Identifiability and Whittle Estimation of Periodic ARMA Models

Alessandro José Queiroz Sarnaglia, Valdério Anselmo Reisen, Pascal Bondon, Carlo Corrêa Solci, Márton Ispány

Identifiability and Whittle Estimation of Periodic ARMA Models

Alessandro J. Q. Sarnaglia, Valdério A. Reisen, Pascal Bondon, Carlo C. Solci, and Márton Ispány

Abstract The Periodic Autoregressive Moving Average (PARMA) models are generally assumed to be identifiable. However, this assumption becomes not true if some model conditions are not specified. This paper fills this gap by providing verifiable conditions for the identifiability of PARMA models and, in addition, the Whittle likelihood estimator (WLE) is proposed to estimate the model parameters. This estimator is strongly consistent and asymptotically normal. The Monte Carlo simulation investigation shows that the WLE is a very attractive alternative to the Gaussian maximum likelihood estimator (MLE) for large data sets. Although the estimators have similar accuracy, the computational cost of the MLE is much higher. The methods are applied to fit a PARMA model to the sulfur dioxide (SO_2) daily average pollutant concentrations measured in the city of Vitória (ES), Brazil.

Key words: Periodic stationarity; PARMA models; Identifiability; Whittle estimation; Sulfur Dioxide.

Alessandro José Queiroz Sarnaglia
DEST and LECON, Federal University of Espírito Santo, Brazil and PGEST, Federal University of Minas Gerais, Brazil

Valdério Anselmo Reisen
PPGEA and PPEco, Federal University of Espírito Santo and PGEST, Federal University of Minas Gerais, Brazil and PGMAT, Federal University of Bahia and Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France

Pascal Bondon
Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France

Carlo Corrêa Solci
PPGEA, Federal University of Espírito Santo, Brazil

Márton Ispány
University of Debrecen, Hungary

1 Introduction

Seasonal phenomena are frequently observed in many fields such as hydrology, climatology, air pollution, radio astronomy, econometrics, communications, signal processing, among others. A standard approach in the literature is to fit a stationary seasonal model after removing any trend. As pointed out by Tiao and Grupe (1980), standard time series tools may indicate stationary models even if the true covariance structure has a periodic (or cyclic) nonstationary behavior. The model mispecification usually deteriorates the forecast performance even if the standard residual diagnostic checking does not reveal any anomaly.

Processes with periodically varying covariances are introduced in the seminal paper of Gladyshev (1961) and are denominated periodically correlated (PC), periodically stationary or cyclostationary. The occurrence of periodic correlation is corroborated by real applications in many areas. For example, Gardner and Franks (1975) investigate cyclostationarity in electrical engineering and Bloomfield et al. (1994) study stratospheric ozone data. For recent reviews on PC processes, see e.g. Gardner et al. (2006) and Hurd and Miamee (2007).

The simplest way to build models for PC processes is to allow the parameters of stationary models to vary periodically with time. In this context, the periodic autoregressive model emerges as an extension of the well-known autoregressive framework. Parameter estimation of a PAR model is already well documented in the literature, see e.g. Sarnaglia et al. (2010) and references therein. However, some data sets require large periodic autoregressive orders to provide an adequate fit. Thus, a more parsimonious model can be built by jointly considering periodic AR and MA coefficients, which leads naturally to the PARMA model. However, this model has not yet been used widely in real applications, perhaps, due to the difficulty and high computational cost of the implementation of standard estimation methods.

The exact Gaussian PARMA likelihood is derived by Li and Hui (1988), and the method requires the Choleski decomposition of a matrix whose dimension is the number of data. Using the innovation algorithm, Lund and Basawa (2000) have proposed a recursive method to calculate this Choleski decomposition which does not require any matrix inversion.

It is well known that the PARMA model has the vector ARMA (VARMA) representation, see e.g. Vecchia (1985). However the VARMA model needs to satisfy the conditions given by Dunsmuir and Hannan (1976) to be identifiable. These conditions are tacitly assumed in the literature on PARMA models, see e.g. Basawa and Lund (2001). Here, we show that the identifiability conditions of Dunsmuir and Hannan (1976) do not transpose trivially to the PARMA model, and one contribution of this paper is to provide identifiability conditions for the PARMA model.

To our knowledge, only time domain estimation methods have been proposed for PARMA models in the literature. In the frequency domain, the

well-known Whittle approximation can be used to circumvent the inversion of the covariance matrix, see e.g. Whittle (1953), Dunsmuir and Hannan (1976), Deistler et al. (1978) and Fox and Taqqu (1986). Here, we propose to apply the Whittle's methodology for estimating the parameters of a PARMA model and we establish the strong consistency and the asymptotic normality of the WLE.

The rest of the paper is organized as follows. In Section 2, PC processes and PARMA models are described and the identifiability results are presented. In Section 3, the WLE of a PARMA model is introduced and its asymptotics properties are derived. In Section 4, we compare via Monte Carlo simulations the MLE and the WLE. The two estimation methods are applied to fit a PARMA model to air pollution data in Section 5. Proofs are deferred to Section 6.

2 Identifiability of the PARMA model

Let $\mathbb{Z}$ be the set of integer numbers and (X_t), $t \in \mathbb{Z}$, be a real-valued stochastic process satisfying $\mathrm{E}(X_t^2) < \infty$ for all $t \in \mathbb{Z}$. Denote the mean and autocovariance functions of (X_t) by $\mu_t = \mathrm{E}(X_t)$ and $\gamma_t(\tau) = \mathrm{Cov}(X_t, X_{t-\tau})$, respectively. (X_t) is said to be PC with period $\mathcal{S} > 0$ if, for every pair $(t, \tau) \in \mathbb{Z}^2$,

$$\mu_{t+\mathcal{S}} = \mu_t \quad \text{and} \quad \gamma_{t+\mathcal{S}}(\tau) = \gamma_t(\tau), \tag{2.1}$$

and there are no smaller values of $\mathcal{S}$ for which (2.1) is satisfied. This definition implies that μ_t and $\gamma_t(\tau)$ are periodic functions in t and need to be known only for $t = 1, \ldots, \mathcal{S}$. If $\mathcal{S} = 1$, (X_t) is weakly stationary in the usual sense.

The natural extension for PC processes of the well-known ARMA model is the PARMA model. (X_t) is said to be a PARMA series with period $\mathcal{S} > 0$ if it is a solution to the difference equation

$$(X_{n\mathcal{S}+\nu} - \mu_\nu) + \sum_{k=1}^{p_\nu} \phi_{\nu,k}(X_{n\mathcal{S}+\nu-k} - \mu_{\nu-k}) = \varepsilon_{n\mathcal{S}+\nu} + \sum_{k=1}^{q_\nu} \theta_{\nu,k}\varepsilon_{n\mathcal{S}+\nu-k}, \tag{2.2}$$

where $X_{n\mathcal{S}+\nu}$ is the series during the νth season, $\nu = 1, \ldots, \mathcal{S}$, of cycle $n \in \mathbb{Z}$, and $(\varepsilon_{n\mathcal{S}+\nu})$ is a sequence of zero mean uncorrelated random variables with $\mathrm{E}(\varepsilon_{n\mathcal{S}+\nu}^2) = \sigma_{n\mathcal{S}+\nu}^2 = \sigma_\nu^2$. The period $\mathcal{S}$ is taken to be the smallest positive integer satisfying (2.2). When $\mathcal{S} = 1$, (2.2) corresponds to the standard ARMA model. During season ν, $p_\nu \geq 0$ and $q_\nu \geq 0$ are the AR and MA orders, respectively, $\phi_\nu = (\phi_{\nu,1}, \ldots, \phi_{\nu,p_\nu})'$ and $\theta_\nu = (\theta_{\nu,1}, \ldots, \theta_{\nu,q_\nu})'$ are the AR and MA parameters, respectively, where A' denotes the transpose of matrix A. The parameter vector of model (2.2) is then $\varphi = (\varphi_\phi', \varphi_\theta', \varphi_\sigma')'$ where $\varphi_\phi = (\phi_1', \ldots, \phi_\mathcal{S}')'$, $\varphi_\theta = (\theta_1', \ldots, \theta_\mathcal{S}')'$ and $\varphi_\sigma = (\sigma_1^2, \ldots, \sigma_\mathcal{S}^2)'$.

In the following, we set

$$p = \max_{1\le\nu\le\mathcal{S}} p_\nu, \qquad \phi_{\nu,k} = 0 \text{ when } p_\nu < k \le p,$$
$$q = \max_{1\le\nu\le\mathcal{S}} q_\nu, \qquad \theta_{\nu,k} = 0 \text{ when } q_\nu < k \le q,$$

for every $\nu = 1, \ldots, \mathcal{S}$, and we refer to (2.2) as the PARMA$(p, q)_\mathcal{S}$ model. For the sake of simplicity, we assume that $\mu_\nu = 0$ for $\nu = 1, \ldots, \mathcal{S}$. Note that, in practical situations, the sample periodic means are, in general, removed from the series before model fitting.

Let $(\mathbf{X}_n)_{n\in\mathbb{Z}}$ be the $\mathcal{S}$-variate process defined by $\mathbf{X}_n = (X_{n\mathcal{S}+1}, \ldots, X_{n\mathcal{S}+\mathcal{S}})'$. It is well known that (X_t) satisfies (2.2) if and only if $(\mathbf{X}_n)$ is a solution to the vector ARMA (VARMA) difference equation

$$\sum_{k=0}^{P} \boldsymbol{\phi}_k \mathbf{X}_{n-k} = \sum_{k=0}^{Q} \boldsymbol{\theta}_k \boldsymbol{\varepsilon}_{n-k}, \tag{2.3}$$

where $\boldsymbol{\varepsilon}_n = (\varepsilon_{n\mathcal{S}+1}, \ldots, \varepsilon_{n\mathcal{S}+\mathcal{S}})'$, the sequence $(\boldsymbol{\varepsilon}_n)$ is uncorrelated and $\mathrm{E}(\boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}_n') = \Sigma$ where Σ is diagonal with element $(\Sigma)_{l,l} = \sigma_l^2$ for $l = 1, \ldots, \mathcal{S}$. The VARMA orders are $P = \lceil p/\mathcal{S} \rceil$ and $Q = \lceil q/\mathcal{S} \rceil$, wherein $\lceil x \rceil$ stands for the smallest integer greater than or equal to x. For every $k = 0, \ldots, P$, the $\mathcal{S} \times \mathcal{S}$ matrix $\boldsymbol{\phi}_k$ has (l, m)th entries

$$(\boldsymbol{\phi}_0)_{l,m} = \begin{cases} 0 & l < m, \\ 1 & l = m, \\ \phi_{l,l-m} & l > m, \end{cases} \quad (\boldsymbol{\phi}_k)_{l,m} = \phi_{l,k\mathcal{S}+l-m}, \quad 1 \le k \le P, \tag{2.4}$$

and the entries of $\boldsymbol{\theta}_k$, for $k = 0, \ldots, Q$, are similarly obtained by replacing $\phi_{l,m}$ by $\theta_{l,m}$ in (2.4). It follows from (2.4) that, for every $\nu = 1, \ldots, \mathcal{S}$,

$$\phi_{\nu,k} = \begin{cases} (\boldsymbol{\phi}_0)_{\nu,\nu-k} & \text{if } 1 \le k < \nu, \\ (\boldsymbol{\phi}_1)_{\nu,\mathcal{S}+\nu-k} & \text{if } \nu \le k < \nu + \mathcal{S}, \\ \quad\vdots & \\ (\boldsymbol{\phi}_P)_{\nu,P\mathcal{S}+\nu-k} & \text{if } \nu + (P-1)\mathcal{S} \le k \le p, \end{cases} \tag{2.5}$$

and $\theta_{\nu,k}$ is similarly obtained by replacing $\boldsymbol{\phi}_k$ by $\boldsymbol{\theta}_k$ in (2.5). Moreover, $\sigma_\nu^2 = (\Sigma)_{\nu,\nu}$. Therefore, φ is uniquely obtained from the $\mathcal{S} \times (P + Q + 3)\mathcal{S}$ matrix defined by $\eta = (\boldsymbol{\phi}_0, \ldots, \boldsymbol{\phi}_P, \boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_Q, \Sigma)$.

