 7.7.1 in Anderson (1971, page 426). Indeed, we have to show that

$$\mathbb{E}\left[\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}(\mathbf{Z}_{n,r} - \mathbb{E}[\mathbf{Z}_{n,r}])\left(\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}(\mathbf{Z}_{n,r} - \mathbb{E}[\mathbf{Z}_{n,r}])\right)'\right] \xrightarrow[r\to\infty]{} 0, \quad \forall N. \qquad (7.11)$$

For $m \in \{1, ..., T(p+q)\}$, we have

$$\text{Var}\left(\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}\mathbf{U}_{n,r}(m)\right) = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{n'=0}^{N-1}\text{Cov}\left(\mathbf{U}_{n,r}(m), \mathbf{U}_{n',r}(m)\right) = \frac{1}{N}\sum_{h=1-N}^{N-1}(N-|h|)c_r(h) \leq \sum_{h=-\infty}^{\infty}|c_r(h)|$$

where

$$
\begin{aligned}
c_r(h) &= \text{Cov}\left(\mathbf{U}_{n,r}(m), \mathbf{U}_{n+h,r}(m)\right) \\
&= \sum_{i=r+1}^{\infty}\sum_{j=0}^{\infty}\sum_{i'=r+1}^{\infty}\sum_{j'=0}^{\infty}\text{Cov}\left(\mathbf{X}'_{n-j}C'_{m,j}(\vec{\alpha}_0)\Sigma_{\omega}^{-1}C_i(\vec{\alpha}_0)\mathbf{X}_{n-i}, \mathbf{X}'_{n+h-j'}C'_{m,j'}(\vec{\alpha}_0)\Sigma_{\omega}^{-1}C_{i'}(\vec{\alpha}_0)\mathbf{X}_{n+h-i'}\right).
\end{aligned}
$$

Consider first the case $h \geq 0$. For $[h/2] \geq r$, using Cauchy-Schwarz inequality, we obtain that,

$$|c_r(h)| \leq M_1\sum_{i=r+1}^{\infty}\|C_i(\vec{\alpha}_0)\|\sum_{j=0}^{\infty}\|C'_{m,j}(\vec{\alpha}_0)\|\sum_{i'=r+1}^{\infty}\|C_{i'}(\vec{\alpha}_0)\|\sum_{j'=0}^{\infty}\|C'_{m,j'}(\vec{\alpha}_0)\| \leq \Delta_1\rho^r$$

37

for some positive constants M_1 and Δ_1. For $r < [h]/2$, using the Cauchy-Schwarz inequality and Davydov inequality, we get that,

$$
\begin{aligned}
|c_r(h)| \;\leq\;\; & \sum_{i=r+1}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=r+1}^{[h/2]-1} \sum_{j'=0}^{[h/2]-1} \left| \mathrm{Cov}\left(\mathbf{X}'_{n-j} C'_{m,j}(\vec{\alpha}_0) \Sigma_\omega^{-1} C_i(\vec{\alpha}_0) \mathbf{X}_{n-i}, \mathbf{X}'_{n+h-j'} C'_{m,j'}(\vec{\alpha}_0) \Sigma_\omega^{-1} C_{i'}(\vec{\alpha}_0) \mathbf{X}_{n+h-i'} \right) \right| \\
& + \sum_{i=r+1}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=[h/2]}^{\infty} \sum_{j'=0}^{\infty} \left| \mathrm{Cov}\left(\mathbf{X}'_{n-j} C'_{m,j}(\vec{\alpha}_0) \Sigma_\omega^{-1} C_i(\vec{\alpha}_0) \mathbf{X}_{n-i}, \mathbf{X}'_{n+h-j'} C'_{m,j'}(\vec{\alpha}_0) \Sigma_\omega^{-1} C_{i'}(\vec{\alpha}_0) \mathbf{X}_{n+h-i'} \right) \right| \\
& + \sum_{i=r+1}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=[h/2]}^{\infty} \left| \mathrm{Cov}\left(\mathbf{X}'_{n-j} C'_{m,j}(\vec{\alpha}_0) \Sigma_\omega^{-1} C_i(\vec{\alpha}_0) \mathbf{X}_{n-i}, \mathbf{X}'_{n+h-j'} C'_{m,j'}(\vec{\alpha}_0) \Sigma_\omega^{-1} C_{i'}(\vec{\alpha}_0) \mathbf{X}_{n+h-i'} \right) \right| \\
\leq\;\; & \Delta_2 \rho^r \left[\underline{\alpha}_{\mathbf{X}}\left(\left[\frac{|h|}{2} \right] \right) \right]^{\frac{\tau}{2+\tau}} + M_2 \rho^r \rho^{|h|/2}
\end{aligned}
$$

