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
Revised version: October 14, 2010

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
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 Obey-
seker (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 esti-
mators of the periodic moving average parameters are often unsatisfactory and that the Yule-Walker
equations become more complicated. Vecchia (1985a, 1985b) investigated Gaussian maximum likeli-
hood 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
for maximum likelihood estimation of Gaussian PARMA models. An extensive simulation study con-
ducted 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 assump-
tion 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 bilin-
ear 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 encoun-
tered 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
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
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 \( \gamma_t(h) = \mathbb{E}[X_tX_{t+h}] \), \( h \in \mathbb{Z} \), 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 \( \mu_\nu \), the autoregressive order and coefficients \( p_\nu, \phi_1(\nu), ..., \phi_{p_\nu}(\nu) \), and the moving average order and coefficients \( q_\nu, \theta_1(\nu), ..., \theta_{q_\nu}(\nu) \), denoted simply PARMA \( T(p_1, ..., p_T; q_1, ..., q_T) \), 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

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

If the errors \( \epsilon_t \) 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 PARMA \( T(p; q) \) instead of PARMA \( T(p_1, ..., p_T; q_1, ..., q_T) \). The terminology periodic autoregressive (PAR) model and periodic moving average (PMA) model are respectively used when the moving average orders are null, and when
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 $E[\xi_t] = 0$, $E[\xi_t^2] = 1$ 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 $E[\epsilon_t] = 0$ for all $t$, $E[\epsilon_t \epsilon_t'] = 0$ for all $t \neq t'$, $E[\epsilon_{nT+\nu}^2] = \sigma_\nu^2 > 0$. It is a $m$-dependent white noise since the variables $\epsilon_t$ 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

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

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

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

where $\{\xi_t\}$ is a sequence of iid centered variables with unit variance, the $\alpha_i$ and $\beta_j$ are nonnegative constants, and $\omega$ 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 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 \( \mathbb{E} [ \xi_t^4 ] < \infty \), then \( \mathbb{E} [ \epsilon_t^4 ] < \infty \) (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

\[
\begin{aligned}
\epsilon_{nT+\nu} &= h_{nT+\nu} \xi_{nT+\nu}, \\
h_{nT+\nu}^2 &= \alpha_{\nu,0} + \alpha_{\nu,1} \xi_{(n-1)T+\nu}^2,
\end{aligned}
\]

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 \( \mathbb{E} [ \xi_t ] = \mathbb{E} [ \xi_t^3 ] = 0, \mathbb{E} [ \xi_t^2 ] = 1 \) 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 \( \tilde{t} \),

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

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 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

\[
\begin{align*}
    X_{nT+1} &= \tilde{\epsilon}_{nT+1} + \alpha X_{nT} \tilde{\epsilon}_{nT-1}, \\
    X_{nT+2} &= \tilde{\epsilon}_{nT+2} + \beta X_{nT+1} \epsilon_n, 
\end{align*}
\]

where the \(\tilde{\epsilon}_t\)'s are independent with \(\mathbb{E}[\tilde{\epsilon}_t] = 0\) for all \(t\), and \(\mathbb{E}[\tilde{\epsilon}_{nT+1}^2] = \tilde{\sigma}_1^2\), \(\mathbb{E}[\tilde{\epsilon}_{nT+2}^2] = \tilde{\sigma}_2^2\). 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 PMA\(_2(3)\) representation.

### 2.2.3 Causal representations of noncausal PARMA models

Let us consider the following PAR\(_2(2)\) model

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

Now, let

\[
\begin{align*}
    \epsilon_{nT+1} &= X_{nT+1} - \alpha^{-1} X_{(n-1)T+1}, \\
    \epsilon_{nT+2} &= X_{nT+2} - \beta^{-1} X_{(n-1)T+2}. 
\end{align*}
\]

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 \(\mathbb{E}[\epsilon_{nT+\nu}^2] = \sigma_\nu^2 > 0\). Thus, \(\{X_t\}\) admits a stationary PAR\(_2(2)\) representation. Moreover, we can check that \(\mathbb{E}[X_{(n-1)T+1}^3] = (1-\alpha^3)^{-1} \mathbb{E}[\tilde{\epsilon}_{nT+1}^3]\) and \(\mathbb{E}[X_{nT+1} X_{(n-1)T+1}^2] = \alpha^{-2}(1-\alpha^3)^{-1} \mathbb{E}[\tilde{\epsilon}_{nT+1}^3]\). This implies that

\[
\mathbb{E}[\epsilon_{nT+1} \epsilon_{(n-1)T+1}^2] = \mathbb{E}[X_{nT+1} X_{(n-1)T+1}^2] - \alpha^{-1} \mathbb{E}[X_{(n-1)T+1}^3] \neq 0,
\]
whenever \( E[\hat{\epsilon}_t^2] \neq 0 \). 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

\[
X_{nT+\nu} = \sum_{k=0}^{\infty} \psi_{\nu,k} \epsilon_{nT+\nu-k},
\]

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

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

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
PARMAPA_T(p, q) process \{X_{nT+\nu}\} satisfies the following difference equations

\[ X_{nT+\nu} - \sum_{i=1}^{p} \phi_i(\nu) X_{nT+\nu-i} = \epsilon_{nT+\nu} - \sum_{j=1}^{q} \theta_j(\nu) \epsilon_{nT+\nu-j}, \]  
(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.
\[ \epsilon_t = X_t - \mathbb{E}[X_t | \mathcal{H}_X(t-1)] \]

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)

\[ \Phi_0 X_n - \sum_{k=1}^{p^*} \Phi_k X_{n-k} = \Theta_0 \epsilon_n - \sum_{l=1}^{q^*} \Theta_l \epsilon_{n-l}, \]  
(3.2)

where

\[ X_n = (X_{nT+1}, ..., X_{nT+T})', \quad \epsilon_n = (\epsilon_{nT+1}, ..., \epsilon_{nT+T})', \]  
(3.3)

\( p^* = \lfloor (p - 1)/T \rfloor + 1, \) \( q^* = \lfloor (q - 1)/T \rfloor + 1, \) the matrix coefficients \( \Phi_k, k = 0, ..., p^* \) and \( \Theta_l, l = 0, ..., q^*, \) are defined by

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

\[ (\Phi_k)_{i,j} = \phi_{kT+i-j}(i), \quad \text{for } k = 1, ..., p^*, \]

and

\[ (\Theta_l)_{i,j} = \theta_{lT+i-j}(i), \quad \text{for } l = 1, ..., q^*. \]

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

\[ \Phi(B) X_n = \Theta(B) \epsilon_n \]  
(3.4)

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 at least two reasons. First, results obtained directly in terms of the univariate PARMA
representation will be more directly usable. Second, the vector ARMA representation (3.2) is not
standard, the matrices $\Phi_0$ and $\Theta_0$ are not in general the identity matrix. Rescaling the vector noise in
(3.2) via $\epsilon^*_n = \Theta_0 \epsilon_n$ and then multiplying both sides by $\Phi_0^{-1}$ leads to a vector ARMA representation.
In doing so, the covariance matrix of $\{\epsilon^*_n\}$ and the MA parameters would depend on both the AR
parameters of the PARMA representation and the vector of variances $\vec{\sigma}^2$. Here, we impose a strong
mixing condition on the process $\{X_n\}$.

In the following, we assume that

(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 (3.4) 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 $\nu$. 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

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

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

$$\epsilon_{nT+\nu}(\vec{\alpha}) = X_{nT+\nu} - \sum_{i=1}^{p} \phi_i(\nu)X_{nT+\nu-i} + \sum_{j=1}^{q} \theta_j(\nu)\epsilon_{nT+\nu-j}(\vec{\alpha}). \tag{3.5}$$
Note that, almost surely, \( e_{nT+\nu}(\tilde{\alpha}) = e_{nT+\nu} \) for all \( n \in \mathbb{Z} \) and \( \nu \in \{1, \ldots, T\} \). Moreover, \( e_{nT+\nu}(\tilde{\alpha}) \) can be approximated recursively in \( t \) via a truncated version of (3.5)

\[
e_{nT+\nu}(\tilde{\alpha}) = X_{nT+\nu} - \sum_{i=1}^{p} \phi_i(\nu)X_{nT+\nu-i} + \sum_{j=1}^{q} \theta_j(\nu)e_{nT+\nu-j}(\tilde{\alpha}),
\]

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

Let \( \delta \) be a strictly positive constant chosen such that \( \tilde{\alpha}_0 \) belongs to the interior of the compact set

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

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

\[
S_N(\hat{\tilde{\alpha}}_{OLS}) = \min_{\tilde{\alpha} \in \Omega_\delta} S_N(\tilde{\alpha}),
\]

where

\[
S_N(\tilde{\alpha}) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} e^2_{nT+\nu}(\tilde{\alpha}).
\]

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

\[
Q^2_N(\hat{\tilde{\alpha}}_{WLS}) = \min_{\tilde{\alpha} \in \Omega_\delta} Q^2_N(\tilde{\alpha})
\]

where

\[
Q^2_N(\tilde{\alpha}) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} \omega^2_{\nu} e^2_{nT+\nu}(\tilde{\alpha}).
\]

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

\[
\hat{\tilde{\alpha}}_{GLS} = \hat{\tilde{\alpha}}_{WLS}^2.
\]

The GLS estimator assumes that \( \sigma_0^2 \) is known. In practice, this parameter has also to be estimated.

Given any consistent estimator \( \hat{\sigma}^2 \) of \( \sigma_0^2 \), a quasi-generalized least squares (QLS) estimator of \( \tilde{\alpha}_0 \) is defined by

\[
\hat{\tilde{\alpha}}_{QLS} = \hat{\tilde{\alpha}}_{WLS}^2.
\]

One possible consistent estimator of \( \sigma_0^2 \) is

\[
\hat{\sigma}^2_{\nu} = \frac{1}{N} \sum_{n=0}^{N-1} e^2_{nT+\nu}(\hat{\tilde{\alpha}}_{OLS}).
\]

To establish the consistency of the least squares estimators, an additional assumption is needed.
(A2) The T-dimensional white noise \( \{ \epsilon_n = (\epsilon_{nT+1}, \ldots, \epsilon_{nT+\nu}, \ldots, \epsilon_{nT+T})', n \in \mathbb{Z} \} \) in (3.4) is strictly stationary and ergodic.

**Theorem 3.1** Suppose that \( \{ X_{nT+\nu} \} \) is a PARMA\(_T\)\((p,q)\) process. Let \( \hat{\alpha}_{OLS}, \hat{\alpha}_{WLS}, \hat{\alpha}_{GLS} \) and \( \hat{\alpha}_{QLS} \) be the least squares estimators defined by (3.7), (3.9), (3.11) and (3.12). Then, under Assumptions (A1) and (A2), and for any \( \bar{\omega}^2 = (\omega_1^2, \ldots, \omega_T^2) > 0 \), where the inequality applies element-wise, we have

\[ \hat{\alpha}_{OLS} \to \bar{\alpha}_0, \quad \hat{\omega}_{WLS}^2 \to \bar{\omega}_0, \quad \hat{\alpha}_{GLS} \to \bar{\alpha}_0, \quad \hat{\alpha}_{QLS} \to \bar{\alpha}_0, \quad \text{almost surely as } N \to \infty. \]

Let \( F_{-\infty}^m \) and \( F_{m+h}^{+\infty} \) be the \( \sigma \)-fields generated by \( \{ X_n, n \leq m \} \) and \( \{ X_n, n \geq m + h \} \) respectively.