Since $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$ are both lower triangular matrices, (2.3) is referred to a "triangular" VARMA representation of $(\mathbf{X}_n)$. Note that this representation does not follow the standard VARMA framework, since $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$ in (2.4) are not the $\mathcal{S} \times \mathcal{S}$ identity matrix $\mathbf{I}$. However, since $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$ are invertible, (2.3) is equivalent to the standard VARMA representation

$$\mathbf{X}_n + \sum_{k=1}^{P} \boldsymbol{\phi}_k^* \mathbf{X}_{n-k} = \boldsymbol{\xi}_n + \sum_{k=1}^{Q} \boldsymbol{\theta}_k^* \boldsymbol{\xi}_{n-k}, \tag{2.6}$$

where

$$\boldsymbol{\phi}_k^* = \boldsymbol{\phi}_0^{-1}\boldsymbol{\phi}_k,\ \boldsymbol{\theta}_k^* = \boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_k\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0, \tag{2.7}$$

and $\boldsymbol{\xi}_n = \boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0\boldsymbol{\varepsilon}_n$. We have $\mathrm{E}(\boldsymbol{\xi}_n\boldsymbol{\xi}_n') = \Sigma^*$ where

$$\Sigma^* = \boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0\Sigma\boldsymbol{\theta}_0'\boldsymbol{\phi}_0'^{-1}. \tag{2.8}$$

Let η^* be the $\mathcal{S}\times(P+Q+1)\mathcal{S}$ matrix defined by $\eta^* = (\boldsymbol{\phi}_1^*, \ldots, \boldsymbol{\phi}_P^*, \boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_Q^*, \Sigma^*)$. The parameters η^* are uniquely defined from the parameters φ, say $\eta^* = f(\varphi)$. However, map f is not necessarily one-to-one as illustrated by the two following examples.

Example 2.1. *Consider a PARMA$(1,1)_2$ process (X_t) with $\phi_{1,1} = \theta_{1,1} = 0$. The nonzero parameters in its triangular VARMA representation* (2.3) *are*

$$\boldsymbol{\phi}_0 = \begin{bmatrix} 1 & 0 \\ \phi_{2,1} & 1 \end{bmatrix},\ \boldsymbol{\theta}_0 = \begin{bmatrix} 1 & 0 \\ \theta_{2,1} & 1 \end{bmatrix},\ \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

The corresponding standard representation (2.6) *reduces to $\mathbf{X}_n = \boldsymbol{\xi}_n$, where the covariance matrix of the bivariate white noise $(\boldsymbol{\xi}_n)$ is*

$$\Sigma^* = \begin{bmatrix} \sigma_1^2 & (\theta_{2,1} - \phi_{2,1})\sigma_1^2 \\ (\theta_{2,1} - \phi_{2,1})\sigma_1^2 & (\theta_{2,1} - \phi_{2,1})^2\sigma_1^2 + \sigma_2^2 \end{bmatrix}. \tag{2.9}$$

It is easy to see that, for any $a \in \mathbb{R}$, the PARMA$(1,1)_2$ process $(X_t^{\blacktriangle})$ with parameters $\phi_{1,1}^{\blacktriangle} = \theta_{1,1}^{\blacktriangle} = 0$, $\phi_{2,1}^{\blacktriangle} = \phi_{2,1} + a$, $\theta_{2,1}^{\blacktriangle} = \theta_{2,1} + a$ and with white noise variances $\sigma_1^{2\blacktriangle} = \sigma_1^2$ and $\sigma_2^{2\blacktriangle} = \sigma_2^2$ has the same representation (2.6) *as $\mathbf{X}_n$, i.e. $\mathbf{X}_n^{\blacktriangle} = \boldsymbol{\xi}_n^{\blacktriangle}$ where $\Sigma^{\blacktriangle *} = \Sigma^*$.*

Example 2.2. *Consider a PARMA$(1,2)_2$ process (X_t) with $\phi_{1,1} = \theta_{1,2} = 0$. Its triangular VARMA representation* (2.3) *has the nonzero parameters*

$$\boldsymbol{\phi}_0 = \begin{bmatrix} 1 & 0 \\ \phi_{2,1} & 1 \end{bmatrix},\ \boldsymbol{\theta}_0 = \begin{bmatrix} 1 & 0 \\ \theta_{2,1} & 1 \end{bmatrix},\ \boldsymbol{\theta}_1 = \begin{bmatrix} 0 & \theta_{1,1} \\ 0 & \theta_{2,2} \end{bmatrix},\ \Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

Its standard VARMA representation (2.6) *is $\mathbf{X}_n = \boldsymbol{\xi}_n + \boldsymbol{\theta}_1^*\boldsymbol{\xi}_{n-1}$ where*

$$\boldsymbol{\theta}_1^* = \begin{bmatrix} \theta_{1,1}(\phi_{2,1} - \theta_{2,1}) & \theta_{1,1} \\ (\theta_{2,2} - \theta_{1,1}\phi_{2,1})(\phi_{2,1} - \theta_{2,1}) & \theta_{2,2} - \theta_{1,1}\phi_{2,1} \end{bmatrix}$$

and Σ^* is given by (2.9)*. One can see that, for any $a \in \mathbb{R}$, the PARMA$(1,2)_2$ process $(X_t^{\blacktriangle})$ with parameters $\phi_{1,1}^{\blacktriangle} = \theta_{1,2}^{\blacktriangle} = 0$, $\phi_{2,1}^{\blacktriangle} = \phi_{2,1} + a$, $\theta_{2,1}^{\blacktriangle} = \theta_{2,1} + a$, $\theta_{1,1}^{\blacktriangle} = \theta_{1,1}$, $\theta_{2,2}^{\blacktriangle} = \theta_{2,2} + \theta_{1,1}a$ and with white noise variances $\sigma_1^{2\blacktriangle} = \sigma_1^2$ and $\sigma_2^{2\blacktriangle} = \sigma_2^2$ has the same VARMA representation* (2.6) *as $\mathbf{X}_n$, i.e. $\mathbf{X}_n^{\blacktriangle} = \boldsymbol{\xi}_n^{\blacktriangle} + \boldsymbol{\theta}_1^*\boldsymbol{\xi}_{n-1}^{\blacktriangle}$ where $\Sigma^{\blacktriangle *} = \Sigma^*$.*

The fact that different PARMA models, with the same orders, may have the same standard VARMA representation induces identifiability problems of PARMA models based on representation (2.6). Therefore, finding conditions to ensure that map f be one-to-one is an important issue. In this context, the following assumptions are introduced:

(A1) The AR orders p_ν's of the PARMA process (X_t) are the same for every $\nu = 1, \ldots, \mathcal{S}$ in (2.2).

(A2) The MA orders q_ν's of the PARMA process (X_t) are the same for every $\nu = 1, \ldots, \mathcal{S}$ in (2.2).

Lemma 2.1. *If* **(A1)** *and/or* **(A2)** *hold, then* φ *is uniquely obtained from* η^*.

Assumptions **(A1)** and **(A2)** are easy to verify and give sufficient conditions to guarantee that map f be one-to-one. **(A1)** and **(A2)** are not necessary conditions as shown by the following example.

Example 2.3. *Consider the class of PARMA*$(1,1)_2$ *processes* (X_t) *satisfying* $\phi_{2,1} = \theta_{1,1} = 0$. *The corresponding triangular VARMA representation* (2.3) *is* $\mathbf{X}_n + \boldsymbol{\phi}_1 \mathbf{X}_{n-1} = \boldsymbol{\theta}_0 \boldsymbol{\varepsilon}_n$, *where*

$$\boldsymbol{\phi}_1 = \begin{bmatrix} 0 & \phi_{1,1} \\ 0 & 0 \end{bmatrix}, \boldsymbol{\theta}_0 = \begin{bmatrix} 1 & 0 \\ \theta_{2,1} & 1 \end{bmatrix},$$

and the standard representation (2.6) *is* $\mathbf{X}_n + \boldsymbol{\phi}_1 \mathbf{X}_{n-1} = \boldsymbol{\xi}_n$, *where*

$$\Sigma^* = \begin{bmatrix} \sigma_1^2 & \theta_{2,1}\sigma_1^2 \\ \theta_{2,1}\sigma_1^2 & \theta_{2,1}^2\sigma_1^2 + \sigma_2^2 \end{bmatrix}.$$

It is readily seen that the parameter vector $(\phi_{1,1}, \theta_{2,1}, \sigma_1^2, \sigma_2^2)$ *is uniquely determined from* $\boldsymbol{\phi}_1$ *and* Σ^*, *while* **(A1)** *and* **(A2)** *are not satisfied.*

For all $z \in \mathbb{C}$, let

$$\begin{aligned} \Phi(z) &= \sum_{k=0}^{P} \boldsymbol{\phi}_k z^k, \quad \Phi^*(z) = \mathbf{I} + \sum_{k=1}^{P} \boldsymbol{\phi}_k^* z^k, \\ \Theta(z) &= \sum_{k=0}^{Q} \boldsymbol{\theta}_k z^k, \quad \Theta^*(z) = \mathbf{I} + \sum_{k=1}^{Q} \boldsymbol{\theta}_k^* z^k. \end{aligned} \tag{2.10}$$

It results from (2.7) that

$$\Phi^*(z) = \boldsymbol{\phi}_0^{-1} \Phi(z) \quad \text{and} \quad \Theta^*(z) = \boldsymbol{\phi}_0^{-1} \Theta(z) \boldsymbol{\theta}_0^{-1} \boldsymbol{\phi}_0. \tag{2.11}$$

Since (X_t) in (2.2) is PC with period $\mathcal{S}$, the vector process $(\mathbf{X}_n)$ in (2.3) is weakly stationary. The autocovariance matrix function of $(\mathbf{X}_n)$ is $\Gamma(\tau) = \mathrm{Cov}(\mathbf{X}_n, \mathbf{X}_{n-\tau})$ and is related to $\gamma_t(\tau)$ by $\Gamma_{l,m}(\tau) = \gamma_l(\tau\mathcal{S} + l - m)$ for every $l, m = 1, \ldots, \mathcal{S}$. The causality and invertibility of (X_t) are equivalent to the

causality and invertibility of $(\mathbf{X}_n)$, respectively. For more details, we refer to Gladyshev (1961) and Hurd and Miamee (2007). Therefore, (X_t) is a causal solution of (2.2) if and only if $(\mathbf{X}_n)$ is a stationary causal solution of (2.6), and this is the case according to (Brockwell and Davis, 2006, Theorem 11.3.1) whenever $\det \Phi^*(z) \neq 0$ for $|z| \leq 1$. Similar arguments jointly with (Brockwell and Davis, 2006, Theorem 11.3.2) show that (X_t) is a PC invertible solution of (2.2) when $\det \Theta^*(z) \neq 0$ for $|z| \leq 1$.