for some positive constants M_2 and Δ_2. The same inequality holds for $h < 0$. Therefore, there exists a constant Δ such that

$$
\sum_{h=-\infty}^{\infty} |c_r(h)| = \sum_{|h| \leq 2r+1} |c_r(h)| + \sum_{|h| \geq 2(r+1)} |c_r(h)| \leq \Delta r \rho^r + \Delta \rho^r + \Delta \rho^r \sum_k [\underline{\alpha}_{\mathbf{X}}(k)]^{\frac{\tau}{2+\tau}} \xrightarrow[r\to\infty]{} 0.
$$

This implies that

$$
\sup_N \mathrm{Var}\left(\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \mathbf{U}_{n,r}(m) \right) \xrightarrow[r\to\infty]{} 0. \tag{7.12}
$$

In a similar way, it can be shown that

$$
\sup_N \mathrm{Var}\left(\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \mathbf{V}_{n,r}(m) \right) \xrightarrow[r\to\infty]{} 0. \tag{7.13}
$$

Finally, (7.11) follows from (7.12) and (7.13). Which completes the proof of Lemma 7.13.

Lemma 7.14 *For all $\vec{\omega}^2 > 0$, almost surely the matrix*

$$
\mathbf{J}(\vec{\alpha}_0, \vec{\omega}^2) = \frac{1}{2} \lim_{N \to \infty} \left[\left(\frac{\partial^2 O_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha} \partial \vec{\alpha}'} \right)_{\vec{\alpha}=\vec{\alpha}_0} \right]
$$

exists and is strictly positive definite.

Proof. We can show that almost surely $\left\| \dfrac{\partial^2 \boldsymbol{\epsilon}_n(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m} - \dfrac{\partial^2 \mathbf{e}_n(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m} \right\| \xrightarrow[N\to\infty]{} 0$. Therefore, $\left(\dfrac{\partial^2 Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m} \right)_{\vec{\alpha}=\vec{\alpha}_0}$ and $\left(\dfrac{\partial^2 O_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m} \right)_{\vec{\alpha}=\vec{\alpha}_0}$ have almost surely the same limit. As in Lemma 7.11, it can be shown that there exists an absolutely summable sequence $(C_{l,m,i}(\vec{\alpha}))_{i\in\mathbb{N}}$ such that

$$
\frac{\partial^2 \boldsymbol{\epsilon}_n(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m} = \sum_{i=0}^{\infty} C_{l,m,i}(\vec{\alpha}) \mathbf{X}_{n-i}.
$$

38

This implies that $\dfrac{\partial^2 \boldsymbol{\epsilon}_n(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m}$ belongs to L^2. In the other hand, we have

$$
\left[\left(\frac{\partial^2 O_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha} \partial \vec{\alpha}'} \right)_{\vec{\alpha}=\vec{\alpha}_0} \right] \quad = \quad 2 \sum_{\nu=1}^{T} \omega_\nu^{-2} \frac{1}{N} \sum_{n=0}^{N-1} \epsilon_{nT+\nu}(\vec{\alpha}_0) \left(\frac{\partial^2 \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha} \partial \vec{\alpha}'} \right)_{\vec{\alpha}=\vec{\alpha}_0}
$$
$$
+ 2 \sum_{\nu=1}^{T} \omega_\nu^{-2} \frac{1}{N} \sum_{n=0}^{N-1} \left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}} \right)'_{\vec{\alpha}=\vec{\alpha}_0}
$$
$$
\underset{N\to\infty}{\longrightarrow} \quad 2 \sum_{\nu=1}^{T} \omega_\nu^{-2} \mathbb{E} \left[\epsilon_{nT+\nu}(\vec{\alpha}_0) \left(\frac{\partial^2 \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha} \partial \vec{\alpha}'} \right)_{\vec{\alpha}=\vec{\alpha}_0} \right]
$$
$$
+ 2 \sum_{\nu=1}^{T} \omega_\nu^{-2} \mathbb{E} \left[\left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}} \right)'_{\vec{\alpha}=\vec{\alpha}_0} \right].
$$