The strong mixing coefficients of the \( T \)-variate stationary process \( \{ X_n, n \in \mathbb{Z} \} \) are defined by

\[ \alpha_X(h) = \sup_{A \in F_{-\infty}^m, B \in F_{m+h}^{+\infty}} |P(A \cap B) - P(A)P(B)|. \]

Let \( ||Z||_r = [E||Z||^r]^{1/r} \) 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 \( \{ X_n \} \) is such that for some \( \tau > 0 \), \( ||X_n||_{4+2\tau} < \infty \) and

\[ \sum_{h=0}^{\infty} [\alpha_X(h)]^{2\tau} < \infty. \]

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 \( I(\bar{\alpha}, \bar{\omega}^2) \) and \( J(\bar{\alpha}, \bar{\omega}^2) \) defined below requires \( ||X_n||_4 < \infty \).

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

\[ \sqrt{N}(\hat{\alpha}_{LS} - \bar{\alpha}_0) \xrightarrow{d} N(0, V_{LS}) \]

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

\[
V_{OLS} = V(\bar{\alpha}_0, (1, \ldots, 1)'), \quad V_{WLS} = V(\bar{\alpha}_0, \bar{\omega}^2), \quad V_{GLS} = V_{QLS} = V(\bar{\alpha}_0, \bar{\omega}_0^2),
\]

with

\[
V(\bar{\alpha}_0, \bar{\omega}^2) = (J(\bar{\alpha}_0, \bar{\omega}^2))^{-1} I(\bar{\alpha}_0, \bar{\omega}^2) (J(\bar{\alpha}_0, \bar{\omega}^2))^{-1}, \quad \text{and} \quad I(\bar{\alpha}_0, \bar{\omega}^2) = \sum_{\nu=1}^{T} \sum_{\nu'=1}^{T} \omega_\nu^{-2} \omega_{\nu'}^{-2} \sum_{k=-\infty}^{\infty} \mathbb{E} \left[ \left( \epsilon_{\nu}(\bar{\alpha}_0) \frac{\partial \epsilon_{\nu}(\bar{\alpha})}{\partial \bar{\alpha}} \right) \right]_{\bar{\alpha} = \bar{\alpha}_0} \left( \epsilon_{kT+\nu}(\bar{\alpha}_0) \frac{\partial \epsilon_{kT+\nu}(\bar{\alpha})}{\partial \bar{\alpha}} \right)_{\bar{\alpha} = \bar{\alpha}_0} \right]
\]

and

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

12
Remark 3.1 In the periodic AR case, the OLS and WLS estimators coincide. Indeed we have in this particular case \( \hat{\alpha} = (\hat{\phi}(1)', \ldots, \hat{\phi}(T)')' \) and

\[
Q^2_N(\hat{\alpha}) = \sum_{\nu=1}^{T} \omega^{-2}_{\nu} \frac{1}{N} \sum_{n=0}^{N-1} \epsilon_{nT+\nu}(\hat{\phi}(\nu)).
\]

Thus the WLS estimator does not depend on the vector of weights \( \omega^2 \):

\[
\hat{\alpha} = (\hat{\phi}(1)', \ldots, \hat{\phi}(T)')', \quad \text{where} \quad \hat{\phi}(\nu) = \arg \min_{\phi(\nu)} \frac{1}{N} \sum_{n=0}^{N-1} \epsilon_{nT+\nu}(\hat{\phi}(\nu)).
\]

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

Remark 3.2 In the strong PARMA setting, i.e., when \( \{\sigma_0^{-1} \epsilon_{nT+\nu}\} \) 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

\[
\mathbf{I}(\hat{\alpha}_0, \sigma_0^2) = \sum_{\nu=1}^{T} \sigma_0^{-2} \mathbb{E} \left[ \frac{\partial \epsilon_{\nu}(\hat{\alpha})}{\partial \hat{\alpha}} \right]_{\hat{\alpha}=\hat{\alpha}_0} \left( \frac{\partial \epsilon_{\nu}(\hat{\alpha})}{\partial \hat{\alpha}} \right)'_{\hat{\alpha}=\hat{\alpha}_0} = \mathbf{J}(\hat{\alpha}_0, \sigma_0^2).
\]

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

\[
\mathbf{V}_{QLS} = (\mathbf{J}(\hat{\alpha}_0, \sigma_0^2))^{-1}.
\] (3.14)

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 \( \mathbb{E} [\epsilon_t^2 | \mathcal{F}_{t-1}] = \mathbb{E} [\epsilon_t^2] \) where \( \mathcal{F}_t \) is the \( \sigma \)-field spanned by \( \{\epsilon_s, s \leq t\} \), the QLS estimator is an optimal LS estimator in the sense that

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

Indeed, let the random vector

\[
S^2_N = \mathbf{J}(\hat{\alpha}_0, \omega^2)^{-1} \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} \omega^{-2}_{\nu} \epsilon_{nT+\nu}(\hat{\alpha}_0) \left( \frac{\partial \epsilon_{nT+\nu}(\hat{\alpha})}{\partial \hat{\alpha}} \right)_{\hat{\alpha}=\hat{\alpha}_0}.
\]
We have

\[ \text{Cov} \left( S_N^2, S_N^2 \right) = \left\{ J(\hat{\alpha}_0, \omega^2) \right\}^{-1} \frac{1}{N} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} \omega_{\nu}^{-2} \sigma_{\nu}^{-2} E \epsilon_{nT+\nu}^2 \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right) \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)' \left\{ J(\hat{\alpha}_0, \sigma_0^2) \right\}^{-1} \]

\[ \rightarrow \left\{ J(\hat{\alpha}_0, \sigma_0^2) \right\}^{-1}. \]

Therefore

\[ \lim_{N \to \infty} \text{Var} \left( S_N^2 - S_N^2 \right) = \left\{ J(\hat{\alpha}_0, \omega^2) \right\}^{-1} - \left\{ J(\hat{\alpha}_0, \sigma_0^2) \right\}^{-1} = V_{\text{WLS}} - V_{\text{QLS}} \]

and the conclusion follows.

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

\[ \hat{J}(\alpha_0, \omega^2) = \sum_{\nu=1}^{T} \omega_{\nu}^{-2} \left[ \frac{1}{N} \sum_{n=0}^{N-1} \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right) \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)' \right] \].

Note that the matrix \((2\pi)^{-1} \Sigma(\alpha_0, \omega^2)\) is the spectral density at frequency zero of the process

\[ \Upsilon_n = \sum_{\nu=1}^{T} \omega_{\nu}^{-2} \epsilon_{nT+\nu}(\tilde{\alpha}_0) \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right) \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)' \].

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 \(p = 0, \ldots, p_{\text{max}}\) to the series \(\hat{\Upsilon}_n, n = 0, \ldots, N-1\), where \(\hat{\Upsilon}_n\) is obtained by replacing \(\epsilon_{nT+\nu}(\tilde{\alpha}_0)\) and its derivatives by \(\epsilon_{nT+\nu}(\hat{\alpha}_0)\) and its derivatives in \(\Upsilon_n\); ii) selecting the order \(p\) which minimizes an information criteria and approximating \(I(\alpha_0, \omega^2)\) by \((2\pi)\) times the spectral density at frequency zero of the estimated \(\text{VAR}(p)\) model. Hereafter, we used the AIC model selection criterion with \(p_{\text{max}} = 25\).

**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 \(\tilde{\phi}(\nu)\) in a weak PAR model with means zero only involves the sequence \(\{\epsilon_{nT+\nu}, n = 0, \ldots, N-1\}\). Therefore, the sample ACF and PACF of the \(\nu\)-th season of the original series can be used to identify the AR order \(p_{\nu}\). 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 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 \(T = 2\):

\[
\epsilon_{nT+\nu} = \sigma_\nu \prod_{j=0}^{m} \xi_{nT+\nu-j}, \quad \nu = 1, 2, \quad (4.1)
\]

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

From a realization \(X_t = \epsilon_t, t = 1, \ldots, NT\), of that weak white noise, suppose that a statistician fits the following PAR\(_2\)(1) model:

\[
\begin{align*}
X_{nT+1} - \phi_1 X_{nT} &= \epsilon_{nT+1}, \\
X_{nT+2} - \phi_2 X_{nT+1} &= \epsilon_{nT+2}.
\end{align*}
\]

The true parameter values are \(\phi_1 = \phi_2 = 0\) and \(\sigma^2 = \sigma_0^2\). According to Theorem 3.2, some moment calculations show that the asymptotic covariance matrix of the QLS estimator of \(\sqrt{N} \left( \hat{\mathbf{\phi}}, \hat{\mathbf{\sigma}}^2 \right)^t\) is given by

\[
V^{(w)}_{QLS} = \left( J(\hat{\mathbf{\alpha}}, \hat{\mathbf{\sigma}}_0^2) \right)^{-1} \mathbf{I}(\hat{\mathbf{\alpha}}, \hat{\mathbf{\sigma}}_0^2) \left( J(\hat{\mathbf{\alpha}}, \hat{\mathbf{\sigma}}_0^2) \right)^{-1} = \kappa^m \begin{pmatrix} \frac{\sigma_0^2}{\sigma_{10}} & 0 \\ 0 & \frac{\sigma_0^2}{\sigma_{20}} \end{pmatrix}.
\]

(4.3)

On the other hand from (3.14), the corresponding asymptotic covariance matrix under the assumption of a strong noise is equal to

\[
V^{(s)}_{QLS} = \left( J(\hat{\mathbf{\alpha}}, \hat{\mathbf{\sigma}}_0^2) \right)^{-1} = \begin{pmatrix} \frac{\sigma_0^2}{\sigma_{10}} & 0 \\ 0 & \frac{\sigma_0^2}{\sigma_{20}} \end{pmatrix}.
\]

(4.4)
It is clear that $V_{QLS}^{(w)}$ and $V_{QLS}^{(s)}$ can be very different. For example, if the iid sequence $\{\xi_t\}$ is $N(0, 1)$, $E[\xi_t^4] = 3$, 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 $\epsilon_t$.

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 $T = 2$:

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

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

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

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 $\kappa = E[\xi_t^4]$ 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) model (4.2) for that series. Once again, the true parameter values are $\phi_1 = \phi_2 = 0$ and $\sigma^2 = \sigma_0^2$. Direct computation of the matrices in (3.13) lead to the following asymptotic covariance matrix of the QLS estimators of $\sqrt{N} (\hat{\phi}_1, \hat{\phi}_2)'$:

$$V_{QLS}^{(w)} = (J(\bar{a}_0, \bar{\sigma}_0^2))^{-1} I(\bar{a}_0, \bar{\sigma}_0^2) (J(\bar{a}_0, \bar{\sigma}_0^2))^{-1} = \left\{ 1 + \frac{(\kappa - 1)\alpha}{1 - \kappa \alpha^2} \right\} \begin{pmatrix} \sigma_{01}^2 & 0 \\ 0 & \sigma_{02}^2 \end{pmatrix} .$$

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 $N(0, 1)$ sequence, and $\alpha = 0.5$, $V_{QLS}^{(w)} = 5V_{QLS}^{(s)}$.