The standard VARMA representation (2.6) is said to be identifiable if Σ^*, $\Phi^*(z)$ and $\Theta^*(z)$ are uniquely determined by the autocovariance matrix function of $(\mathbf{X}_n)$, or equivalently by the spectral density matrix of $(\mathbf{X}_n)$. When the model is non identifiable, the Gaussian likelihood may have more than one maximum. Causality and invertibility properties do not ensure the identifiability of model (2.6), see e.g. (Brockwell and Davis, 2006, page 431) and (Reinsel, 1997, section 2.3). Further restrictions have to be imposed in order to obtain an identifiable model.

Following Dunsmuir and Hannan (1976), two $\mathcal{S} \times \mathcal{S}$ matrices of polynomials $g(z)$ and $h(z)$ are said to be left prime when they have no common left factors other than unimodular ones, that is, if $g(z) = e(z)g_1(z)$ and $h(z) = e(z)h_1(z)$ where $e(z)$, $g_1(z)$, $h_1(z)$ are matrices of polynomials, then $e(z)$ has constant determinant. It is known (Heymann, 1975) that $g(z)$ and $h(z)$ are left prime if and only if the $\mathcal{S} \times 2\mathcal{S}$ matrix $[g(z), h(z)]$ has rank $\mathcal{S}$ for all $z \in \mathbb{C}$.

Now, following Deistler et al. (1978), for every $i = 1, \ldots, \mathcal{S}$, let $g_i(z)$ and $h_i(z)$ be the ith column of $g(z)$ and $h(z)$, p_i and q_i be the maximum degrees of $g_i(z)$ and $h_i(z)$, $g_i(j)$ and $h_i(j)$ be the vectors of coefficients of z^j in $g_i(z)$ and $h_i(z)$, respectively. We define the $\mathcal{S} \times 2\mathcal{S}$ matrix

$$\mathcal{G}(g(z), h(z)) = [g_1(p_1), \ldots, g_{\mathcal{S}}(p_{\mathcal{S}}), h_1(q_1), \ldots, h_{\mathcal{S}}(q_{\mathcal{S}})].$$

Deistler et al. (1978) show that the standard VARMA representation (2.6) is identifiable if the following three assumptions are satisfied:

(A3*) $\det \Phi^*(z) \neq 0$ and $\det \Theta^*(z) \neq 0$ for $|z| \leq 1$,
(A4*) $\Phi^*(z)$ and $\Theta^*(z)$ are left prime,
(A5*) $\operatorname{rank} \mathcal{G}(\Phi^*(z), \Theta^*(z)) = \mathcal{S}$.

Remark 2.1. *It is easy to verify that the PARMA models in Examples 2.1 and 2.2 satisfy* **(A4*)** *and* **(A5*)**. *Moreover,* **(A3*)** *is always satisfied in Example 2.1 and is satisfied in Example 2.2 when* $|\theta_{1,1}\theta_{2,1} - \theta_{2,2}| < 1$ *since* $\det \Theta^*(z) = \det(\boldsymbol{\phi}_0^{-1}\Theta(z)\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0) = \det \Theta(z) = 1 + (\theta_{2,2} - \theta_{1,1}\theta_{2,1})z$. *Then, under this restriction, the standard VARMA representations* (2.6) *in Examples 2.1 and 2.2 are identifiable whereas, as shown above, the corresponding PARMA models* (2.2) *are not.*

The following assumptions are introduced:

(A3) $\det \Phi(z) \neq 0$ and $\det \Theta(z) \neq 0$ for $|z| \leq 1$,
(A4) $\Phi(z)$ and $\Theta(z)$ are left prime,
(A5) $\operatorname{rank} \mathcal{G}(\Phi(z), \Theta(z)\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0) = \mathcal{S}$.

Lemma 2.2. *Assumptions* **(A3)**, **(A4)** *and* **(A5)** *are equivalent to* **(A3*)**, **(A4*)** *and* **(A5*)**, *respectively.*

The following theorem gives sufficient conditions of identifiability for model in (2.2).

Theorem 2.1. *If either* **(A1)** *or* **(A2)** *and* **(A5)** *hold, and in addition* **(A3)** *and* **(A4)** *are satisfied, then the parameter vector* φ *of model* (2.2) *is uniquely determined by the autocovariance matrix function or by the spectral density matrix of* $(\mathbf{X}_n)$.

Remark 2.2. *If* $p = 0$ *in* (2.2), $\boldsymbol{\phi}_0 = \mathbf{I}$ *in* (2.4) *and it follows from* (2.7) *and* (2.8) *that* $\boldsymbol{\phi}_k = \boldsymbol{\phi}_k^*$, $\boldsymbol{\theta}_k = \boldsymbol{\theta}_k^*\boldsymbol{\theta}_0$ *and* $\boldsymbol{\theta}_0\Sigma\boldsymbol{\theta}_0'$ *is the Cholesky decomposition of* Σ^*. *Therefore,* Σ *and* $\boldsymbol{\theta}_0$ *are uniquely obtained from* Σ^*. *Then* φ *is uniquely determined from* η^* *and the PARMA model* (2.2) *is identifiable when the standard VARMA model* (2.6) *is. In the same way, when* $q = 0$ *in* (2.2), $\boldsymbol{\theta}_0 = \mathbf{I}$, $\boldsymbol{\phi}_k = \boldsymbol{\phi}_0\boldsymbol{\phi}_k^*$, $\boldsymbol{\theta}_k = \boldsymbol{\phi}_0\boldsymbol{\theta}_k^*\boldsymbol{\phi}_0^{-1}$ *and* $\boldsymbol{\phi}_0^{-1}\Sigma\boldsymbol{\phi}_0'^{-1}$ *is the Cholesky decomposition of* Σ^*. *Therefore,* Σ *and* $\boldsymbol{\phi}_0$ *are uniquely obtained from* Σ^* *and* φ *is uniquely determined from* η^*. *Again, the PARMA model* (2.2) *is identifiable when the standard VARMA model* (2.6) *is.*

3 Whittle estimation

To simplify the notation, we denote by the same symbol $\mathcal{P}$ the set of points $(\varphi_\phi', \varphi_\theta')' \subset \mathbb{R}^{(p+q)\mathcal{S}}$ which satisfy either **(A1)**, **(A3)** and **(A4)**, or **(A2)**, **(A3)**, **(A4)** and **(A5)**. We denote the true parameter vector of model (2.2) by $\varphi_0 = (\varphi_{\phi_0}', \varphi_{\theta_0}', \varphi_{\sigma_0}')'$ and we assume that $(\varphi_{\phi_0}', \varphi_{\theta_0}')' \in \mathcal{P}$. We suppose also that the sample contains N full periods of data which are indexed from 0 to $N-1$ and we set $\mathbf{X} = (X_1, \ldots, X_{N\mathcal{S}})' = (\mathbf{X}_0', \ldots, \mathbf{X}_{N-1}')'$.

We denote by $\mathbb{R}_{>0}$ the set of positive real numbers. For any $\varphi \in \mathcal{P} \times \mathbb{R}_{>0}^{\mathcal{S}}$, let $\Gamma_N(\varphi)$ be the $N\mathcal{S} \times N\mathcal{S}$ matrix with $\Gamma(m-l)$ in the (l,m)th block of $\mathcal{S} \times \mathcal{S}$ elements. Then, $\Gamma_N(\varphi_0) = \mathrm{Cov}(\mathbf{X}, \mathbf{X})$. Let

$$\hat{\mathcal{L}}_N(\varphi) = N^{-1} \log\det \Gamma_N(\varphi) + N^{-1}\mathbf{X}'\Gamma_N^{-1}(\varphi)\mathbf{X},$$

be the Gaussian log likelihood with the scaling factor $-2N^{-1}$. The Gaussian MLE of φ_0 is

$$\hat{\varphi}_N = \underset{\varphi \in \mathcal{P} \times \mathbb{R}_{>0}^{\mathcal{S}}}{\operatorname{argmin}} \ \hat{\mathcal{L}}_N(\varphi).$$

In most cases the minimization of $\hat{\mathcal{L}}_N(\varphi)$ is performed through optimization algorithms, which can demand high computational effort, since *a priori* it is necessary to invert $\Gamma_N(\varphi)$. One alternative is to resort to the recursive likelihood evaluation technique proposed by Lund and Basawa (2000). However, as illustrated in Section 4, the computational cost of this method may be impor-

tant for large sample sizes. To circumvent this difficulty, we use the multivariate version of Whittle's methodology to approximate $\hat{\mathcal{L}}_N(\varphi)$. The multivariate periodogram of $\mathbf{X}$ at frequency $\omega \in [-\pi, \pi]$ is $I(\omega) = W(e^{-\mathbf{i}\omega})W'(e^{\mathbf{i}\omega})$, where

$$W(e^{\mathbf{i}\omega}) = (2\pi N)^{-1/2} \sum_{n=0}^{N-1} \mathbf{X}_n e^{\mathbf{i}n\omega}.$$

The spectral density matrix of $(\mathbf{X}_n)$ is $\mathbf{f}(\omega, \varphi_0)$, where

$$\mathbf{f}(\omega, \varphi) = \frac{1}{2\pi} \Phi^{*-1}(e^{-\mathbf{i}\omega}) \Theta^*(e^{-\mathbf{i}\omega}) \Sigma^* \Theta^{*\prime}(e^{\mathbf{i}\omega}) \Phi^{*\prime-1}(e^{\mathbf{i}\omega}). \tag{3.1}$$

Following Dunsmuir and Hannan (1976), we approximate $\hat{\mathcal{L}}_N(\varphi)$ by

$$\tilde{\mathcal{L}}_N(\varphi) = \log\det \Sigma^* + N^{-1} \sum_{j=0}^{N-1} \operatorname{tr}[\mathbf{f}^{-1}(\omega_j, \varphi) I(\omega_j)], \tag{3.2}$$

where $\omega_j = 2\pi j/N$ and $\operatorname{tr} A$ is the trace of matrix A. This approximation is particularly interesting from a computational point of view. According to (2.8),

$$\log\det \Sigma^* = \log\det \Sigma = \sum_{l=1}^{\mathcal{S}} \log \sigma_l^2, \tag{3.3}$$

and it follows from (2.8), (2.11) and (3.1) that

$$\mathbf{f}(\omega, \varphi) = \frac{1}{2\pi} \Phi^{-1}(e^{-\mathbf{i}\omega}) \Theta(e^{-\mathbf{i}\omega}) \Sigma \Theta'(e^{\mathbf{i}\omega}) \Phi'^{-1}(e^{\mathbf{i}\omega}). \tag{3.4}$$