Since $\left(\dfrac{\partial^2 \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha} \partial \vec{\alpha}'} \right)_{\vec{\alpha}=\vec{\alpha}_0}$ belongs to the Hilbert space $\mathcal{H}_X(nT+\nu-1)$, the random variables $\epsilon_{nT+\nu}(\vec{\alpha}_0)$

and $\left(\dfrac{\partial^2 \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha} \partial \vec{\alpha}'} \right)_{\vec{\alpha}=\vec{\alpha}_0}$ are orthogonal and the first term in the limit is zero. Therefore,

$$
\mathbf{J}(\vec{\alpha}_0, \vec{\omega}^2) = \sum_{\nu=1}^{T} \omega_\nu^{-2} \mathbb{E} \left[\left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \left(\frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}} \right)'_{\vec{\alpha}=\vec{\alpha}_0} \right].
$$

Again, invoking the fact that $\omega_1^2, ..., \omega_T^2$ are all strictly positive and the identifiability Assumption (A1), we can conclude that $\mathbf{J}(\vec{\alpha}_0, \vec{\omega}^2)$ is a strictly positive definite matrix.

Proof of Theorem 3.2. Consider the first-order Taylor expansion around $\vec{\alpha}_0$, we obtain

$$
0 = \sqrt{N} \left(\frac{\partial Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\hat{\vec{\alpha}}_{\mathrm{WLS}}} = \sqrt{N} \left(\frac{\partial Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} + \left[\left(\frac{\partial^2 Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m} \right)_{\vec{\alpha}=\vec{\alpha}_{N,l,m}^\star} \right] \sqrt{N} \left(\hat{\vec{\alpha}}_{\mathrm{WLS}} - \vec{\alpha}_0 \right)
$$

where $\vec{\alpha}_{N,l,m}^\star$ is between $\hat{\vec{\alpha}}_{\mathrm{WLS}}$ and $\vec{\alpha}_0$. Using again a Taylor expansion, we obtain

$$
\left| \left(\frac{\partial^2 Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m} \right)_{\vec{\alpha}=\vec{\alpha}_{N,l,m}^\star} - \left(\frac{\partial^2 Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m} \right)_{\vec{\alpha}=\vec{\alpha}_0} \right| \quad \leq \quad \sup_{\vec{\alpha} \in \Omega_\delta} \left\| \frac{\partial}{\partial \vec{\alpha}} \left(\frac{\partial^2 Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m} \right) \right\| \| \vec{\alpha}_{N,l,m}^\star - \vec{\alpha}_0 \|
$$
$$
\underset{N\to\infty}{\longrightarrow} \quad 0 \quad \text{a.s..}
$$

This implies that, as $N \to \infty$

$$
\sqrt{N} \left(\hat{\vec{\alpha}}_{\mathrm{WLS}} - \vec{\alpha}_0 \right) = - \left[\left(\frac{\partial^2 Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \alpha_l \partial \alpha_m} \right)_{\vec{\alpha}=\vec{\alpha}_0} \right]^{-1} \sqrt{N} \left(\frac{\partial Q_N^{\vec{\omega}^2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} + o_p(1).
$$