With a series of 2000 observations ($N = 1000$) and if $\sigma_{01}^2 = \sigma_{02}^2$, the standard error of the QLS estimators of $\phi_1$ and $\phi_2$ 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 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
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 ($\chi^2$), 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 $\phi_1$ and $\phi_2$ 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 $\phi_1$, $\phi_2$, $\sigma_1$ and $\sigma_2$. 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 $\hat{\phi}_1$ and $\hat{\phi}_2$ 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 (3.14) or for a weak noise as provided by (3.13). 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.
Table 1. Empirical and asymptotic standard errors of the OLS estimators of $\phi_1$ and $\phi_2$ 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.

<table>
<thead>
<tr>
<th></th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N$  $t$  LN  $\chi^2$  Exp  Gam</td>
<td>WN1  WN2</td>
<td>$\alpha = 0.3$  $\alpha = 0.4$  $\alpha = 0.5$</td>
</tr>
<tr>
<td>$NT = 100$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical stand.</td>
<td>$S_{\hat{\phi}_1}$: .1487</td>
<td>$S_{\hat{\phi}_1}$: .1406</td>
<td></td>
</tr>
<tr>
<td></td>
<td>exps $0.0609$</td>
<td>exps $0.0609$</td>
<td></td>
</tr>
<tr>
<td>Asymptotic</td>
<td>$\sqrt{V(\hat{\phi}_i)}$:</td>
<td>$\sqrt{V(\hat{\phi}_i)}$:</td>
<td></td>
</tr>
<tr>
<td>stand. errors</td>
<td>.1414 .1414 .1414 .1414 .1414</td>
<td>.1414 .1414 .1414 .1414 .1414</td>
<td></td>
</tr>
<tr>
<td>Rejection</td>
<td>$H_0^{(1)}$: .0640 .0700 .0640</td>
<td>$H_0^{(1)}$: .0640 .0700 .0640</td>
<td></td>
</tr>
<tr>
<td>frequencies</td>
<td>$H_0^{(2)}$: .0660 .0660 .0660</td>
<td>$H_0^{(2)}$: .0660 .0660 .0660</td>
<td></td>
</tr>
<tr>
<td>$NT = 200$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical stand.</td>
<td>$S_{\hat{\phi}_1}$: .1013</td>
<td>$S_{\hat{\phi}_1}$: .0994</td>
<td></td>
</tr>
<tr>
<td></td>
<td>exps $0.0609$</td>
<td>exps $0.0609$</td>
<td></td>
</tr>
<tr>
<td>Asymptotic</td>
<td>$\sqrt{V(\hat{\phi}_i)}$:</td>
<td>$\sqrt{V(\hat{\phi}_i)}$:</td>
<td></td>
</tr>
<tr>
<td>stand. errors</td>
<td>.1000 .1000 .1000 .1000 .1000</td>
<td>.1000 .1000 .1000 .1000 .1000</td>
<td></td>
</tr>
<tr>
<td>Rejection</td>
<td>$H_0^{(1)}$: .0570 .0630 .0580</td>
<td>$H_0^{(1)}$: .0570 .0630 .0580</td>
<td></td>
</tr>
<tr>
<td>frequencies</td>
<td>$H_0^{(2)}$: .0560 .0600 .0550</td>
<td>$H_0^{(2)}$: .0560 .0600 .0550</td>
<td></td>
</tr>
<tr>
<td>$NT = 2000$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empirical stand.</td>
<td>$S_{\hat{\phi}_1}$: .0311</td>
<td>$S_{\hat{\phi}_1}$: .0322</td>
<td></td>
</tr>
<tr>
<td></td>
<td>exps $0.0609$</td>
<td>exps $0.0609$</td>
<td></td>
</tr>
<tr>
<td>Asymptotic</td>
<td>$\sqrt{V(\hat{\phi}_i)}$:</td>
<td>$\sqrt{V(\hat{\phi}_i)}$:</td>
<td></td>
</tr>
<tr>
<td>stand. errors</td>
<td>.0316 .0316 .0316 .0316 .0316</td>
<td>.0316 .0316 .0316 .0316 .0316</td>
<td></td>
</tr>
<tr>
<td>Rejection</td>
<td>$H_0^{(1)}$: .0400 .0460 .0510</td>
<td>$H_0^{(1)}$: .0400 .0460 .0510</td>
<td></td>
</tr>
<tr>
<td>frequencies</td>
<td>$H_0^{(2)}$: .0510 .0450 .0480</td>
<td>$H_0^{(2)}$: .0510 .0450 .0480</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0510 .0450 .0480 .0520 .0530</td>
<td>.0510 .0450 .0480 .0520 .0530</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.0460 .0490 .0520 .0530 .0460</td>
<td>.0460 .0490 .0520 .0530 .0460</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.2460 .4840 .2460 .4840 .2460</td>
<td>.2460 .4840 .2460 .4840 .2460</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.1200 .2120 .1200 .2120 .1200</td>
<td>.1200 .2120 .1200 .2120 .1200</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.1730 .2840 .1730 .2840 .1730</td>
<td>.1730 .2840 .1730 .2840 .1730</td>
<td></td>
</tr>
</tbody>
</table>
5.2 Using an estimated covariance structure

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

\[ X_{nT+\nu} = \mu_\nu + \phi_\nu X_{nT+\nu-1} + \epsilon_{nT+\nu} - \theta_\nu \epsilon_{nT+\nu-1}, \quad \nu = 1, 2, \quad (5.1) \]

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 is a 3-dependent process as defined by Equation (2.2) with \( m = 3 \), the \( \xi_t \) 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 (MASE$^s$) and a weak noise (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, MASE$^s$ and MASE$^w$ are close to each other and also reasonably close to the corresponding ESE for all parameters but \( \mu_2 \). For this latter parameter, both asymptotic standard errors underestimate the true standard error even with 2000 observations. With a weak noise, MASE$^s$ is quite far from the corresponding ESE for all parameters except \( \mu_1 \). 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 \( \mu_1 \), MASE$^w$ is much closer to ESE than 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 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%.

<table>
<thead>
<tr>
<th>NT</th>
<th>Noise</th>
<th>Statistics</th>
<th>$\mu_1$</th>
<th>$\mu_2$</th>
<th>$\phi_1$</th>
<th>$\phi_2$</th>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>S</td>
<td>Bias</td>
<td>.007</td>
<td>.007</td>
<td>-.007</td>
<td>-.006</td>
<td>-.013</td>
<td>-.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ESE</td>
<td>.258</td>
<td>.235</td>
<td>.040</td>
<td>.031</td>
<td>.103</td>
<td>.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(s)</td>
<td>.281</td>
<td>.137</td>
<td>.044</td>
<td>.023</td>
<td>.120</td>
<td>.035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(w)</td>
<td>.307</td>
<td>.149</td>
<td>.043</td>
<td>.023</td>
<td>.119</td>
<td>.035</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>Bias</td>
<td>-.018</td>
<td>-.015</td>
<td>-.008</td>
<td>-.005</td>
<td>-.022</td>
<td>.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ESE</td>
<td>1.218</td>
<td>1.141</td>
<td>.072</td>
<td>.053</td>
<td>.269</td>
<td>.110</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(s)</td>
<td>.240</td>
<td>.119</td>
<td>.037</td>
<td>.020</td>
<td>.101</td>
<td>.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(w)</td>
<td>.245</td>
<td>.121</td>
<td>.040</td>
<td>.026</td>
<td>.270</td>
<td>.070</td>
</tr>
<tr>
<td>1000</td>
<td>S</td>
<td>Bias</td>
<td>.000</td>
<td>.000</td>
<td>-.005</td>
<td>-.004</td>
<td>-.010</td>
<td>-.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ESE</td>
<td>.173</td>
<td>.158</td>
<td>.026</td>
<td>.021</td>
<td>.072</td>
<td>.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(s)</td>
<td>.204</td>
<td>.099</td>
<td>.031</td>
<td>.016</td>
<td>.085</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(w)</td>
<td>.216</td>
<td>.105</td>
<td>.031</td>
<td>.016</td>
<td>.084</td>
<td>.025</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>Bias</td>
<td>.003</td>
<td>.004</td>
<td>-.008</td>
<td>-.005</td>
<td>-.015</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ESE</td>
<td>.157</td>
<td>.144</td>
<td>.055</td>
<td>.041</td>
<td>.222</td>
<td>.089</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(s)</td>
<td>.184</td>
<td>.091</td>
<td>.028</td>
<td>.015</td>
<td>.077</td>
<td>.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(w)</td>
<td>.190</td>
<td>.094</td>
<td>.052</td>
<td>.023</td>
<td>.233</td>
<td>.061</td>
</tr>
<tr>
<td>2000</td>
<td>S</td>
<td>Bias</td>
<td>.002</td>
<td>.002</td>
<td>-.003</td>
<td>-.002</td>
<td>.000</td>
<td>-.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ESE</td>
<td>.116</td>
<td>.107</td>
<td>.019</td>
<td>.015</td>
<td>.049</td>
<td>.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(s)</td>
<td>.146</td>
<td>.071</td>
<td>.022</td>
<td>.011</td>
<td>.060</td>
<td>.018</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(w)</td>
<td>.151</td>
<td>.073</td>
<td>.021</td>
<td>.011</td>
<td>.060</td>
<td>.018</td>
</tr>
<tr>
<td></td>
<td>W</td>
<td>Bias</td>
<td>-.006</td>
<td>-.005</td>
<td>-.004</td>
<td>-.002</td>
<td>-.016</td>
<td>.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ESE</td>
<td>.113</td>
<td>.104</td>
<td>.044</td>
<td>.032</td>
<td>.170</td>
<td>.072</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(s)</td>
<td>.139</td>
<td>.068</td>
<td>.021</td>
<td>.011</td>
<td>.057</td>
<td>.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MASE(w)</td>
<td>.143</td>
<td>.070</td>
<td>.042</td>
<td>.018</td>
<td>.192</td>
<td>.050</td>
</tr>
</tbody>
</table>

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
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.

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>NT</th>
<th>Test Statistic</th>
<th>Nominal Level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Strong noise</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$H_0$</td>
<td>500</td>
<td>WS</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WW</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>WS</td>
<td>.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WW</td>
<td>.7</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>WS</td>
<td>.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WW</td>
<td>1.1</td>
</tr>
<tr>
<td>$H_1$</td>
<td>500</td>
<td>WS</td>
<td>6.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WW</td>
<td>8.1</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>WS</td>
<td>26.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WW</td>
<td>26.8</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>WS</td>
<td>78.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WW</td>
<td>78.4</td>
</tr>
</tbody>
</table>

Table 3. Rejection frequencies (%) 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 PARMA2(1, 1) model (5.1). Under $H_0$, $\alpha = (0, 0.8, 0.8, -0.5, -0.5)'$ and under $H_1$, $\alpha = (0.05, -0.05, 0.8, 0.75, -0.5, -0.45)'$.