Then

$$\operatorname{tr}[\mathbf{f}^{-1}(\omega_j, \varphi) I(\omega_j)] = W'(e^{\mathbf{i}\omega_j}) \mathbf{f}^{-1}(\omega_j, \varphi) W(e^{-\mathbf{i}\omega_j}) = 2\pi \sum_{l=1}^{\mathcal{S}} \sigma_l^{-2} \left| \left(\Theta^{-1}(e^{-\mathbf{i}\omega_j}) \Phi(e^{-\mathbf{i}\omega_j}) W(e^{-\mathbf{i}\omega_j}) \right)_l \right|^2, \tag{3.5}$$

and replacing (3.3) and (3.5) in (3.2), we get that

$$\tilde{\mathcal{L}}_N(\varphi) = \sum_{l=1}^{\mathcal{S}} \left[\log \sigma_l^2 + \frac{2\pi}{N\sigma_l^2} \sum_{j=0}^{N-1} \left| \left(\Theta^{-1}(e^{-\mathbf{i}\omega_j}) \Phi(e^{-\mathbf{i}\omega_j}) W(e^{-\mathbf{i}\omega_j}) \right)_l \right|^2 \right]. \tag{3.6}$$

The WLE of φ_0 is

$$\tilde{\varphi}_N = \operatorname*{argmin}_{\varphi \in \mathcal{P} \times \mathbb{R}^{\mathcal{S}}_{>0}} \tilde{\mathcal{L}}_N(\varphi).$$

For every $l = 1, \ldots, \mathcal{S}$, the minimum of (3.6) with respect to σ_l^2 is

$$\tilde{\sigma}^2_{l,N}(\varphi_\phi, \varphi_\theta) = \frac{2\pi}{N} \sum_{j=0}^{N-1} \left| \left(\Theta^{-1}(e^{-\mathbf{i}\omega_j})\Phi(e^{-\mathbf{i}\omega_j})W(e^{-\mathbf{i}\omega_j})\right)_l \right|^2. \tag{3.7}$$

Replacing (3.7) in (3.6), we get that the WLE of $(\varphi'_{\phi_0}, \varphi'_{\theta_0})'$ is

$$(\tilde{\varphi}'_{\phi_N}, \tilde{\varphi}'_{\theta_N})' = \underset{(\varphi'_\phi, \varphi'_\theta)' \in \mathcal{P}}{\operatorname{argmin}} \sum_{l=1}^{\mathcal{S}} \log \tilde{\sigma}^2_{l,N}(\varphi_\phi, \varphi_\theta).$$

Therefore, $\tilde{\varphi}_N = (\tilde{\varphi}'_{\phi_N}, \tilde{\varphi}'_{\theta_N}, \tilde{\varphi}'_{\sigma_N})'$, where $(\tilde{\varphi}_{\sigma_N})_l = \tilde{\sigma}^2_{l,N}(\tilde{\varphi}_{\phi_N}, \tilde{\varphi}_{\theta_N})$. Observe that $\tilde{\varphi}_N$ is easier to calculate than $\hat{\varphi}_N$ since $\hat{\mathcal{L}}_N(\varphi)$ involves $(p+q+1)\mathcal{S}$ parameters while $\tilde{\sigma}^2_{l,N}(\varphi_\phi, \varphi_\theta)$ is a function of $(p+q)\mathcal{S}$ parameters.

Theorem 3.1. *For any $\varphi_0 \in \mathcal{P} \times \mathbb{R}^{\mathcal{S}}_{>0}$, $\tilde{\varphi}_N$ converges almost surely (a.s.) to φ_0 as N tends to infinity.*

To establish asymptotic normality, we now introduce the following additional assumption:

(A6) For all $t \in \mathbb{Z}$, $t = n\mathcal{S} + \nu$ where $n \in \mathbb{Z}$ and $\nu = 1, \ldots, \mathcal{S}$, $\mathrm{E}(\varepsilon_t^4) < \infty$ and

a. $\mathrm{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$,
b. $\mathrm{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_\nu^2$,
c. $\mathrm{E}(\varepsilon_t^3|\mathcal{F}_{t-1}) = \beta_\nu$,
d. $\mathrm{E}(\varepsilon_t^4) = \gamma_\nu$,

where σ_ν^2, β_ν and γ_ν are constants and $\mathcal{F}_t$ is the σ-algebra generated by $\{\varepsilon_s; s \le t\}$.

In the following, it is convenient to let φ_k denote the kth component of φ for $k = 1, \ldots, (p+q+1)\mathcal{S}$, and $\mathbf{f}_0(\omega)$, $\frac{\partial \mathbf{f}_0(\omega)}{\partial \varphi_k}$ denote respectively $\mathbf{f}(\omega, \varphi)$ and $\frac{\partial \mathbf{f}(\omega,\varphi)}{\partial \varphi_k}$ evaluated at $\varphi = \varphi_0$. Let Ω be the $(p+q+1)\mathcal{S} \times (p+q+1)\mathcal{S}$ matrix with (k,l) entries

$$\Omega_{k,l} = \frac{1}{2\pi} \int_0^{2\pi} \mathrm{tr}\left[\mathbf{f}_0^{-1}(\omega) \frac{\partial \mathbf{f}_0(\omega)}{\partial \varphi_k} \mathbf{f}_0^{-1}(\omega) \frac{\partial \mathbf{f}_0(\omega)}{\partial \varphi_l}\right] d\omega,$$

and Π be the diagonal $(p+q+1)\mathcal{S} \times (p+q+1)\mathcal{S}$ matrix with (k,k) entries

$$\Pi_{k,k} = \begin{cases} 0 & \text{if } k = 1, \ldots, (p+q)\mathcal{S}, \\ \kappa_\nu/\sigma_\nu^8 & \text{if } k = (p+q)\mathcal{S} + \nu \text{ and } \nu = 1, \ldots, \mathcal{S}, \end{cases} \tag{3.8}$$

where $\kappa_\nu = \gamma_\nu - 3\sigma_\nu^4$ is the fourth order cumulant of $\varepsilon_{n\mathcal{S}+\nu}$. These matrices coincide with the matrices Ω and Π introduced in (Dunsmuir, 1979, Theorem 2.2), and (3.8) is obtained from (Dunsmuir, 1979, Remark 3). It follows from (3.4) that $\Omega_{k,l} = 0$ when $k, l = (p+q)\mathcal{S}+1, \ldots, (p+q)\mathcal{S}+\mathcal{S}$ and $k \neq l$. We

deduce from (Dunsmuir, 1979, Remark 4) that the matrix $\Omega^{-1}(2\Omega+\Pi)\Omega^{-1}$ reduces to

$$\begin{bmatrix} 2\Omega^{(1)^{-1}} & 0 \\ 0 & \Omega^{(2)^{-1}}(2\Omega^{(2)}+\Pi^{(2)})\Omega^{(2)^{-1}} \end{bmatrix}, \tag{3.9}$$

where $\Omega^{(1)} = (\Omega_{k,l})_{k,l=1,\ldots,(p+q)\mathcal{S}}$ and $\Omega^{(2)^{-1}}(2\Omega^{(2)}+\Pi^{(2)})\Omega^{(2)^{-1}}$ is the diagonal $\mathcal{S}\times\mathcal{S}$ matrix with diagonal elements $(2\Omega_{k,k}+\Pi_{k,k})/\Omega_{k,k}^2$ for $k=(p+q)\mathcal{S}+1,\ldots,(p+q)\mathcal{S}+\mathcal{S}$.

Theorem 3.2. *For any $\varphi_0 \in \mathcal{P}\times\mathbb{R}_{>0}^{\mathcal{S}}$, under assumption* **(A6)**, $N^{1/2}(\tilde{\varphi}_N - \varphi_0)$ *converges in law to a normal distribution with zero mean vector and covariance matrix* (3.9)*, as N tends to infinity.*

4 Monte Carlo study

We compare by Monte Carlo simulations the finite sample properties of the WLE and the exact MLE obtained with the algorithm in Lund and Basawa (2000). For this purpose, we simulate PARMA models with $\mathcal{S}=2$, $N=50,100,150$ and 200 and the series has complete periods. The empirical bias, root mean squared error (RMSE) and the computation time of the WLE and MLE are based on the mean of $M=1000$ replications.

Consider a PARMA$(1,1)_2$ model. The non-zero parameters in its triangular VARMA representation (2.3) are

$$\boldsymbol{\phi}_0 = \begin{bmatrix} 1 & 0 \\ \phi_{2,1} & 1 \end{bmatrix}, \boldsymbol{\phi}_1 = \begin{bmatrix} 0 & \phi_{1,1} \\ 0 & 0 \end{bmatrix}, \boldsymbol{\theta}_0 = \begin{bmatrix} 1 & 0 \\ \theta_{2,1} & 1 \end{bmatrix}, \boldsymbol{\theta}_1 = \begin{bmatrix} 0 & \theta_{1,1} \\ 0 & 0 \end{bmatrix}$$

and

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}.$$

Then, for all $z\in\mathbb{C}$,

$$\Phi(z) = \begin{bmatrix} 1 & \phi_{1,1}z \\ \phi_{2,1} & 1 \end{bmatrix} \quad \text{and} \quad \Theta(z) = \begin{bmatrix} 1 & \theta_{1,1}z \\ \theta_{2,1} & 1 \end{bmatrix}.$$

Condition (**A1**), respectively (**A2**), is equivalent to $\phi_{1,1}\phi_{2,1}\neq 0$, respectively, $\theta_{1,1}\theta_{2,1}\neq 0$. Condition (**A3**) writes $|\phi_{1,1}\phi_{2,1}|<1$ and $|\theta_{1,1}\theta_{2,1}|<1$. When $\phi_{1,1}\neq\theta_{1,1}$ or $\phi_{2,1}\neq\theta_{2,1}$, (**A4**) holds since $[\Phi(z),\Theta(z)]$ has rank 2 for all $z\in\mathbb{C}$. In order to

$$\mathcal{G}(\Phi(z),\Theta(z)\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0) = \begin{bmatrix} 1 & \phi_{1,1} & \theta_{1,1}(\phi_{2,1}-\theta_{2,1}) & \theta_{1,1} \\ \phi_{2,1} & 0 & 0 & 0 \end{bmatrix}$$

be full rank, it is sufficient that (**A1**) holds. If (**A1**) does not hold but (**A2**) holds, it is additionally required that $\phi_{2,1}\neq 0$.

Monte Carlo experiments are carried out with four PARMA$(1,1)_2$ models whose parameters are provided in Table 1 and the innovation processes (ε_t) are Gaussian. These models satisfy **(A1)**, **(A2)**, **(A3)**, **(A4)**, **(A5)**, **(A6)**, and are chosen to evaluate the effect caused by the proximity of the parameters to the regions of noncausality and noninvertibility.