Lemmas 7.13 and 7.14 complete the proof of Theorem 3.2 for LS = WLS, and thus for LS = OLS and LS = GLS. Finally, the asymptotic normality of the QLS estimator is obtained (i) by showing that an

equivalent version of Lemma 7.8 can be obtained when $O_N^{\vec{\omega}^2}(\vec{\alpha}) - Q_N^{\vec{\omega}^2}(\vec{\alpha})$ is replaced by its first and second order derivatives, and (ii) by noting that

$$\sqrt{N}\left(\frac{\partial O_N^{\hat{\vec{\sigma}}^2}(\vec{\alpha})}{\partial\vec{\alpha}} - \frac{\partial O_N^{\vec{\sigma_0}^2}(\vec{\alpha})}{\partial\vec{\alpha}}\right)_{\vec{\alpha}=\vec{\alpha}_0} = \sum_{\nu=1}^{T}\left(\frac{1}{\hat{\sigma}_\nu^2} - \frac{1}{\sigma_{0\nu}^2}\right)\frac{1}{\sqrt{N}}\sum_{n=0}^{N-1}\epsilon_{nT+\nu}(\alpha_0)\left(\frac{\partial\epsilon_{nT+\nu}}{\partial\alpha}\right)_{\vec{\alpha}=\vec{\alpha}_0} = o_P(1).$$

$\square$

Acknowledgements

References

Aknouche, A. and Bibi, A. (2009) Quasi-maximum likelihood estimation of periodic GARCH and periodic ARMA-GARCH processes. *Journal of Time Series Analysis* 30, 19-46.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York.

Balaban, B., Bayar, A. and Kan, Ö. (2001). Stock returns, seasonality and asymmetric conditional volatility in world equity markets. *Applied Economic Letters* 8, 263-268.

Berlinet, A. and Francq, C. (1997). On Bartlett's formula for nonlinear processes. *Journal of Time Series Analysis* 18, 535-552.

Basawa, I. V. and Lund, R. (2001). Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis* 22, 651-663.

Bibi, A. and Gautier, A. (2006) Propriétés dans L^2 et estimation des processus purement bilinéaires et strictement superdiagonaux à coefficients périodiques. *Revue Canadienne de Statistique / Canadian Journal of Statistics* 34, 131-148.

Bloomfield, P., Hurd, H. L., and Lund, R. (1994). Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis* 15, 127-150.

Bollerslev, T., and Ghysels, E. (1996). Periodic autoregressive conditional heteroscedasticity. *Journal of Business & Economic Statistics* 14, 139-51.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. 2nd ed., Springer, New York.

Cheng, Q. (1999). On time-reversibility of linear processes. *Biometrika* 86, 483-486.

Davydov, Y. A. (1968). On convergence of distributions generated by stationary processes. *Theory of*

Probability and its Applications 13, 691-696.

den Haan, W., and Levin, A. (1997). A practitioners guide to robust covariance matrix estimation. *in Handbook of Statistics* 15, G. Maddala and C. Rao, Eds, 309-327. Elsevier, Amsterdam.

Dunsmuir, W. (1979) A central limit theorem for parameters in stationary vector time series and its application to model for a signal observed with noise. *Annals of Statistics* 7, 490-506.

Dunsmuir, W. and Hannan, E. J. (1976). Vector linear time series models. *Advances in Applied Probability* 8, 339-364.

Francq, C., and Zakoïan, J. M. (1998a). Estimating linear representations of nonlinear processes. *Journal of Statistical Planning and Inference* 68, 145-165.

Francq, C., and Zakoïan, J. M. (1998b). Estimating the order of weak ARMA models. *Prague Stochastic'98 Proceedings*, 1, 165-168.

Francq, C., and Zakoïan, J. M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernouilli* 10, 605-637.

Francq, C., and Zakoïan, J. M. (2009). Bartlett's formula for a general class of nonlinear processes. *Journal of Time Series Analysis* 30, 449-465.

Francq, C., Roy, R. and Zakoïan, J. M. (2005). Diagnostic checking in ARMA models with uncorrelated errors. *Journal of the American Statistical Association* 100, 532–544.

Franses, P. H. and Paap, R. (2000). Modelling day-of-the-week seasonality in the S&P 500 index. *Applied Financial Economics* 10, 483-488.

Franses, P. H. and Paap, R. (2004). *Periodic Time Series Models.* Oxford University Press, Oxford.

Gardner, W., and C. Spooner (1994). The cumulant theory of cyclostationary time-series, Part I: Foundation. *IEEE Transactions on Signal Processing* 42, 3387-3408.

Gladyshev, E. G. (1961). Periodically correlated random sequences. *Soviet Mathematics* 2, 385-388.