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 (Swiss), 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 $\sigma_t^2$ is the so-called volatility. For the GARCH-type models, $\sigma_t$ 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(1) model to each series:

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

(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

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

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.
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).

<table>
<thead>
<tr>
<th>Index</th>
<th>$H_{01}$</th>
<th>$H_{02}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WS WW</td>
<td>WS WW</td>
</tr>
<tr>
<td>CAC 40</td>
<td>.026 .005</td>
<td>.022 .013</td>
</tr>
<tr>
<td>DAX</td>
<td>.013 .038</td>
<td>.009 .035</td>
</tr>
<tr>
<td>FTSE 120</td>
<td>.165 .001</td>
<td>.180 .014</td>
</tr>
<tr>
<td>SMI</td>
<td>.001 .005</td>
<td>.002 .008</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Index</th>
<th>CAC 40</th>
<th>DAX</th>
<th>FTSE 120</th>
<th>SMI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4165</td>
<td>4200</td>
<td>4165</td>
<td>4185</td>
</tr>
<tr>
<td>Day</td>
<td>$\hat{\mu}<em>\nu$ $\hat{\phi}</em>\nu$ $\hat{\sigma}_\nu$</td>
<td>$\hat{\mu}<em>\nu$ $\hat{\phi}</em>\nu$ $\hat{\sigma}_\nu$</td>
<td>$\hat{\mu}<em>\nu$ $\hat{\phi}</em>\nu$ $\hat{\sigma}_\nu$</td>
<td>$\hat{\mu}<em>\nu$ $\hat{\phi}</em>\nu$ $\hat{\sigma}_\nu$</td>
</tr>
<tr>
<td>Monday</td>
<td>.045 (0.056) -.081 (.124) 1.91 (.072)</td>
<td>.132c (.072) -.060 (.104) 2.06 (.046)</td>
<td>.061 (.046) -.042 (.072) 1.51 (.050)</td>
<td>.038 (.050) -.008 (.120) 1.66 (.041)</td>
</tr>
<tr>
<td>Tuesday</td>
<td>-.012 (.040) -.045 (.029) 1.35 (.039)</td>
<td>.013 (.039) -.031 (.042) 1.41 (.041)</td>
<td>.003 (.041) -.056 (.037) 1.09 (.037)</td>
<td>.000 (.037) -.016 (.042) 1.13 (.042)</td>
</tr>
<tr>
<td>Wednesday</td>
<td>-.055 (.056) -.055 (.036) 1.41 (.058)</td>
<td>-.090 (.058) -.066c (.038) 1.43 (.050)</td>
<td>-.071 (.050) -.015 (.031) 1.12 (.044)</td>
<td>-.038 (.044) -.008 (.052) 1.14 (.042)</td>
</tr>
<tr>
<td>Thursday</td>
<td>.001 (.067) -.008 (.067) 1.46 (.062)</td>
<td>-.044 (.062) -.008 (.072) 1.47 (.043)</td>
<td>.004 (.043) .002 (.085) 1.15 (.055)</td>
<td>-.011 (.055) .061 (.071) 1.24 (.040)</td>
</tr>
<tr>
<td>Friday</td>
<td>.021 (.042) .092b (.042) 1.35 (.043)</td>
<td>-.010 (.043) .086c (.048) 1.38 (.028)</td>
<td>.004 (.028) .052 (.047) 1.18 (.036)</td>
<td>.010 (.036) .159a (.040) 1.15 (.036)</td>
</tr>
</tbody>
</table>

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 \( K \), \( \Delta \) 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 \( \epsilon_t(\tilde{\alpha}) \) rather than the truncated ones \( (\epsilon_t(\tilde{\alpha})) \) does not alter the asymptotic results and we will use the criterion

\[
O_N^2(\tilde{\alpha}) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{\nu=1}^{T} \omega_{\nu}^{-2} \epsilon_{nT+\nu}(\tilde{\alpha}).
\]  

(7.1)

instead of \( Q_N^2(\tilde{\alpha}) \) defined by (3.10).

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

\[
\epsilon_n(\tilde{\alpha}) = \sum_{i=0}^{\infty} C_i(\tilde{\alpha}) X_{n-i}.
\]
Then, there exists a constant $K$ such that for all $i \in \mathbb{N}$,
\[
\sup_{\tilde{\alpha} \in \Omega_{\delta}} \| C_i(\tilde{\alpha}) \| \leq K_i^{T q^*} \left( \frac{1}{1 + \delta} \right)^i.
\]

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

Now, consider the case $q^* > 0$. Define the $q^* T \times 1$ vectors
\[
\tilde{X}_n = \begin{pmatrix} X_n \\ 0_{T \times 1} \\ \vdots \\ 0_{T \times 1} \end{pmatrix}, \quad \mathbf{e}_n(\tilde{\alpha}) = \begin{pmatrix} \mathbf{e}_n(\tilde{\alpha}) \\ \mathbf{e}_{n-1}(\tilde{\alpha}) \\ \vdots \\ \mathbf{e}_{n-q^*+1}(\tilde{\alpha}) \end{pmatrix}, \quad \mathbf{e}_n(\tilde{\alpha}) = \begin{pmatrix} \mathbf{e}_n(\tilde{\alpha}) \\ \mathbf{e}_{n-1}(\tilde{\alpha}) \\ \vdots \\ \mathbf{e}_{n-q^*+1}(\tilde{\alpha}) \end{pmatrix},
\]
and the $q^* T \times q^* T$ companion matrices
\[
A_i = \begin{pmatrix} \Theta_0^{-1} \Phi_i & 0_{T \times T} & \ldots & 0_{T \times T} \\ 0_{T \times T} & 0_{T \times T} & \ldots & 0_{T \times T} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{T \times T} & 0_{T \times T} & \ldots & 0_{T \times T} \end{pmatrix}, \quad i = 1, \ldots, p^*,
\]
\[
D = \begin{pmatrix} -\Theta_0^{-1} \Theta_1 & -\Theta_0^{-1} \Theta_2 & \ldots & -\Theta_0^{-1} \Theta_{q^*-1} & -\Theta_0^{-1} \Theta_{q^*} \\ \text{I}_{T \times T} & 0_{T \times T} & \ldots & 0_{T \times T} & 0_{T \times T} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{T \times T} & 0_{T \times T} & \ldots & \text{I}_{T \times T} & 0_{T \times T} \end{pmatrix}.
\]

It is easy to check that
\[
A_0 \tilde{X}_n + \sum_{i=1}^{p^*} A_i \tilde{X}_{n-i} = \mathbf{e}_n(\tilde{\alpha}) - D \mathbf{e}_{n-1}(\tilde{\alpha}).
\]

This implies that
\[
\mathbf{e}_n(\tilde{\alpha}) = \sum_{j=0}^{\infty} D^j \left( A_0 \tilde{X}_{n-j} + A_1 \tilde{X}_{n-j-1} + \ldots + A_p \tilde{X}_{n-j-p} \right) = \sum_{i=0}^{\infty} C_i(\tilde{\alpha}) \tilde{X}_{n-i},
\]
where
\[
C_i(\tilde{\alpha}) = \sum_{j=0}^{\min(i, p^*)} D^{i-j} A_j.
\]

Again, using a multiplicative norm, it can be shown that there exists a constant $K_1$ independent of $\tilde{\alpha}$ such that
\[
\| C_i(\tilde{\alpha}) \| \leq \sum_{j=0}^{p^*} \| D^{i-j} \| \| A_j \| \leq K_1^{T q^*} \left( \frac{1}{1 + \delta} \right)^i.
\]
The conclusion follow from the fact that $C_i(\tilde{\alpha})$ is the matrix in the first row and the first column of the block matrix $\tilde{C}_i(\tilde{\alpha})$.

**Lemma 7.2** We have

$$
\mathbb{E} \left[ \sup_{\tilde{\alpha} \in \Omega_2} \epsilon_{nT+\nu}^2(\tilde{\alpha}) \right] < \infty.
$$

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

$$
\sup_{\tilde{\alpha} \in \Omega_2} |\epsilon_{nT+\nu}(\tilde{\alpha})| \leq K \sum_{i=0}^{\infty} i^{Tq^*} \left( \frac{1}{1+\delta} \right)^i |X_{nT+\nu-i}|.
$$

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 $\epsilon_n(\tilde{\alpha})$ as defined by (3.3). For any $\tilde{\alpha} \in \Omega$,

$$
\epsilon_n(\tilde{\alpha}) = \epsilon_n(\tilde{\alpha}_0) \text{ a.s. } \Rightarrow \tilde{\alpha} = \tilde{\alpha}_0.
$$

**Proof.** Since the covariance matrix $\Sigma_{0\epsilon} = \text{Diag}(\sigma_{01}^2, \ldots, \sigma_{0T}^2)$ 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

$$
\sum_{i=0}^{\infty} \Psi_iX_{n-i} = 0 \text{ a.s. } \Rightarrow \Psi_i = 0, i \geq 0.
$$

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

$$
\epsilon_n(\tilde{\alpha}) = \sum_{i=0}^{\infty} C_i(\tilde{\alpha})X_{n-i} \quad \text{and} \quad \epsilon_n(\tilde{\alpha}_0) = \sum_{i=0}^{\infty} C_i(\tilde{\alpha}_0)X_{n-i}.
$$

Then,

$$
\epsilon_n(\tilde{\alpha}) - \epsilon_n(\tilde{\alpha}_0) = 0 \text{ a.s. } \Rightarrow \sum_{i=0}^{\infty} (C_i(\tilde{\alpha}) - C_i(\tilde{\alpha}_0))X_{n-i} = 0 \text{ a.s.}
$$

$$
\Rightarrow C_i(\tilde{\alpha}) = C_i(\tilde{\alpha}_0), i \geq 0.
$$

The conclusion hold by invoking the identifiability assumption.

**Lemma 7.4** For any $\tilde{\alpha} \in \Omega$ and any $\tilde{\omega}^2 = (\omega_1^2, \ldots, \omega_T^2) > 0$, let

$$
O_{\infty}^{\tilde{\omega}^2}(\tilde{\alpha}) = \sum_{\nu=1}^{T} \omega_{\nu}^{-2} \mathbb{E} \left[ \epsilon_{\nu}^2(\tilde{\alpha}) \right].
$$

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

$$
O_{\infty}^{\tilde{\omega}^2}(\tilde{\alpha}_0) = \sum_{\nu=1}^{T} \frac{\sigma_{0\nu}^2}{\omega_{\nu}^2} < O_{\infty}^{\tilde{\omega}^2}(\tilde{\alpha}).
$$
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 \( \epsilon_\nu(\tilde{\alpha}) - \epsilon_\nu(\tilde{\alpha}_0) \) belongs to the Hilbert space \( \mathcal{H}_N(\nu-1) \). Therefore the linear innovation \( \epsilon_\nu(\tilde{\alpha}_0) \) is not correlated with \( \epsilon_\nu(\tilde{\alpha}) \). Thus

\[
\mathbb{E} \left[ \epsilon_\nu^2(\tilde{\alpha}) \right] = \mathbb{E} \left[ (\epsilon_\nu(\tilde{\alpha}) - \epsilon_\nu(\tilde{\alpha}_0) + \epsilon_\nu(\tilde{\alpha}_0))^2 \right] \\
= \mathbb{E} \left[ \epsilon_\nu^2(\tilde{\alpha}_0) \right] + \mathbb{E} \left[ (\epsilon_\nu(\tilde{\alpha}) - \epsilon_\nu(\tilde{\alpha}_0))^2 \right] + 2\text{Cov} (\epsilon_\nu(\tilde{\alpha}_0), \epsilon_\nu(\tilde{\alpha}) - \epsilon_\nu(\tilde{\alpha}_0)) \\
= \sigma_\nu^2 + \mathbb{E} \left[ (\epsilon_\nu(\tilde{\alpha}) - \epsilon_\nu(\tilde{\alpha}_0))^2 \right].
\]

If \( \tilde{\alpha} \neq \tilde{\alpha}_0 \), 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,

\[
O_{\infty}^2(\tilde{\alpha}_0) = \sum_{\nu=1}^{T} \frac{1}{\omega_\nu^2} \mathbb{E} \left[ \epsilon_\nu^2(\tilde{\alpha}) \right] > \sum_{\nu=1}^{T} \frac{\sigma_\nu^2}{\omega_\nu^2}
\]

if \( \tilde{\alpha} \neq \tilde{\alpha}_0 \).