Model 1 is far from both the noncausality and noninvertibility regions. Models 2 and 3 are close to noncausality and noninvertibility regions, respectively. Model 4 is close from both noncausality and noninvertibility regions. The numerical optimization procedures are initialized with the true values of the parameters.

Table 1: PARMA$(1,1)_2$ models.

	Parameters					
	$\nu = 1$			$\nu = 2$		
Model	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
1	-0.7	0.4	1.0	-0.5	0.8	1.0
2	-1.0	0.4	1.0	-0.7	0.8	1.0
3	-0.7	0.6	1.0	-0.5	1.1	1.0
4	-1.0	0.6	1.0	-0.7	1.1	1.0

4.1 Bias

Let $\hat{\varphi}_{N,k}$ and $\tilde{\varphi}_{N,k}$ be respectively the MLE and the WLE of φ_0 obtained in the kth experiment, $k = 1, \ldots, M$. The empirical bias of the MLE and the WLE are respectively,

$$M^{-1}\sum_{k=1}^{M}\hat{\varphi}_{N,k} - \varphi_0 \quad \text{and} \quad M^{-1}\sum_{k=1}^{M}\tilde{\varphi}_{N,k} - \varphi_0.$$

Tables 2, 3, 4 and 5 display the empirical bias of the MLE and WLE, for Models 1, 2, 3 and 4, respectively. The results show that the bias decreases as the sample size increases for both estimators, which is an expected result based on the asymptotic theories previously discussed. Furthermore, both estimation methods overestimate the AR parameters. However, the estimators behave differently in the estimation of the MA parameters and the white noise variances. The MA parameters are overestimated and underestimated by MLE and WLE approaches, respectively. On the other hand, the white noise variances are underestimated and overestimated by the MLE and WLE methods, respectively. The empirical phenomena of underestimation and overestimation are expected results, since the variance of (X_t) (Eq. 2.2) is defined as a balanced relation between the white noise variance and the parameters of the process. The proximity of noncausality or noninvertibility regions does

not seem to have a significant effect on the bias of the MLE. However, this empirical property does not hold for the WLE method. In this case, the bias of the MA parameter and the variance increase substantially (compare Tables 4 and 5). In general, for models with parameters in a very stable region, both methods perform similarly for time series with size $N \geq 200$ which is a sample size quite common in many areas of applications.

Table 2: Empirical bias of the MLE and the WLE for Model 1.

		Bias					
		$\nu = 1$			$\nu = 2$		
	N	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	50	0.014	0.027	-0.023	0.017	0.036	-0.051
	100	0.007	0.011	-0.025	0.009	0.016	-0.024
	150	0.005	0.011	-0.010	0.006	0.009	-0.012
	200	0.001	0.001	-0.007	0.003	0.008	-0.010
WLE	50	0.026	-0.003	0.067	0.014	-0.034	0.004
	100	0.015	-0.004	0.025	0.009	-0.022	0.006
	150	0.010	-0.002	0.025	0.005	-0.019	0.010
	200	0.005	-0.006	0.018	0.003	-0.011	0.006

Table 3: Empirical bias of the MLE and the WLE for Model 2.

		Bias					
		$\nu = 1$			$\nu = 2$		
	N	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	50	0.015	0.031	-0.036	0.015	0.035	-0.057
	100	0.007	0.014	-0.016	0.010	0.013	-0.021
	150	0.004	0.008	-0.011	0.003	0.010	-0.017
	200	0.004	0.005	-0.008	0.004	0.009	-0.013
WLE	50	0.035	-0.023	0.266	0.014	-0.126	0.069
	100	0.017	-0.021	0.166	0.010	-0.089	0.059
	150	0.011	-0.015	0.104	0.003	-0.062	0.038
	200	0.009	-0.015	0.082	0.004	-0.051	0.035

4.2 Root mean squared error

The empirical RMSE of the MLE and the WLE are, respectively,

$$\left(M^{-1}\sum_{k=1}^{M}(\hat{\varphi}_{N,k}-\varphi_0)^2\right)^{1/2} \quad \text{and} \quad \left(M^{-1}\sum_{k=1}^{M}(\tilde{\varphi}_{N,k}-\varphi_0)^2\right)^{1/2}.$$

Table 4: Empirical bias of the MLE and the WLE for Model 3.

		Bias					
		$\nu = 1$			$\nu = 2$		
	N	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	50	0.012	0.030	-0.038	0.019	0.040	-0.062
	100	0.002	0.011	-0.020	0.003	0.017	-0.023
	150	0.004	0.008	-0.017	0.005	0.008	-0.014
	200	0.003	0.006	-0.011	0.005	0.007	-0.020
WLE	50	0.025	-0.063	0.118	0.018	-0.105	0.092
	100	0.010	-0.044	0.077	0.003	-0.077	0.077
	150	0.008	-0.031	0.043	0.005	-0.051	0.049
	200	0.007	-0.026	0.039	0.005	-0.043	0.035

Table 5: Empirical bias of the MLE and the WLE for Model 4.

		Bias					
		$\nu = 1$			$\nu = 2$		
	N	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	50	0.012	0.040	-0.037	0.015	0.049	-0.056
	100	0.007	0.015	-0.025	0.009	0.016	-0.028
	150	0.003	0.010	-0.012	0.003	0.010	-0.010
	200	0.000	0.007	-0.008	0.000	0.006	-0.015
WLE	50	0.034	-0.118	0.471	0.017	-0.263	0.273
	100	0.018	-0.087	0.246	0.009	-0.182	0.183
	150	0.010	-0.071	0.174	0.004	-0.140	0.154
	200	0.006	-0.062	0.142	0.001	-0.120	0.121

Tables 6, 7, 8 and 9 display the empirical RMSE of the MLE and WLE methods, for models 1, 2, 3 and 4, respectively. Tables show that the RMSE decreases as the sample size increases for both estimators, which corroborates the asymptotic theory discussed earlier. In general, the approximation of the parameters to noncausality or noninvertibility regions seems to have no significant effect in the RMSE of the MLE, which is an expected result based on the bias performance. We observe that the RMSE are smaller (especially for the estimation of the AR parameters) for Models 2, 3 and 4 than for Model 1. The same phenomenon appears with the WLE for the estimation of the AR parameters. The proximity of noncausality and noninvertibility regions increases significantly the RMSE of the WLE of the MA parameters and white noise variances, and the worse results are obtained for Model 4.

4.3 Computation time

For each estimator, the mean computation time is the average of the computation times obtained in each Monte Carlo experiment. For each simulation,

Table 6: Empirical RMSE of the MLE and the WLE for Model 1.

		RMSE					
		$\nu = 1$			$\nu = 2$		
	N	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	50	0.098	0.156	0.210	0.120	0.178	0.200
	100	0.067	0.099	0.140	0.083	0.128	0.144
	150	0.055	0.080	0.116	0.069	0.100	0.114
	200	0.046	0.070	0.100	0.058	0.085	0.097
WLE	50	0.104	0.161	0.258	0.124	0.195	0.216
	100	0.071	0.104	0.162	0.084	0.137	0.153
	150	0.057	0.082	0.131	0.069	0.105	0.119
	200	0.047	0.072	0.110	0.059	0.090	0.101

Table 7: Empirical RMSE of the MLE and the WLE for Model 2.

		RMSE					
		$\nu = 1$			$\nu = 2$		
	N	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	50	0.058	0.139	0.205	0.064	0.151	0.202
	100	0.044	0.089	0.145	0.050	0.108	0.145
	150	0.035	0.070	0.113	0.037	0.083	0.118
	200	0.029	0.059	0.106	0.031	0.069	0.102
WLE	50	0.077	0.151	0.559	0.066	0.240	0.267
	100	0.050	0.102	0.351	0.050	0.177	0.185
	150	0.038	0.078	0.229	0.038	0.130	0.146
	200	0.032	0.066	0.190	0.031	0.109	0.123

Table 8: Empirical RMSE of the MLE and the WLE for Model 3.

		RMSE					
		$\nu = 1$			$\nu = 2$		
	N	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	50	0.080	0.129	0.205	0.111	0.169	0.208
	100	0.055	0.080	0.143	0.075	0.109	0.138
	150	0.045	0.061	0.115	0.060	0.084	0.116
	200	0.039	0.055	0.102	0.053	0.081	0.103
WLE	50	0.090	0.165	0.319	0.116	0.239	0.284
	100	0.059	0.111	0.215	0.077	0.166	0.201
	150	0.047	0.079	0.148	0.061	0.121	0.145
	200	0.041	0.072	0.129	0.054	0.108	0.130

the computation time is defined as the time required by the optimization algorithm to converge. Each optimization is performed by the function constrOptim.nl of the package "alabama" of the free software environment R.

Figures 1a and 1b display, as a function of N, the mean computation time of each estimator and their ratio, respectively. For both estimators, the computation time is nearly the same for each model, the longest computation

Table 9: Empirical RMSE of the MLE and the WLE for Model 4.

		RMSE					
		$\nu = 1$			$\nu = 2$		
	N	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	50	0.056	0.112	0.210	0.065	0.140	0.202
	100	0.038	0.067	0.142	0.044	0.085	0.144
	150	0.030	0.053	0.118	0.033	0.068	0.115
	200	0.026	0.044	0.100	0.030	0.059	0.101
WLE	50	0.075	0.197	0.840	0.069	0.377	0.459
	100	0.045	0.138	0.456	0.046	0.272	0.317
	150	0.034	0.116	0.323	0.034	0.213	0.257
	200	0.028	0.100	0.259	0.030	0.185	0.220

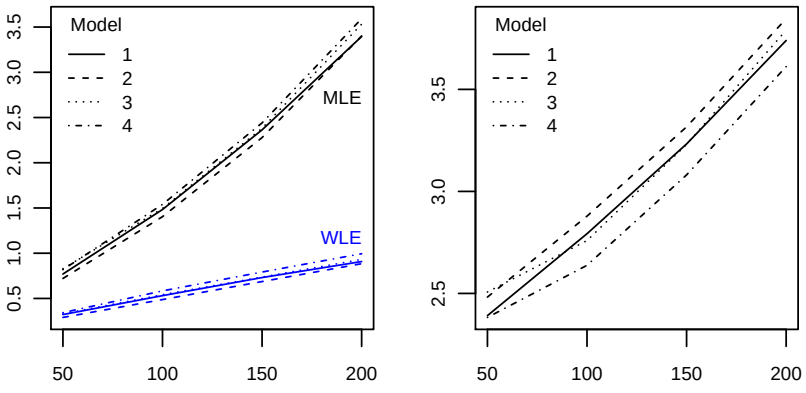
time being obtained for Model 4. The computation time is longer for the MLE than the WLE. This is certainly because the MLE of the white noise variances σ_ν^2 for $\nu = 1, \ldots, \mathcal{S}$ are obtained by minimizing $\hat{\mathcal{L}}_N(\varphi)$, while their WLE are obtained by calculation and do not require any numerical optimization. As expected, the computation time increases monotonously with N, but the slope is much more important for the MLE than the WLE. For instance for Model 1, the ratio of the mean computation times of the MLE and the WLE is 239 when $N = 50$ whereas it is 374 when $N = 200$. Therefore, the larger the sample size is, the greater the benefit of the WLE is. Now, for small sample sizes where the computation time of the MLE is reasonable, this is preferable to use the MLE, especially for models with parameters close to noncausality or noninvertibility regions. Hauser (1999) comes to the same conclusion for the estimation of ARMA models.