Hannan, E. J. and Deisltler, M. (1988). *The Statistical Theory of Linear Systems.* Wiley, New York.

Hipel, K. W. and McLeod, A. I. (1994) *Time Series Modelling of Water Resources and Environmental Systems.* Elsevier, Amsterdam.

Hosoya, Y. and Taniguchi, M. (1982). A central limit theorem for stationary processes and the parameter estimation of linear processes. *Annals of Statistics* 10, 132-153. Correction (1993), 21, 1115-1117.

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability and its Applications* 7, 349-382.

Jimenez, C., McLeod, A. I., and Hipel, K. W. (1989). Kalman filter estimation for periodic autoregressive-moving average models. *Stochastic Hydrology and Hydraulics* 3, 227-240.

Jones, R. and Brelsford, W. (1967). Time series with periodic structure. *Biometrika* 54, 403-408.

Ling, S. and McAleer M. (2002). Necessary and sufficient moment conditions for the GARCH(r,s) and asymmetric power GARCH(r,s) models. *Econometric Theory* 18, 722-729.

Lund, R. (2006). A seasonal analysis of riverflow trends. *Journal of Statistical Computation and Simulation* 76, 397-405.

Lund, R. and Basawa, I. V. (2000). Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis* 21, 75-93.

Lund, R., Hurd, H., Bloomfield, P., and Smith, R. (1995). Climatological time series with periodic correlation. *Journal of Climate* 8, 2787-2809.

Lund, R., Shao, Q. and Basawa, I. (2006). Parsimonious periodic time series modeling. *Australian and New Zealand Journal of Statistics* 48, 33-47.

McLeod, A. I. (1994). Diagnostic checking periodic autoregression models with application. *Journal of Time Series Analysis* 15, 221-233. Addendum, 16, 647-648.

Osborn, D., and Smith, J. (1989). The performance of periodic autoregressive models in forecasting seasonal U.K. consumption. *Journal of Business and Economic Statistics* 7, 117-127.

Pagano, M. (1978). On periodic and multiple autoregressions. *Annals of Statistics* 6, 1310-1317.

Peiró, A. (1994). Daily seasonality in stock returns: Further international evidence. *Economics Letters* 45, 227-232.

Reinsel, G. C. (1997). *Elements of Multivariate Time Series Analysis*. 2nd ed., Springer, New York.

Romano, J. P., and Thombs, L. A. (1996). Inference for autocorrelations under weak assumptions. *Journal of the American Statistical Association* 91, 590-600.

Roy, R. and Saidi A. (2008). Temporal aggregation and systematic sampling in PARMA processes. *Computational Statistics and Data analysis* 52, 4287-4304.

Salas, J. D., and Obeysekera, J. T. B. (1992). Conceptual basis of seasonal streamflow time series models. *Journal of Hydraulic Engineering* 118, 1186-1194.

Smadi, A. A. (2005). LS estimation of periodic autoregressive models with non-Gaussian errors: A simulation study. *Journal of Statistical Computation and Simulation* 75, 207-216.

Shao, Q., and Lund, R. (2004). Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. *Journal of Time Series Analysis* 25, 359-372.

Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*. Springer, New York.

Tesfaye, Y. G., Meerschaert, M. M., and Anderson, P. L. (2006). Identification of PARMA models and their application to the modeling of river flows. *Water Resources Research* 42, W01419, doi:10.1029/2004WR003772

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

Troutman, B. (1979). Some results in periodic autoregression. *Biometrika* 66, 219-228.

Vecchia, A. V. (1985a). Maximum likelihood estimation for periodic autoregressive-moving average models. *Technometrics* 27, 375-384.

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

Wang, W., Van Gelder, P. H. A. J. M., Vrijling, J. K. and Ma, J. (2005). Testing and modelling autoregressive conditional heteroskedasticity of streamflow processes. *Nonlinear Processes in Geophysics* 12, 55-66.

Wang, W., Vrijling, J. K., Van Gelder, P. H. A. J. M. and Ma, J. (2006). Testing for nonlinearity of streamflow at different timescales processes. *Journal of Hydrology* 322, 247-268.