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

\[
\lim_{N \to \infty} \inf_{\tilde{\alpha} \in V(\tilde{\alpha}^*)} O_N^2(\tilde{\alpha}) > \lim_{N \to \infty} O_N^2(\tilde{\alpha}_0) = \sum_{\nu=1}^{T} \frac{\sigma_\nu^2}{\omega_\nu^2} \text{ a.s.}
\]

where \( O_N^2(\tilde{\alpha}) \) is given by (7.1).

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

\[
S_m(n) = \inf_{\tilde{\alpha} \in V_m(\tilde{\alpha}^*) \cap \Omega} \sum_{\nu=1}^{T} \omega_\nu^{-2} \epsilon_{nT+\nu}^2(\tilde{\alpha}).
\]

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

\[
\inf_{\tilde{\alpha} \in V_m(\tilde{\alpha}^*) \cap \Omega} O_N^2(\tilde{\alpha}) = \inf_{\tilde{\alpha} \in V_m(\tilde{\alpha}^*) \cap \Omega} \{ \frac{1}{N} \sum_{\nu=1}^{T} \omega_\nu^{-2} \sum_{n=0}^{N-1} \epsilon_{nT+\nu}^2(\tilde{\alpha}) \} \geq \frac{1}{N} \sum_{n=0}^{N-1} S_m(n) \underset{N \to \infty}{\longrightarrow} \mathbb{E} [S_m(0)].
\]

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

\[
\lim_{m \to \infty} \mathbb{E} [S_m(0)] = \sum_{\nu=1}^{T} \omega_\nu^{-2} \mathbb{E} [\epsilon_\nu^2(\tilde{\alpha}^*)] = O_{\infty}^2(\tilde{\alpha}^*) > \sum_{\nu=1}^{T} \frac{\sigma_\nu^2}{\omega_\nu^2}.
\]
This implies that
\[
\liminf_{m \to \infty} \liminf_{N \to \infty} \inf_{\vec{\alpha} \in V_m(\vec{\alpha}^*)} O_N^2(\vec{\alpha}) \geq O_\infty^2(\vec{\alpha}^*) > \sum_{\nu=1}^{T} \frac{\sigma_\nu^2}{\omega_\nu^2},
\]
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 \( \vec{\alpha}^* \in \Omega \), \( \vec{\alpha}^* \neq \vec{\alpha}_0 \), there exists a neighbourhood \( V(\vec{\alpha}^*) \) of \( \vec{\alpha}^* \) such that
\[
\liminf_{N \to \infty} \inf_{\vec{\alpha} \in V(\vec{\alpha}^*)} O_N^2(\vec{\alpha}) > \lim_{N \to \infty} O_N^2(\vec{\alpha}_0) = \sum_{\nu=1}^{T} \frac{\sigma_\nu^2}{\omega_\nu^2} \text{ a.s.}
\]

**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
\[
\limsup_{N \to \infty} \sup_{\vec{\alpha} \in \Omega} \left| O_N^2(\vec{\alpha}) - O_N^2(\vec{\alpha}) \right| \leq \limsup_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \frac{1}{\hat{\omega}_\nu^2} - \frac{1}{\omega_\nu^2} \sup_{\vec{\alpha} \in \Omega} \left| e_{nT+\nu}^2(\vec{\alpha}) \right| \\
\leq \varepsilon \sum_{\nu=1}^{T} E \sup_{\vec{\alpha} \in \Omega} \left| e_{\nu}^2(\vec{\alpha}) \right|.
\]
Since the inequality holds for any \( \varepsilon > 0 \), we have
\[
\limsup_{N \to \infty} \sup_{\vec{\alpha} \in \Omega} \left| O_N^2(\vec{\alpha}) - O_N^2(\vec{\alpha}) \right| = 0 \quad \text{a.s.}
\]
In view of Lemma 7.5, the conclusion follows.

**Lemma 7.7** Let \( e_{nT+\nu}(\vec{\alpha}) \) given by (3.5) and \( e_{nT+\nu}(\vec{\alpha}) \) given by (3.6). Almost surely, there exist \( K > 0 \) and \( \rho \in (0, 1) \) such that
\[
\sup_{\vec{\alpha} \in \Omega_3} |e_{nT+\nu}(\vec{\alpha}) - e_{nT+\nu}(\vec{\alpha})| \leq K \rho^n
\]
for all \( \nu \in \{1, 2, \ldots, T\} \) and all \( n \geq 1 \).

**Proof.** For \( q^* = 0 \), the result is obvious. Otherwise, consider and the \( q^*T \times 1 \) vectors \( \vec{e}_n(\vec{\alpha}) \), \( \vec{e}_n(\vec{\alpha}) \) defined by (7.2) and the \( q^*T \times q^*T \) companion matrix \( D \) given by (7.3). We have
\[
\vec{e}_n(\vec{\alpha}) - \vec{e}_n(\vec{\alpha}) = D^{n-p^*} (\vec{e}_{p}(\vec{\alpha}) - \vec{e}_{p}(\vec{\alpha})) \quad \forall n > p^*.
\]
Consider the Jordan decomposition $D = P \Lambda P^{-1}$, where the matrix $\Lambda$ takes the form

$$
\Lambda = \begin{pmatrix}
\Lambda_1 & 0 & \ldots & 0 \\
0 & \Lambda_2 & \ldots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & \Lambda_s
\end{pmatrix}
$$

with

$$
\Lambda_h = \begin{pmatrix}
\lambda_h & 1 & 0 & \ldots & 0 \\
0 & \lambda_h & 1 & \ddots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & \lambda_h & 1 \\
0 & 0 & \ldots & 0 & \lambda_h
\end{pmatrix}.
$$

Using this decomposition, it can be shown that $D^t = P \Lambda^t P^{-1}$ where $\Lambda^t = \text{Diag}(\Lambda^t_1, \ldots, \Lambda^t_s)$ with

$$
\Lambda^t_h = \begin{pmatrix}
\lambda^t_h & (t-1)\lambda^t_{h-1} & \ldots & (t-r_h-1)\lambda^t_{h-r_h+1} \\
0 & \lambda^t_h & \ldots & (t-r_h-2)\lambda^t_{h-r_h+2} \\
\vdots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & \lambda^t_h
\end{pmatrix}.
$$

Since the nonzero eigenvalues $\lambda_i$ 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 $\bar{\alpha}$ such that

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

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

**Lemma 7.8** Let $Q_{N}^{2^2}(\bar{\alpha})$ given by (3.10) and $O_{N}^{2^2}(\bar{\alpha})$ given by (7.1). For any constant $\omega > 0$, we have almost surely

$$
\sup_{\bar{\alpha} \in \Omega_3, \|\bar{\alpha}\| > \omega} \sup_{\bar{\alpha} \in \Omega_3} |Q_{N}^{2^2}(\bar{\alpha}) - Q_{N}^{2^2}(\bar{\alpha})| = O(N^{-1}) \quad \text{as} \quad N \to \infty.
$$

**Proof.** Using Lemma 7.7, we obtain that

$$
N \sup_{\bar{\alpha} \in \Omega_3, \|\bar{\alpha}\| > \omega} |Q_{N}^{2^2}(\bar{\alpha}) - Q_{N}^{2^2}(\bar{\alpha})| \leq T \omega^{-2} \sup_{\bar{\alpha} \in \Omega_3} \left| \sum_{\nu=1}^{T} \sum_{n=0}^{N-1} (\epsilon_{nT+\nu}(\bar{\alpha}) - \epsilon_{nT+\nu}(\bar{\alpha})) \right| \leq K t^{Tq^*} \left( \frac{1}{\delta + 1} \right)^t
$$

30
for $K > 0$ and $\rho \in (0, 1)$. By Lemma 7.2 and the Beppo-Levy theorem, we have
\[
E \sum_{n=0}^{\infty} \rho^n \left(1 + \sup_{\tilde{a} \in \Omega_0} |\epsilon_{nT+\nu}(\tilde{a})|\right) < \infty.
\]
Therefore, the series $\sum_{n=0}^{\infty} \rho^n \left(1 + \sup_{\tilde{a} \in \Omega_0} |\epsilon_{nT+\nu}(\tilde{a})|\right)$ converges almost surely, which completes the proof.

**Lemma 7.9** For any $\tilde{a}^* \in \Omega$, $\tilde{a}^* \neq \tilde{a}_0$, and any $\hat{\omega}^2 > 0$, there exists a neighbourhood $V(\tilde{a}^*)$ of $\tilde{a}^*$ such that $V(\tilde{a}^*) \subset \Omega$ and
\[
\lim_{N \to \infty} \inf_{\tilde{a} \in V(\tilde{a}^*)} Q_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a}) \geq \sum_{\nu=1}^{T} \frac{\sigma_{\nu}^2}{\omega_{\nu}^2} \text{ a.s.}
\]
For any neighbourhood $V(\tilde{a}_0)$ of $\tilde{a}_0$,
\[
\lim_{N \to \infty} \sup_{\tilde{a} \in V(\tilde{a}_0)} Q_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a}) \leq \sum_{\nu=1}^{T} \frac{\sigma_{\nu}^2}{\omega_{\nu}^2} \text{ a.s.}
\]

**Proof.** Since $V(\tilde{a}^*)$ can be chosen as being included in $\Omega_0$ for some $\delta > 0$. We have
\[
\inf_{\tilde{a} \in V(\tilde{a}^*)} Q_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a}) \geq \inf_{\tilde{a} \in V(\tilde{a}^*)} O_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a}) - \sup_{\tilde{a} \in \Omega_0} |O_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a}) - Q_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a})|
\]
and
\[
\inf_{\tilde{a} \in V(\tilde{a}_0)} Q_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a}) \leq O_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a}_0) + \sup_{\tilde{a} \in \Omega_0} |O_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a}) - Q_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a})|
\]
The conclusion follows from Lemmas 7.5 and 7.8.