In this Monte Carlo study we have taken $\mathcal{S} = 2$ to limit the number of parameters to estimate. However, it is worth noting that the difference between the computation time of the MLE and the WLE increases with $\mathcal{S}$, and in practice, the calculation of the MLE may become impracticable. For example, this may be the case in the context of automatic model selection through information criteria like Akaike and Schwarz criteria.

5 Application

We analyze the daily mean concentrations of sulfur dioxide (SO_2) observed from January 1, 2005 to December 31, 2009 at the monitoring station of environment and water resources state institute located in Vitória, Espírito Santo, Brazil. Figure 2 displays the data.

Since the data was measured daily, a PARMA model, with period $\mathcal{S} = 7$, seems to be appropriated. We fit a PARMA model to the mean-corrected data obtained by subtracting the sample periodic mean from the original data. The



(a) (b)

Fig. 1: (a) Mean computation time in seconds of the MLE and the WLE as a function of N. (b) Ratio of the mean computation times of the MLE and the WLE as a function of N.

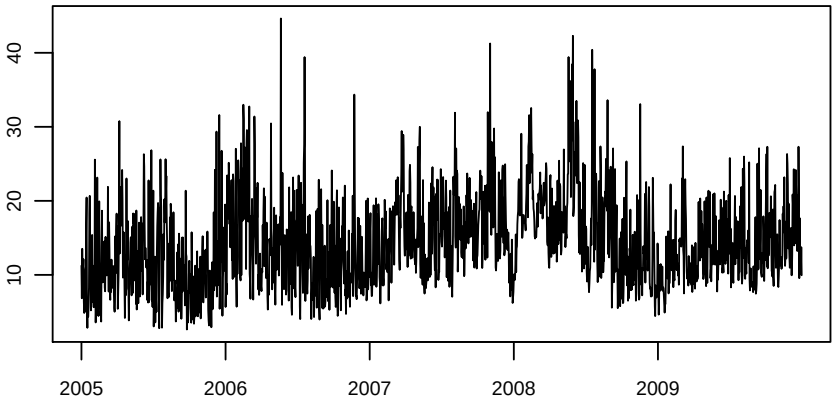


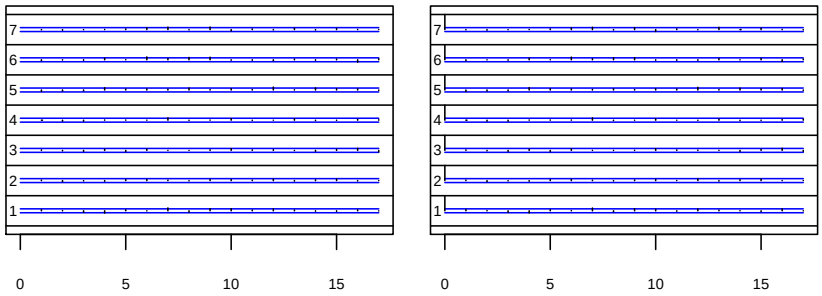
Fig. 2: Daily mean concentrations of SO_2 in Vitória, ES, Brazil.

sample periodic autocorrelation and partial autocorrelation functions suggest the ARMA orders $p_\nu = 1, 1, 1, 1, 1, 1, 1$ and $q_\nu = 1, 1, 1, 1, 0, 0, 1$ (observe that this model satisfies **(A1)**). We set the initial AR and MA parameters as zero. The initial values for the white noise variances $(\sigma_1^2, \ldots, \sigma_{\mathcal{S}}^2)$ has an impact on the computation time of the MLE, while this is not the case for the WLE. We consider the initial values $(1, \ldots, 1)$, $(\hat{\sigma}_X^2, \ldots, \hat{\sigma}_X^2)$ and $(\hat{\sigma}_{X,1}^2, \ldots, \hat{\sigma}_{X,\mathcal{S}}^2)$ where $\hat{\sigma}_X^2$ is the empirical variance of the data (X_t) for $t = 1, \ldots, N\mathcal{S}$ and $\hat{\sigma}_{X,\nu}^2$ is the empirical variance of the data $(X_{n\mathcal{S}+\nu})$ for $n = 0, \ldots, N-1$. Then, for the MLE, the computation time is 381.9 seconds, 261.1 seconds and 148.9

seconds, respectively, while for the WLE, the computation time is 2.9 seconds for all initial values. These different initial values do not have influence on the values of the MLE, even for the estimate of σ_ν^2. Therefore, the WLE is at least 50 times faster than the MLE. This huge difference discourages the use of the MLE in a repetitive context such as automatic model selection through information criteria like Akaike and Schwarz criteria. The estimates obtained by both methods are presented in Table 10 and their values are very similar. The sample autocorrelation functions of the residuals for each season plotted in Figure 3 confirm that these residuals are uncorrelated.

Table 10: Fitted PARMA model to SO_2 data.

	MLE			WLE		
ν	$\phi_{\nu,1}$	$\theta_{\nu,1}$	σ_ν^2	$\phi_{\nu,1}$	$\theta_{\nu,1}$	σ_ν^2
1	-0.72	-0.49	28.97	-0.72	-0.48	28.95
2	-1.14	-0.75	28.38	-1.13	-0.74	28.41
3	-0.80	-0.54	23.49	-0.80	-0.54	23.49
4	-0.89	-0.50	19.56	-0.89	-0.50	19.57
5	-0.58	—	25.93	-0.58	—	25.94
6	-0.61	—	32.85	-0.61	—	32.85
7	-0.69	-0.36	32.40	-0.70	-0.36	32.40



(a) MLE (b) WLE

Fig. 3: Sample autocorrelation functions of the residuals for each season.

6 Proofs

Proof of Lemma 2.1. Since φ is uniquely obtained from η, it is sufficient to prove that η is uniquely defined by η^* under **(A1)** or **(A2)**. Since the product $\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$ is unit lower triangular and Σ is diagonal, (2.8) is the Cholesky

decomposition of Σ^*. Therefore, Σ and the product $\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$ are uniquely obtained from the Cholesky decomposition of Σ^*. We shall prove that $\boldsymbol{\phi}_0^{-1}$ can be uniquely determined from $(\boldsymbol{\phi}_1^*, \ldots, \boldsymbol{\phi}_P^*)$ when **(A1)** holds. Then $\boldsymbol{\theta}_0$ is obtained from $\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$ and it follows from (2.7) that for every positive integer k, $\boldsymbol{\phi}_k$ and $\boldsymbol{\theta}_k$ are uniquely determined from $\boldsymbol{\phi}_k^*$ and $\boldsymbol{\theta}_k^*$ by the relations $\boldsymbol{\phi}_k = \boldsymbol{\phi}_0\boldsymbol{\phi}_k^*$ and $\boldsymbol{\theta}_k = \boldsymbol{\phi}_0\boldsymbol{\theta}_k^*\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$. In the proof we distinguish the cases where $p = \mathcal{S}$, $p > \mathcal{S}$ and $p < \mathcal{S}$. In a similar way, when **(A2)** holds, it can be shown by distinguishing the cases $q = S$, $q > S$ and $q < S$ that $\boldsymbol{\theta}_0^{-1}$ can be uniquely determined from $(\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0\boldsymbol{\theta}_1^*\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0\boldsymbol{\theta}_Q^*\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0)$ (this proof is omitted). Then $\boldsymbol{\phi}_0$ is obtained from $\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$ and the matrices $\boldsymbol{\phi}_k$ and $\boldsymbol{\theta}_k$ are uniquely determined from $\boldsymbol{\phi}_k^*$ and $\boldsymbol{\theta}_k^*$ as above.

Case $p = \mathcal{S}$. Then the AR order $P = \lceil p/\mathcal{S} \rceil$ of the VARMA representation is equal to 1, and it follows from (2.4) that

$$\boldsymbol{\phi}_0 = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ \phi_{2,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{\mathcal{S},\mathcal{S}-1} & \phi_{\mathcal{S},\mathcal{S}-2} & \cdots & 1 \end{bmatrix}, \boldsymbol{\phi}_1 = \begin{bmatrix} \phi_{1,\mathcal{S}} & \phi_{1,\mathcal{S}-1} & \cdots & \phi_{1,1} \\ 0 & \phi_{2,\mathcal{S}} & \cdots & \phi_{2,2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{\mathcal{S},\mathcal{S}} \end{bmatrix}. \tag{6.1}$$

Setting $\mathbf{L} = \boldsymbol{\phi}_0^{-1}$ and $\mathbf{U} = \boldsymbol{\phi}_1$, we have $\boldsymbol{\phi}_1^* = \mathbf{LU}$ which is a LU decomposition of $\boldsymbol{\phi}_1^*$ since $\mathbf{L}$ is lower triangular with unit diagonal and $\mathbf{U}$ is upper triangular. It follows from **(A1)** that the diagonal elements of $\mathbf{U}$ are nonzero. Then $\boldsymbol{\phi}_1^*$ is nonsingular and the LU decomposition is unique, see e.g. Golub and Van Loan (2013, Theorem 3.2.1). This implies that $\boldsymbol{\phi}_0^{-1}$ is uniquely determined from $\boldsymbol{\phi}_1^*$.