**Lemma 7.10** Let $\hat{\omega}^2 = (\omega_1^2, \ldots, \omega_T^2)'$ be a vector of strictly positive constants, and let $\hat{\omega}^2$ be a sequence of random vectors such that $\hat{\omega}^2 \to \hat{\omega}^2$ almost surely as $N \to \infty$. For any $\tilde{a}^* \in \Omega$, $\tilde{a}^* \neq \tilde{a}_0$, there exists a neighbourhood $V(\tilde{a}^*)$ of $\tilde{a}^*$ such that $V(\tilde{a}^*) \subset \Omega$ and
\[
\lim_{N \to \infty} \inf_{\tilde{a} \in V(\tilde{a}^*)} Q_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a}) \geq \sum_{\nu=1}^{T} \frac{\sigma_{\nu}^2}{\omega_{\nu}^2} \text{ a.s.}
\]
For any neighbourhood $V(\tilde{a}_0)$ of $\tilde{a}_0$
\[
\lim_{N \to \infty} \sup_{\tilde{a} \in V(\tilde{a}_0)} Q_{\tilde{a}}^{\hat{\omega}^2}(\tilde{a}) \leq \sum_{\nu=1}^{T} \frac{\sigma_{\nu}^2}{\omega_{\nu}^2} \text{ a.s.}
\]

**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 $V(\tilde{a}_0)$ be a neighbourhood of $\tilde{a}_0$. Clearly $\Omega_0$ is covered by $V(\tilde{a}_0)$ and the union of all the $V(\tilde{a}^*)$, $\tilde{a}^* \in \Omega_0 - V(\tilde{a}_0)$ where $V(\tilde{a}^*)$ is defined in Lemma 7.9. By the compactness
of $\Omega_\delta$, there exist $\bar{\alpha}_1, \ldots, \bar{\alpha}_k$ such that $\Omega_\delta$ is covered by $V(\bar{\alpha}_0), V(\bar{\alpha}_1), \ldots, V(\bar{\alpha}_k)$. Lemma 7.9 shows that, almost surely,
\[
\inf_{\tilde{\alpha} \in \Omega_\delta} Q^2_N(\tilde{\alpha}) = \min_{i=0,1,\ldots,k} \inf_{\tilde{\alpha} \in V(\bar{\alpha}_i) \cap \Omega_\delta} Q^2_N(\tilde{\alpha}) = \inf_{\tilde{\alpha} \in V(\bar{\alpha}_0) \cap \Omega_\delta} Q^2_N(\tilde{\alpha})
\]
for $N$ large enough. Therefore, $\hat{\omega}^{\tilde{\omega}^2}$ almost surely belongs to $V(\bar{\alpha}_0)$ for $N$ large enough. Since $V(\bar{\alpha}_0)$ can be arbitrarily small, $\hat{\omega}^{\tilde{\omega}^2} \rightarrow \bar{\alpha}_0$ almost surely as $N \rightarrow \infty$. The OLS and GLS estimators correspond to WLS estimators, for particular choices of the weights vector $\bar{\omega}^2$. Thus the first three consistencies of Theorem 3.1 are shown. Using Lemma 7.10, the previous arguments show that $\hat{\omega}^{\tilde{\omega}^2} \rightarrow \bar{\alpha}_0$ almost surely whenever $\hat{\omega}^2 \rightarrow \bar{\sigma}^2_0$ as $N \rightarrow \infty$. It follows that the QLS estimator is also consistent.

Theorem 3.2 can be established using the following lemmas.

**Lemma 7.11** For any $\bar{\alpha}^* \in \Omega$, and any $m \in \{1, 2, \ldots, (p+q)T\}$, there exist absolutely summable sequences $(C_i(\bar{\alpha}))_{i \in \mathbb{N}}$ and $(C_{m,i}(\bar{\alpha}))_{i \in \mathbb{N}}$ such that
\[
\epsilon_n(\bar{\alpha}) = \sum_{i=0}^\infty C_i(\bar{\alpha})X_{n-i} \quad \text{and} \quad \frac{\partial \epsilon_n(\bar{\alpha})}{\partial \alpha_m} = \sum_{i=0}^\infty C_{m,i}(\bar{\alpha})X_{n-i}. \tag{7.7}
\]
Moreover, there exist $\rho \in (0,1)$ and $K \in [0, \infty)$ such that, for all $i \geq 0$,
\[
\sup_{\tilde{\alpha} \in \Omega_\delta} \|C_i(\tilde{\alpha})\| \leq K \rho^i \quad \text{and} \quad \sup_{\tilde{\alpha} \in \Omega_\delta} \|C_{m,i}(\tilde{\alpha})\| \leq K \rho^i. \tag{7.8}
\]

**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, $\epsilon_n(\bar{\alpha}) = \sum_{i=0}^\infty \hat{C}_i(\bar{\alpha})\hat{X}_{n-i}$, where $\hat{X}_n$ and $\epsilon_n(\bar{\alpha})$ are given by (7.2) and $\hat{C}_i(\bar{\alpha}) = \sum_{j=0}^\infty B^{i-j}A_j$ as in (7.4). Then,
\[
\frac{\partial \epsilon_n(\bar{\alpha})}{\partial \alpha_m} = \sum_{i=0}^\infty \frac{\partial \hat{C}_i(\bar{\alpha})}{\partial \alpha_m} \hat{X}_{n-i},
\]
and since,
\[
\epsilon_n(\bar{\alpha}) = \sum_{i=0}^\infty L_1^i \hat{C}_i(\bar{\alpha})L_1X_{n-i} = \sum_{i=0}^\infty C_i(\bar{\alpha})X_{n-i},
\]
we have
\[
\frac{\partial \epsilon_n(\bar{\alpha})}{\partial \alpha_m} = \sum_{i=0}^\infty L_1^i \frac{\partial \hat{C}_i(\bar{\alpha})}{\partial \alpha_m} L_1X_{n-i} = \sum_{i=0}^\infty C_{m,i}(\bar{\alpha})X_{n-i},
\]
where $C_i(\bar{\alpha}) = L_1^i \hat{C}_i(\bar{\alpha})L_1$ and $C_{i,m}(\bar{\alpha}) = L_1^i \frac{\partial \hat{C}_i(\bar{\alpha})}{\partial \alpha_m} L_1$ with $L_1 = (I_{T \times T}, 0_{T \times T}, \ldots, 0_{T \times T})^T$. 

32
Also, from Lemma 7.1 there exists a positive constant $K_1$ independent of $\vec{\alpha}$ such that
\[
\sup_{\vec{\alpha} \in \Omega} \| C_i(\vec{\alpha}) \| \leq K_1 i T q^* \left( \frac{1}{1 + \delta} \right)^i 
\leq K_1 i T q^* \left( \frac{1}{1 + (\delta/2)} \right)^i \left( \frac{1 + \delta}{1 + \delta} \right)^i 
\leq K \rho^i,
\]
where $\rho = \frac{1 + (\delta/2)}{1 + \delta} \in (0, 1)$ and $K = K_1 \max_i \left\{ i T q^* \left( \frac{1}{1 + (\delta/2)} \right)^i \right\} \leq K_1 e^{-T q^* \left[ \frac{q^* T}{\log(1 + (\delta/2))} \right]^T q^*}.$

Now, consider the second inequality in (7.8). We have
\[
\partial \tilde{C}_i(\vec{\alpha}) / \partial \alpha_m = \min_{j=0}^{\min(i, p^*)} \left\{ \partial D_{i-j} / \partial \alpha_m A_j + D_{i-j} / \partial \alpha_m \partial A_j / \partial \alpha_m \right\}
\]
and
\[
\sup_{\vec{\alpha} \in \Omega} \left\| \partial \tilde{C}_i(\vec{\alpha}) / \partial \alpha_m \right\| \leq \sum_{j=0}^{\min(1, p^*)} \left\{ \| \partial D_{i-j} / \partial \alpha_m \| \| A_j \| + \| D_{i-j} / \partial \alpha_m \| \partial A_j / \partial \alpha_m \| \right\}. \quad (7.9)
\]
Using the Jordan decomposition $D = P \Lambda P^{-1}$, we have
\[
\partial D^t / \partial \alpha_m = \partial P / \partial \alpha_m \Lambda^t P^{-1} + P \partial \Lambda^t / \partial \alpha_m P^{-1} + P \Lambda^t \partial P^{-1} / \partial \alpha_m.
\]
This implies that
\[
\left\| \partial D^t / \partial \alpha_m \right\| \leq \left\| \partial P / \partial \alpha_m \| \| \Lambda^t \| \| P^{-1} \| \right\| + \left\| \partial \Lambda^t / \partial \alpha_m \| \| P^{-1} \| \right\| + \left\| \| \partial P^{-1} / \partial \alpha_m \| \| \Lambda^t \| \right\| \| P^{-1} \|
\]
with $\partial \Lambda^t / \partial \alpha_m = \text{Diag} \left( \partial \Lambda^t_1 / \partial \alpha_m, \ldots, \partial \Lambda^t_s / \partial \alpha_m \right)$ where
\[
\frac{\partial \Lambda^t_h}{\partial \alpha_m} = \frac{\partial \Lambda_h}{\partial \alpha_m} = \begin{pmatrix}
    n \lambda_h^{t-1} & (t-1)\lambda_h^{t-2} & \cdots & (t-r_h-1)\lambda_h^{t-r_h} \\
    0 & n \lambda_h^{t-1} & \cdots & (t-r_h-2)\lambda_h^{t-r_h+1} \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & t \lambda_h^{t-1}
\end{pmatrix}.
\]
Since $\max_h |\lambda_h| \leq \frac{1}{1 + \delta}$, there exists a positive constant $K_2$ independent of $\vec{\alpha}$ such that
\[
\left\| \partial D^t / \partial \alpha_m \right\| \leq K_2 i T q^* \left( \frac{1}{1 + \delta} \right)^i. \quad (7.10)
\]
From (7.6), (7.9) and (7.10), we obtain that
\[
\sup_{\vec{\alpha} \in \Omega} \| C_{m,i}(\vec{\alpha}) \| \leq K_3 \partial q^* \left( \frac{1}{1 + \delta} \right)^i \leq K_3 i T q^* \left( \frac{1}{1 + (\delta/2)} \right)^i \left( \frac{1 + \delta}{1 + \delta} \right)^i \leq K \rho^i
\]

where ρ = \frac{1 + (δ/2)}{1 + δ} \in (0, 1) and K = K_{3\max} \left\{ i^{T}q^* \left( \frac{1}{1 + (δ/2)} \right) \right\} \leq K_{3}e^{-Tq^* \left[ \frac{q^* T}{\log(1 + (δ/2))} \right]^{T}q^*}.

Which completes the proof of Lemma 7.11.

**Lemma 7.12** Let the assumptions of Theorem 2 be satisfied. For all ⃗α ∈ Ωδ and all ⃗ω^2 > 0, the matrix

\[ I(\bar{α}, \omega^2) = \frac{1}{4} \lim_{N \to \infty} \text{Var} \left( \sqrt{N} \frac{\partial O^2_N (\bar{α})}{\partial \bar{α}} \right) \]

exists.

**Proof.** Let the diagonal matrix Σω = \text{Diag} (ω^2) and

\[ Y_n = \frac{\partial e_n'(\bar{α})}{\partial \bar{α}} \Sigma^{-1}_ω e_n(\bar{α}) = \sum_{\nu=1}^{T} \omega^{-2}_{\nu} \epsilon_{nT+\nu}(\bar{α}) \frac{\partial \epsilon_{nT+\nu}(\bar{α})}{\partial \bar{α}}. \]

The process \( \{ Y_n \} \) is strictly stationary and ergodic. Moreover,

\[ \frac{1}{4} \text{Var} \left( \sqrt{N} \frac{\partial O^2_N (\bar{α})}{\partial \bar{α}} \right) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{n'=0}^{N-1} \text{Cov} (Y_n, Y_{n'}). \]

Denote by

\[ I_N(l, m) = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{n'=0}^{N-1} \text{Cov} (Y_n(l), Y_{n'}(m)) = \frac{1}{N} \sum_{k=1-N}^{N-1} (N - |k|)c(k) \]

where

\[ c(k) = \text{Cov} (Y_n(l), Y_{n-k}(m)) \]

with

\[ Y_n(l) = \frac{\partial e_n'(\bar{α})}{\partial \alpha_l} \Sigma^{-1}_ω e_n(\bar{α}) \text{ and } Y_{n-k}(m) = \frac{\partial e_{n-k}'(\bar{α})}{\partial \alpha_m} \Sigma^{-1}_ω e_{n-k}(\bar{α}). \]

Suppose that \( k \geq 0 \). From Lemma 7.11,

\[ |c(k)| \leq \sum_{i,j,i',j'=0}^{\infty} \sum_{i,j,i',j'=0}^{\infty} \sum_{i,j,i',j'=0}^{\infty} \text{Cov} (X'_{n-j}C'_{i,j}(\bar{α})\Sigma^{-1}_ω C_i(\bar{α})X_{n-i}, X'_{n-k-j'}C'_{m,j'}(\bar{α})\Sigma^{-1}_ω C_{i'}(\bar{α})X_{n-k-i'}) \]

\[ \leq \sum_{i,j,i',j'=0}^{\infty} \left| \text{Cov} (X'_{n-j}C'_{i,j}(\bar{α})\Sigma^{-1}_ω C_i(\bar{α})X_{n-i}, X'_{n-k-j'}C'_{m,j'}(\bar{α})\Sigma^{-1}_ω C_{i'}(\bar{α})X_{n-k-i'}) \right| \]

\[ \leq g_1 + g_2 + g_3 + g_4 + g_5 \]

34
where
\[
g_1 = \sum_{i=k/2}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \left| \operatorname{Cov} \left( X'_{n-j} C'_{i, j, l} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_i (\tilde{\alpha}) X_{n-i}, X'_{n-k-j} C'_{m, j'} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_{i'} (\tilde{\alpha}) X_{n-k-i'} \right) \right|, 
\]
\[
g_2 = \sum_{j' > k/2}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \left| \operatorname{Cov} \left( X'_{n-j} C'_{i, j, l} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_i (\tilde{\alpha}) X_{n-i}, X'_{n-k-j} C'_{m, j'} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_{i'} (\tilde{\alpha}) X_{n-k-i'} \right) \right|, 
\]
\[
g_3 = \sum_{j > k/2}^{\infty} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \left| \operatorname{Cov} \left( X'_{n-j} C'_{i, j, l} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_i (\tilde{\alpha}) X_{n-i}, X'_{n-k-j} C'_{m, j'} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_{i'} (\tilde{\alpha}) X_{n-k-i'} \right) \right|, 
\]
\[
g_4 = \sum_{i > k/2}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \left| \operatorname{Cov} \left( X'_{n-j} C'_{i, j, l} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_i (\tilde{\alpha}) X_{n-i}, X'_{n-k-j} C'_{m, j'} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_{i'} (\tilde{\alpha}) X_{n-k-i'} \right) \right|, 
\]
\[
g_5 = \sum_{i=k/2}^{\infty} \sum_{j=k/2}^{\infty} \sum_{i'=k/2}^{\infty} \sum_{j'=k/2}^{\infty} \left| \operatorname{Cov} \left( X'_{n-j} C'_{i, j, l} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_i (\tilde{\alpha}) X_{n-i}, X'_{n-k-j} C'_{m, j'} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_{i'} (\tilde{\alpha}) X_{n-k-i'} \right) \right|. 
\]

By the Cauchy-Schwarz inequality, we obtain
\[
g_1 \leq \sum_{i=k/2}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \left\{ \mathbb{E} \left[ \left( X'_{n-j} C'_{i, j, l} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_i (\tilde{\alpha}) X_{n-i} \right)^2 \right] \right\} \mathbb{E} \left[ \left( X'_{n-k-j} C'_{m, j', l} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_{i'} (\tilde{\alpha}) X_{n-k-i'} \right)^2 \right] \right\}^{1/2} \]
\[
\leq \sum_{i=k/2}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=0}^{\infty} \sum_{j'=0}^{\infty} \left\{ \mathbb{E} \left[ \left\| X_{n-j} \right\|^2 \left\| C'_{i, j, l} (\tilde{\alpha}) \right\|^2 \left\| \Sigma_{\omega}^{-1} \right\|^2 \left\| C_i (\tilde{\alpha}) \right\|^2 \left\| X_{n-i} \right\|^2 \right] \right\} \mathbb{E} \left[ \left\| X_{n-k-j} \right\|^2 \left\| C'_{m, j', l} (\tilde{\alpha}) \right\|^2 \left\| \Sigma_{\omega}^{-1} \right\|^2 \left\| C_{i'} (\tilde{\alpha}) \right\|^2 \left\| X_{n-k-i'} \right\|^2 \right] \right\}^{1/2} \]
\[
\leq \mathbb{E} \left[ \left\| X_n \right\|^2 \right] \left\| \Sigma_{\omega}^{-1} \right\|^2 \sum_{i=k/2}^{\infty} \left\| C_i (\tilde{\alpha}) \right\| \sum_{j=0}^{\infty} \left\| C_{i, j, l} (\tilde{\alpha}) \right\| \sum_{i'=0}^{\infty} \left\| C_{i'} (\tilde{\alpha}) \right\| \sum_{j'=0}^{\infty} \left\| C_{m, j'} (\tilde{\alpha}) \right\| \right\}^{1/2} \]
\[
\leq M_1 \sum_{i=k/2}^{\infty} \left\| C_i (\tilde{\alpha}) \right\| \leq \Delta_1 \rho^{k/2} 
\]
for some positive constants $M_1$ and $\Delta_1$. Using the same arguments we obtain that $g_i$ ($i = 2, 3, 4$) is bounded by $\Delta_1 \rho^{k/2}$.

In the other hand, Davydov inequality (Davydov, 1968) implies that, there exists a positive constant $M_5$ such that
\[
g_5 \leq \sum_{i=k/2}^{\infty} \sum_{j=k/2}^{\infty} \sum_{i'=k/2}^{\infty} \sum_{j'=k/2}^{\infty} M_5 \left\| X'_{n-j} C'_{i, j, l} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_i (\tilde{\alpha}) X_{n-i} \right\|_{2+r} \left\| X'_{n-k-j} C'_{m, j', l} (\tilde{\alpha}) \Sigma_{\omega}^{-1} C_{i'} (\tilde{\alpha}) X_{n-k-i'} \right\|_{2+r} \left( \alpha_\chi (\min \{ k + i' - j, k + j' - j, k + i' - i, k + j' - i \}) \right)^{2+r} \]
\[
\leq M \left[ \alpha_{\chi} (k/2) \right]^{2+r} \]
for some positive constant $M$. Therefore, for $k \geq 0$, there exist positive constants $M$ and $\Delta$ such that
\[
|c(k)| \leq \Delta \rho^{k/2} + M \left[ \alpha_{\chi} (k/2) \right]^{2+r}. 
\]
A similar inequality hold for \( k < 0 \). Therefore
\[
\sum_{k=\infty}^{\infty} |c(k)| < \infty.
\]
Then, the dominated convergence theorem gives
\[
I_N(l, m) = \frac{1}{N} \sum_{k=1}^{N-1} (N - |k|)c(k) \xrightarrow{N \to \infty} \sum_{k=\infty}^{\infty} |c(k)|.
\]
This implies that
\[
I(\vec{\alpha}, \vec{\omega}) = T \sum_{\nu=1}^{T} \omega_{\nu}^{-2} \omega_{\nu'}^{-2} \sum_{k=\infty}^{\infty} \mathbb{E} \left[ \left( \epsilon_{nT+\nu}(\vec{\alpha}) \frac{\partial \epsilon_{nT+\nu}(\vec{\alpha})}{\partial \vec{\alpha}} \right) \left( \epsilon_{(n-k)T+\nu'}(\vec{\alpha}) \frac{\partial \epsilon_{(n-k)T+\nu'}(\vec{\alpha})}{\partial \vec{\alpha}} \right) \right]
\]
exists. Which completes the proof of Lemma 7.12.

**Lemma 7.13** Under the assumptions of Theorem 2, the random vector
\[
\sqrt{N} \left( \frac{\partial Q_{N}^{2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4I(\vec{\alpha}_0, \vec{\omega}^2)) \quad \text{as } N \to \infty.
\]

**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
\[
\left( \frac{\partial Q_{N}^{2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \text{ and } \left( \frac{\partial O_{N}^{2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0}
\]
have the same asymptotic distribution. Therefore, it remains to show that
\[
\sqrt{N} \left( \frac{\partial Q_{N}^{2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4I(\vec{\alpha}_0, \vec{\omega}^2)) \quad \text{as } N \to \infty.
\]
Moreover, note that
\[
\sqrt{N} \left( \frac{\partial Q_{N}^{2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} = 2 \sqrt{N} \sum_{n=0}^{N-1} Y_n = 2 \sqrt{N} \sum_{n=0}^{N-1} \frac{\partial \epsilon_n(\vec{\alpha})}{\partial \vec{\alpha}} \Sigma^\omega \epsilon_n(\vec{\alpha})
\]
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 \( H_X(nT + \nu - 1) \), 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
\[
\mathbb{E} \left[ \sqrt{N} \left( \frac{\partial Q_{N}^{2}(\vec{\alpha})}{\partial \vec{\alpha}} \right)_{\vec{\alpha}=\vec{\alpha}_0} \right] = 0.
\]
\[36\]
Now, for any positive integer \( r \), we have

\[
\sqrt{N} \left( \partial \Omega^2_N (\tilde{\alpha}) \right)_{\tilde{\alpha} = \tilde{\alpha}_0} = \frac{2}{\sqrt{N}} \sum_{n=0}^{N-1} (Y_{n,r} - \mathbb{E} [Y_{n,r}]) + \frac{2}{\sqrt{N}} \sum_{n=0}^{N-1} (Z_{n,r} - \mathbb{E} [Z_{n,r}])
\]

where

\[
Z_{n,r} = U_{n,r} + V_{n,r} \quad \quad U_{n,r} = \sum_{i=r+1}^{\infty} \sum_{j=0}^{\infty} (I_d \times X'_{n-j}) E_j(\tilde{\alpha}_0) \Sigma^{-1} C_i(\tilde{\alpha}_0) X_{n-i} \quad \quad V_{n,r} = \sum_{i=0}^{r} \sum_{j=r+1}^{\infty} (I_d \times X'_{n-j}) E_j(\tilde{\alpha}_0) \Sigma^{-1} C_i(\tilde{\alpha}_0) X_{n-i}
\]

and

\[
Y_{n,r} = \sum_{i=0}^{r} \sum_{j=0}^{r} (I_d \times X'_{n-j}) E_j(\tilde{\alpha}_0) \Sigma^{-1} C_i(\tilde{\alpha}_0) X_{n-i}.
\]