Case $p > \mathcal{S}$. Then $P > 1$. If $p/\mathcal{S}$ is an integer, we have $p = P\mathcal{S}$ and

$$\boldsymbol{\phi}_P = \begin{bmatrix} \phi_{1,p} & \phi_{1,p-1} & \cdots & \phi_{1,p-\mathcal{S}+1} \\ 0 & \phi_{2,p} & \cdots & \phi_{2,p-\mathcal{S}+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{\mathcal{S},p} \end{bmatrix}. \tag{6.2}$$

In this case, we define $\mathbf{U}$ by $\mathbf{U} = \boldsymbol{\phi}_P$. If $p/\mathcal{S}$ is not an integer, we have $P - 1 < p/\mathcal{S} < P$. Setting $\kappa = (P-1)\mathcal{S}$, we have

$$\boldsymbol{\phi}_P = \begin{bmatrix} 0 \cdots & \phi_{1,p} & \phi_{1,p-1} & \cdots & \phi_{1,\kappa+1} \\ 0 \cdots & 0 & \phi_{2,p} & \cdots & \phi_{2,\kappa+2} \\ \vdots \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 \cdots & 0 & 0 & \cdots & \phi_{p-\kappa,p} \\ 0 \cdots & 0 & 0 & \cdots & 0 \\ \vdots \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{6.3}$$

and

$$\boldsymbol{\phi}_{P-1} = \begin{bmatrix} \phi_{1,\kappa} & \cdots & \phi_{1,p-\mathcal{S}+1} & \cdots & \phi_{1,\kappa-\mathcal{S}+1} \\ \phi_{2,\kappa+1} & \cdots & \phi_{2,p-\mathcal{S}+2} & \cdots & \phi_{2,\kappa-\mathcal{S}+2} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \phi_{p-\kappa,p-1} & \cdots & \phi_{p-\kappa,2p-P\mathcal{S}} & \cdots & \phi_{p-\kappa,p-\mathcal{S}} \\ \phi_{p-\kappa+1,p} & \cdots & \phi_{p-\kappa+1,2p-P\mathcal{S}+1} & \cdots & \phi_{p-\kappa+1,p-\mathcal{S}+1} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_{\mathcal{S},p} & \cdots & \phi_{\mathcal{S},\kappa} \end{bmatrix}. \tag{6.4}$$

We define $\mathbf{U}$ as the $\mathcal{S} \times \mathcal{S}$ matrix formed by the concatenation of the last $p - \kappa$ columns of $\boldsymbol{\phi}_P$ and the first $\kappa - p + \mathcal{S}$ columns of $\boldsymbol{\phi}_{P-1}$. We see that $\mathbf{U}$ coincides with the right hand side of (6.2). Therefore, the expression of $\mathbf{U}$ is the same for all $p > \mathcal{S}$ and $\mathbf{U}$ is upper triangular. Now, we define $\mathbf{U}^*$ exactly as $\mathbf{U}$ by replacing $\boldsymbol{\phi}_P$ and $\boldsymbol{\phi}_{P-1}$ by $\boldsymbol{\phi}_P^*$ and $\boldsymbol{\phi}_{P-1}^*$, respectively. Since $\boldsymbol{\phi}_P^* = \boldsymbol{\phi}_0^{-1}\boldsymbol{\phi}_P$ and $\boldsymbol{\phi}_{P-1}^* = \boldsymbol{\phi}_0^{-1}\boldsymbol{\phi}_{P-1}$, we have $\mathbf{U}^* = \boldsymbol{\phi}_0^{-1}\mathbf{U}$. Setting $\mathbf{L} = \boldsymbol{\phi}_0^{-1}$, we see that $\mathbf{U}^* = \mathbf{L}\mathbf{U}$ is a LU decomposition of $\mathbf{U}^*$. According to **(A1)**, the diagonal elements of $\mathbf{U}$ are nonzero. Then $\mathbf{U}^*$ is nonsingular, the LU decomposition is unique and $\boldsymbol{\phi}_0^{-1}$ is uniquely determined from $\mathbf{U}^*$.

Case $p < \mathcal{S}$. Then $P = 1$. To simplify the notations, let $\mathbf{L} = \boldsymbol{\phi}_0^{-1}$, $\mathbf{U} = \boldsymbol{\phi}_1$ and $\mathcal{M} = \boldsymbol{\phi}_1^* = \mathbf{L}\mathbf{U}$. If $p = 0$, $\boldsymbol{\phi}_0 = \mathbf{I}$. We assume that $p > 0$ and we partition the matrices as follows,

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0}_{p\times\mathcal{S}-p} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} \mathbf{0}_{p\times\mathcal{S}-p} & \mathbf{U}_{12} \\ \mathbf{0}_{\mathcal{S}-p\times\mathcal{S}-p} & \mathbf{0}_{\mathcal{S}-p\times p} \end{bmatrix}$$

and

$$\mathcal{M} = \begin{bmatrix} \mathbf{0}_{p\times\mathcal{S}-p} & \mathcal{M}_{12} \\ \mathbf{0}_{\mathcal{S}-p\times\mathcal{S}-p} & \mathcal{M}_{22} \end{bmatrix},$$

where the unit lower triangular matrices $\mathbf{L}_{11}$ and $\mathbf{L}_{22}$ have dimensions $p \times p$ and $\mathcal{S} - p \times \mathcal{S} - p$, respectively, and the $p \times p$ matrix $\mathbf{U}_{12}$ is

$$\mathbf{U}_{12} = \begin{bmatrix} \phi_{1,p} & \phi_{1,p-1} & \cdots & \phi_{1,1} \\ 0 & \phi_{2,p} & \cdots & \phi_{2,2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{p,p} \end{bmatrix}.$$

We have $\mathcal{M}_{12} = \mathbf{L}_{11}\mathbf{U}_{12}$ where $\mathbf{U}_{12}$ is upper triangular and all diagonal elements of $\mathbf{U}_{12}$ are nonzero according to **(A1)**. Then $\mathbf{L}_{11}\mathbf{U}_{12}$ is the unique LU decomposition of $\mathcal{M}_{12}$. Since $\mathcal{M}_{22} = \mathbf{L}_{21}\mathbf{U}_{12}$, $\mathbf{L}_{21} = \mathcal{M}_{22}\mathbf{U}_{12}^{-1}$. Thus $\mathbf{L}_{11}$ and $\mathbf{L}_{21}$ are uniquely determined from $\boldsymbol{\phi}_1^*$. To identify $\boldsymbol{\phi}_0^{-1}$, it remains to determine $\mathbf{L}_{22}$. For this, we shall distinguish the cases where $p = \mathcal{S}/2$, $\mathcal{S}/2 < p < \mathcal{S}$ and $0 < p < \mathcal{S}/2$. We set $\mathbf{F} = \boldsymbol{\phi}_0$.

Assume that $p = \mathcal{S}/2$. Then we can rewrite

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{0}_{p\times p} \\ \mathbf{F}_{21} & \mathbf{F}_{22} \end{bmatrix},$$

where all blocks are $p \times p$ matrices, $\mathbf{F}_{11}$ and $\mathbf{F}_{22}$ are unit lower triangular and

$$\mathbf{F}_{21} = \begin{bmatrix} \phi_{p+1,p} & \phi_{p+1,p-1} & \cdots & \phi_{p+1,1} \\ 0 & \phi_{p+2,p} & \cdots & \phi_{p+2,2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{2p,p} \end{bmatrix}.$$

Since $\mathbf{LF} = \mathbf{I}$, we have $\mathbf{F}_{11} = \mathbf{L}_{11}^{-1}$ and $-\mathbf{L}_{21}\mathbf{F}_{11} = \mathbf{L}_{22}\mathbf{F}_{21}$. Since $\mathbf{F}_{21}$ is upper triangular and invertible by **(A1)**, $\mathbf{L}_{22}\mathbf{F}_{21}$ is the unique LU decomposition of $-\mathbf{L}_{21}\mathbf{L}_{11}^{-1}$, and thus $\mathbf{L}_{22}$ is uniquely determined from $\boldsymbol{\phi}_1^*$.

Consider now the case where $\mathcal{S}/2 < p < \mathcal{S}$. We rewrite

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{I}_{12} & \mathbf{0}_{p\times \mathcal{S}-p} \\ \mathbf{0}_{\mathcal{S}-p\times \mathcal{S}-p} & \mathbf{0}_{\mathcal{S}-p\times 2p-\mathcal{S}} & \mathbf{I}_{23} \end{bmatrix} \quad \text{and} \quad \mathbf{F} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{F}_{12} & \mathbf{0}_{p\times \mathcal{S}-p} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \mathbf{F}_{23} \end{bmatrix},$$

where $[\mathbf{I}_{11}, \mathbf{I}_{12}]$ and $\mathbf{I}_{23}$ are the $p \times p$ and $\mathcal{S} - p \times \mathcal{S} - p$ identity matrices, respectively, $[\mathbf{F}_{11}, \mathbf{F}_{12}]$ and $\mathbf{F}_{23}$ are unit lower triangular matrices with dimensions $p \times p$ and $\mathcal{S} - p \times \mathcal{S} - p$, respectively, and the $\mathcal{S} - p \times \mathcal{S} - p$ matrix $\mathbf{F}_{21}$ is

$$\mathbf{F}_{21} = \begin{bmatrix} \phi_{p+1,p} & \phi_{p+1,p-1} & \cdots & \phi_{p+1,2p-\mathcal{S}+1} \\ 0 & \phi_{p+2,p} & \cdots & \phi_{p+2,2p-\mathcal{S}+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{\mathcal{S},p} \end{bmatrix}.$$

Since $\mathbf{LF} = \mathbf{I}$, $[\mathbf{F}_{11}, \mathbf{F}_{12}] = \mathbf{L}_{11}^{-1}$ and $-\mathbf{L}_{21}\mathbf{F}_{11} = \mathbf{L}_{22}\mathbf{F}_{21}$ where $\mathbf{L}_{22}$ is unit lower triangular and $\mathbf{F}_{21}$ is invertible by **(A1)**. Then $\mathbf{L}_{22}\mathbf{F}_{21}$ is the unique LU decomposition of $-\mathbf{L}_{21}\mathbf{F}_{11}$, and thus $\mathbf{L}_{22}$ is uniquely determined from $\boldsymbol{\phi}_1^*$.

Suppose that $0 < p < \mathcal{S}/2$. Remember that the first p columns of $\mathbf{L}$ are uniquely determined from $\boldsymbol{\phi}_1^*$ and partition the matrices as follows,

$$\mathbf{I} = \begin{bmatrix} \mathbf{I}_{11} & \mathbf{0}_{p\times \mathcal{S}-2p} & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \mathbf{I}_{22} & \mathbf{I}_{23} \\ \mathbf{0}_{\mathcal{S}-2p\times p} & \mathbf{I}_{32} & \mathbf{I}_{33} \end{bmatrix}, \mathbf{L} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0}_{p\times p} & \mathbf{0}_{p\times \mathcal{S}-2p} \\ \mathbf{L}_{21} & \mathbf{L}_{22} & \mathbf{0}_{p\times \mathcal{S}-2p} \\ \mathbf{L}_{31} & \mathbf{L}_{32} & \mathbf{L}_{33} \end{bmatrix},$$

and

$$\mathbf{F} = \begin{bmatrix} \mathbf{F}_{11} & \mathbf{0}_{p\times \mathcal{S}-2p} & \mathbf{0}_{p\times p} \\ \mathbf{F}_{21} & \mathbf{F}_{22} & \mathbf{F}_{23} \\ \mathbf{0}_{\mathcal{S}-2p\times p} & \mathbf{F}_{32} & \mathbf{F}_{33} \end{bmatrix},$$

where $\mathbf{L}_{11}$, $\mathbf{L}_{22}$, $\mathbf{L}_{33}$ and $\mathbf{F}_{11}$ are unit lower triangular matrices with dimensions $p \times p$, $p \times p$, $\mathcal{S} - 2p \times \mathcal{S} - 2p$ and $p \times p$, respectively, the $p \times p$ matrix $\mathbf{F}_{21}$ and the $\mathcal{S} - 2p \times \mathcal{S} - 2p$ matrix $\mathbf{F}_{32}$ are upper triangular and their diagonal elements are nonzero according to **(A1)**. Since $\mathbf{LF} = \mathbf{I}$, we have $\mathbf{F}_{11} = \mathbf{L}_{11}^{-1}$, $-\mathbf{L}_{21}\mathbf{F}_{11} = \mathbf{L}_{22}\mathbf{F}_{21}$ is the unique LU decomposition of $-\mathbf{L}_{21}\mathbf{F}_{11}$

so that $\mathbf{L}_{22}$ and $\mathbf{F}_{21}$ are uniquely determined from $\boldsymbol{\phi}_1^*$, $\mathbf{L}_{32} = -\mathbf{L}_{31}\mathbf{F}_{11}\mathbf{F}_{21}^{-1}$, $\mathbf{F}_{22} = \mathbf{L}_{22}^{-1}\mathbf{I}_{22}$, $\mathbf{I}_{32} - \mathbf{L}_{32}\mathbf{F}_{22} = \mathbf{L}_{33}\mathbf{F}_{32}$ is the unique LU decomposition of $\mathbf{I}_{32} - \mathbf{L}_{32}\mathbf{F}_{22}$ so that $\mathbf{L}_{33}$ is uniquely determined from $\boldsymbol{\phi}_1^*$. Therefore, all the elements of $\mathbf{L}$ are identified in a unique way from $\boldsymbol{\phi}_1^*$.