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

\[
\frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} (Y_{n,r} - \mathbb{E} [Y_{n,r}]) \xrightarrow{N \to \infty} \mathcal{N}(0, I_r).
\]

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

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

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

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

where

\[
c_r(h) = \text{Cov} (U_{n,r}(m), U_{n+h,r}(m)) = \sum_{i=r+1}^{\infty} \sum_{j=0}^{\infty} \sum_{i'=r+1}^{\infty} \sum_{j'=0}^{\infty} \text{Cov} (X'_{n-j} C_{m,j}(\tilde{\alpha}_0) \Sigma^{-1} C_i(\tilde{\alpha}_0) X_{n-i}, X'_{n+h-i'} C_{m',j'}(\tilde{\alpha}_0) \Sigma^{-1} C_{i'}(\tilde{\alpha}_0) X_{n+h-i'}).\]

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

\[
|c_r(h)| \leq M_1 \sum_{i=r+1}^{\infty} \|C_i(\tilde{\alpha}_0)\| \sum_{j=0}^{\infty} \|C'_{m,j}(\tilde{\alpha}_0)\| \sum_{i'=r+1}^{\infty} \|C_{i'}(\tilde{\alpha}_0)\| \sum_{j'=0}^{\infty} \|C'_{m',j'}(\tilde{\alpha}_0)\| \leq A_1 \rho^r
\]
for some positive constants $M_1$ and $\Delta_1$. For $r < [h]/2$, using the Cauchy-Schwarz inequality and Davydov inequality, we get that,

\[
|c_r(h)| \leq \sum_{i=r+1}^{\infty} \sum_{j=0}^{\infty} \sum_{j'=0}^{\infty} \left| \text{Cov} \left( X_{n-i} C_{l} (\alpha_{0}) X_{n} \right) \right|
\]

\[
+ \sum_{i=r+1}^{\infty} \sum_{j=0}^{\infty} \sum_{j'=0}^{\infty} \left| \text{Cov} \left( X_{n-i} C_{l} (\alpha_{0}) X_{n} \right) \right|
\]

\[
+ \sum_{i=r+1}^{\infty} \sum_{j=0}^{\infty} \sum_{j'=0}^{\infty} \left| \text{Cov} \left( X_{n-i} C_{l} (\alpha_{0}) X_{n} \right) \right|
\]

\[
\leq \Delta_2 r^{\frac{c}{2}} \left( \left[ \frac{|h|}{2} \right] \right)^{\frac{c}{2}} + M_2 r^{h/2}
\]

for some positive constants $M_2$ and $\Delta_2$. The same inequality holds for $h < 0$. Therefore, there exists a constant $\Delta$ such that

\[
\sum_{h=-\infty}^{\infty} |c_r(h)| = \sum_{|h| \leq 2r+1} |c_r(h)| + \sum_{|h| \geq 2(r+1)} |c_r(h)| \leq \Delta r^c + \Delta r^{\frac{c}{2}} \sum_{k} [\alpha_k(k)]^{\frac{c}{2}} \rightarrow 0.
\]

This implies that

\[
\sup_{N} \text{Var} \left( \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} U_{n,r}(m) \right) \rightarrow 0. \tag{7.12}
\]

In a similar way, it can be shown that

\[
\sup_{N} \text{Var} \left( \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} V_{n,r}(m) \right) \rightarrow 0. \tag{7.13}
\]

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

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

\[
J(\bar{\alpha}, \bar{\omega}^2) = \frac{1}{2} \lim_{N \rightarrow \infty} \left[ \frac{\partial^2 Q_N^2 (\bar{\alpha})}{\partial \bar{\alpha} \partial \bar{\omega}^2} \right]_{\bar{\alpha} = \bar{\alpha}_0}
\]

exists and is strictly positive definite.

**Proof.** We can show that almost surely \( \left\| \frac{\partial^2 \epsilon_n(\bar{\alpha})}{\partial \alpha_i \partial \alpha_m} - \frac{\partial^2 \epsilon_n(\bar{\alpha})}{\partial \alpha_l \partial \alpha_m} \right\| \rightarrow 0 \). Therefore, \( \frac{\partial^2 Q_N^2 (\bar{\alpha})}{\partial \alpha_i \partial \alpha_m} \) and \( \frac{\partial^2 Q_N^2 (\bar{\alpha})}{\partial \alpha_l \partial \alpha_m} \), almost surely the same limit. As in Lemma 7.11, it can be shown that there exists an absolutely summable sequence \((C_{l,m,i}(\bar{\alpha}))_{i \in \mathbb{N}}\) such that

\[
\frac{\partial^2 \epsilon_n(\bar{\alpha})}{\partial \alpha_i \partial \alpha_m} = \sum_{i=0}^{\infty} C_{l,m,i}(\bar{\alpha}) X_{n-i}.
\]

38
This implies that \( \frac{\partial^2 \epsilon_n(\tilde{\alpha})}{\partial \alpha_i \partial \alpha_m} \) belongs to \( L^2 \). In the other hand, we have
\[
\left[ \left( \frac{\partial^2 Q_N^{\omega^2}(\tilde{\alpha})}{\partial \tilde{\alpha} \partial \tilde{\alpha}'} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} \right] = 2 \sum_{\nu=1}^{T} \omega^2_{\nu} - \frac{1}{N} \sum_{\nu=1}^{N} \epsilon_{nT+\nu}(\tilde{\alpha}_0) \left( \frac{\partial^2 \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha} \partial \tilde{\alpha}'} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} + 2 \sum_{\nu=1}^{T} \omega^2_{\nu} - \frac{1}{N} \sum_{\nu=1}^{N} \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} \rightarrow_{N \to \infty} 2 \sum_{\nu=1}^{T} \omega^2_{\nu} \mathbb{E} \left[ \epsilon_{nT+\nu}(\tilde{\alpha}_0) \left( \frac{\partial^2 \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha} \partial \tilde{\alpha}'} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} \right] + 2 \sum_{\nu=1}^{T} \omega^2_{\nu} \mathbb{E} \left[ \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} \right].
\]

Since \( \left( \frac{\partial^2 \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha} \partial \tilde{\alpha}'} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} \) belongs to the Hilbert space \( \mathcal{H}_X(nT+\nu-1) \), the random variables \( \epsilon_{nT+\nu}(\tilde{\alpha}_0) \) and \( \left( \frac{\partial^2 \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha} \partial \tilde{\alpha}'} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} \) are orthogonal and the first term in the limit is zero. Therefore,
\[
J(\tilde{\alpha}_0, \omega^2) = \sum_{\nu=1}^{T} \omega^2_{\nu} \mathbb{E} \left[ \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} \left( \frac{\partial \epsilon_{nT+\nu}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha}=\tilde{\alpha}_0} \right].
\]

Again, invoking the fact that \( \omega^2_1, ..., \omega^2_T \) are all strictly positive and the identifiability Assumption (A1), we can conclude that \( J(\tilde{\alpha}_0, \omega^2) \) is a strictly positive definite matrix.

**Proof of Theorem 3.2.** Consider the first-order Taylor expansion around \( \tilde{\alpha}_0 \), we obtain
\[
0 = \sqrt{N} \left( \frac{\partial Q_N^{\omega^2}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha} = \hat{\alpha}_{WLS}} = \sqrt{N} \left( \frac{\partial Q_N^{\omega^2}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha} = \tilde{\alpha}_0} + \left[ \left( \frac{\partial^2 Q_N^{\omega^2}(\tilde{\alpha})}{\partial \alpha_i \partial \alpha_m} \right)_{\tilde{\alpha} = \tilde{\alpha}_{N,l,m}} \right] \sqrt{N} (\hat{\alpha}_{WLS} - \tilde{\alpha}_0)
\]

where \( \tilde{\alpha}_{N,l,m} \) is between \( \hat{\alpha}_{WLS} \) and \( \tilde{\alpha}_0 \). Using again a Taylor expansion, we obtain
\[
\left| \left( \frac{\partial^2 Q_N^{\omega^2}(\tilde{\alpha})}{\partial \alpha_i \partial \alpha_m} \right)_{\tilde{\alpha} = \tilde{\alpha}_{N,l,m}} - \left( \frac{\partial^2 Q_N^{\omega^2}(\tilde{\alpha})}{\partial \alpha_i \partial \alpha_m} \right)_{\tilde{\alpha} = \tilde{\alpha}_0} \right| \leq \sup_{\tilde{\alpha} \in \Omega_t} \left| \frac{\partial}{\partial \tilde{\alpha}} \left( \frac{\partial^2 Q_N^{\omega^2}(\tilde{\alpha})}{\partial \alpha_i \partial \alpha_m} \right) \right| ||\tilde{\alpha}_{N,l,m} - \tilde{\alpha}_0|| \rightarrow_{N \to \infty} 0 \text{ a.s..}
\]

This implies that, as \( N \to \infty \)
\[
\sqrt{N} (\hat{\alpha}_{WLS} - \tilde{\alpha}_0) = - \left[ \left( \frac{\partial^2 Q_N^{\omega^2}(\tilde{\alpha})}{\partial \alpha_i \partial \alpha_m} \right)_{\tilde{\alpha} = \tilde{\alpha}_0} \right]^{-1} \sqrt{N} \left( \frac{\partial Q_N^{\omega^2}(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha} = \tilde{\alpha}_0} + o_p(1).
\]

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 $O^2_N(\tilde{\alpha}) - Q^2_N(\tilde{\alpha})$ is replaced by its first and second order derivatives, and (ii) by noting that
\[
\sqrt{N} \left( \frac{\partial O^2_N(\tilde{\alpha})}{\partial \tilde{\alpha}} - \frac{\partial Q^2_N(\tilde{\alpha})}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha} = \tilde{\alpha}_0} = \sum_{\nu=1}^{T} \left( \frac{1}{\hat{\sigma}_\nu^2} - \frac{1}{\sigma_\nu^2} \right) \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} \epsilon_{nT+\nu}(\tilde{\alpha}_0) \left( \frac{\partial \epsilon_{nT+\nu}}{\partial \tilde{\alpha}} \right)_{\tilde{\alpha} = \tilde{\alpha}_0} = o_P(1).
\]

Acknowledgements
The authors are grateful to the Associate Editor and three referees whose comments led to a greatly improved presentation. This work was partially supported by grants to the second author from the Natural Science and Engineering Research Council of Canada and the Network of Centres of Excellence on The Mathematics of Information Technology and Complex Systems (MITACS).

References


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.