Proof of Lemma 2.2. According to (2.4), $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$ are unit lower triangular. Then it follows from (2.11) that $\det \Phi(z) = \det \Phi^*(z)$ and $\det \Theta(z) = \det \Theta^*(z)$, which shows that **(A3)** is equivalent to **(A3*)**. Assume **(A4)** and that $\Phi^*(z) = e^*(z)\Phi_1^*(z)$ and $\Theta^*(z) = e^*(z)\Theta_1^*(z)$ where $e^*(z)$, $\Phi_1^*(z)$ and $\Theta_1^*(z)$ are $\mathcal{S} \times \mathcal{S}$ matrices of polynomials. It follows from (2.11) that $\Phi(z) = e(z)\Phi_1(z)$ where $e(z) = \boldsymbol{\phi}_0 e^*(z)$, $\Phi_1(z) = \Phi_1^*(z)$, and $\Theta(z) = e(z)\Theta_1(z)$ where $\Theta_1(z) = \Theta_1^*(z)\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$. According to **(A4)**, $e(z)$ has constant determinant. Since $\det e^*(z) = \det e(z)/\det \boldsymbol{\phi}_0$, $\det e^*(z)$ is constant and **(A4)** implies **(A4*)**. The converse is shown in the same way. To show that **(A5)** and **(A5*)** are equivalent, it is sufficient to show that for any $\mathcal{S} \times \mathcal{S}$ matrix of polynomials $g(z)$ and any constant invertible matrix A, we have $\mathcal{G}(Ag(z)) = A\mathcal{G}(g(z))$. For every $i = 1, \ldots, \mathcal{S}$, let $g_i(z)$ be the ith column of $g(z)$, p_i be the maximum degree of $g_i(z)$ and $g_i(j)$ be the vectors of coefficients of z^j in $g_i(z)$. The ith column of $Ag(z)$ is $Ag_i(z) = \sum_{j=0}^{p_i} Ag_i(j)z^j$. Since A is invertible and $g_i(p_i) \neq 0$, $Ag_i(p_i) \neq 0$. Then p_i is the maximum degree of $Ag_i(z)$ and $Ag_i(p_i)$ is the ith column of $\mathcal{G}(Ag(z))$. Therefore, $\mathcal{G}(Ag(z)) = A\mathcal{G}(g(z))$.

Proof of Theorem 2.1. First, we show that **(A1)** implies that $\operatorname{rank} \mathcal{G}(\Phi(z)) = \mathcal{S}$. If $p = 0$, $P = 0$ and $\Phi(z) = \boldsymbol{\phi}_0 = \mathbf{I}$. Then, $\mathcal{G}(\Phi(z)) = \mathbf{I}$ and $\operatorname{rank} \mathcal{G}(\Phi(z)) = \mathcal{S}$. If $p = \mathcal{S}$, $P = 1$ and $\boldsymbol{\phi}_1$ is given by (6.1). According to **(A1)**, $\phi_{\nu,\mathcal{S}} \neq 0$ for every $\nu = 1, \ldots, \mathcal{S}$. Then $\mathcal{G}(\Phi(z)) = \boldsymbol{\phi}_1$, and since $\boldsymbol{\phi}_1$ is full rank, $\operatorname{rank} \mathcal{G}(\Phi(z)) = \mathcal{S}$. If $p > \mathcal{S}$ and $p/\mathcal{S}$ is an integer, $p = P\mathcal{S}$ and $\boldsymbol{\phi}_P$ is given by (6.2). **(A1)** implies that $\mathcal{G}(\Phi(z)) = \boldsymbol{\phi}_P$, and since $\boldsymbol{\phi}_P$ is full rank, $\operatorname{rank} \mathcal{G}(\Phi(z)) = \mathcal{S}$. If $p > \mathcal{S}$ and $p/\mathcal{S}$ is not an integer, $\boldsymbol{\phi}_P$ and $\boldsymbol{\phi}_{P-1}$ are given by (6.3) and (6.4), respectively. **(A1)** implies that $\mathcal{G}(\Phi(z))$ is the concatenation of the first $\kappa - p + \mathcal{S}$ columns of $\boldsymbol{\phi}_{P-1}$ and the last $p - \kappa$ columns of $\boldsymbol{\phi}_P$, i.e.,

$$\mathcal{G}(\Phi(z)) = \begin{bmatrix} \phi_{1,\kappa} & \cdots & \phi_{1,p-\mathcal{S}+1} & \phi_{1,p} & \cdots & \phi_{1,\kappa+1} \\ \phi_{2,\kappa+1} & \cdots & \phi_{2,p-\mathcal{S}+2} & 0 & \cdots & \phi_{2,\kappa+2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \phi_{p-\kappa,p-1} & \cdots & \phi_{p-\kappa,2p-P\mathcal{S}} & 0 & \cdots & \phi_{p-\kappa,p} \\ \phi_{p-\kappa+1,p} & \cdots & \phi_{p-\kappa+1,2p-P\mathcal{S}+1} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \phi_{\mathcal{S},p} & 0 & \cdots & 0 \end{bmatrix}. \tag{6.5}$$

By permutation of the first $\kappa - p + \mathcal{S}$ columns and the last $p - \kappa$ columns of $\mathcal{G}(\Phi(z))$, we see that $\operatorname{rank} \mathcal{G}(\Phi(z)) = \mathcal{S}$. Lastly, if $0 < p < \mathcal{S}$, $P = 1$, $\Phi(z) = \boldsymbol{\phi}_0 + \boldsymbol{\phi}_1 z$ where $\boldsymbol{\phi}_0$ and $\boldsymbol{\phi}_1$ are given by (6.1). We see that $\boldsymbol{\phi}_1$ has the same structure as (6.3) where $\kappa = 0$, and that $\boldsymbol{\phi}_0$ has the same structure as (6.4) where $\kappa = 0$, $\phi_{\nu,0} = 1$ and $\phi_{\nu,k} = 0$ for every $\nu = 1, \ldots, \mathcal{S}$ and every

negative integer k. Therefore, **(A1)** implies that $\mathcal{G}(\Phi(z))$ is given by (6.5) where $\kappa = 0$, and we conclude as before that $\operatorname{rank} \mathcal{G}(\Phi(z)) = \mathcal{S}$. Now, since $\operatorname{rank} \mathcal{G}(\Phi(z)) \leq \operatorname{rank} \mathcal{G}(\Phi(z), \Theta(z)\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0) \leq \mathcal{S}$, **(A1)** implies **(A5)**. It follows from Lemma 2.2 that the combination of **(A1)**, **(A3)** and **(A4)** is equivalent to the combination of **(A1)**, **(A3*)**, **(A4*)** and **(A5*)**, and the combination of **(A2)**, **(A3)**, **(A4)** and **(A5)** is equivalent to the combination of **(A2)**, **(A3*)**, **(A4*)** and **(A5*)**. In both cases, we deduce respectively from Lemma 2.1 and Deistler et al. (1978) that function f is one-to-one and that the standard VARMA representation (2.6) is identifiable. Then φ is uniquely determined by η^* which is itself uniquely determined by the autocovariance matrix function of $(\mathbf{X}_n)$. This proves Theorem 2.1.

Proof of Theorem 3.1. According to Lemma 2.1, map f defined by $\eta^* = f(\varphi)$ is a bijective continuous map between $\mathcal{P} \times \mathbb{R}^{\mathcal{S}}_{>0}$ and $\mathcal{P}^* = f(\mathcal{P} \times \mathbb{R}^{\mathcal{S}}_{>0})$. Moreover, $\varphi = f^{-1}(\eta^*)$ is obtained by applying Cholesky and LU decompositions. Therefore, f^{-1} is continuous and f is an homeomorphism between $\mathcal{P} \times \mathbb{R}^{\mathcal{S}}_{>0}$ and $\mathcal{P}^*$. The WLE of $\eta_0^* = f(\varphi_0)$ is

$$\tilde{\eta}_N^* = \underset{\eta^* \in \mathcal{P}^*}{\operatorname{argmin}} \tilde{\mathcal{L}}_N^*(\eta^*)$$

where $\tilde{\mathcal{L}}_N^*(\eta^*) = \tilde{\mathcal{L}}_N(f^{-1}(\eta^*))$. According to Lemma 2.2, the parameter space $\mathcal{P}^*$ satisfies assumptions of Deistler et al. (1978, Theorem 4'). Therefore, $\tilde{\eta}_N^*$ converges a.s. to η_0^* as $N \to \infty$. Since $\tilde{\varphi}_N = f^{-1}(\tilde{\eta}_N^*)$, $\tilde{\varphi}_N$ converges a.s. to φ_0 as $N \to \infty$.

Proof of Theorem 3.2. Let T be the set of entries of the matrices $(\boldsymbol{\phi}_1^*, \ldots, \boldsymbol{\phi}_P^*, \boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_Q^*)$ which satisfy **(A3*)**, **(A4*)** and **(A5*)**. According to Deistler et al. (1978, Theorem 3), T is an open set in $\mathbb{R}^{(P+Q)\mathcal{S}^2}$. Let h be the map, $h(\varphi_\phi', \varphi_\theta') = (\boldsymbol{\phi}_1^*, \ldots, \boldsymbol{\phi}_P^*, \boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_Q^*)$ defined by (2.11). Since h is continuous, the set $h^{-1}(T)$ is an open set in $\mathbb{R}^{(p+q)\mathcal{S}}$. But it follows from Lemma 2.2 that $h^{-1}(T)$ is the set of points which satisfy **(A3)**, **(A4)** and **(A5)**. Therefore, the set $\mathcal{P}$ is an open set in $\mathbb{R}^{(p+q)\mathcal{S}}$ and $\mathcal{P} \times \mathbb{R}^{\mathcal{S}}_{>0}$ is an open set in $\mathbb{R}^{(p+q+1)\mathcal{S}}$. All the entries of the spectral matrix $\mathbf{f}(\omega, \varphi)$ in (3.4) are divisions of polynomials functions of the components of φ, and therefore are twice continuously differentiable functions of $\varphi \in \mathcal{P} \times \mathbb{R}^{\mathcal{S}}_{>0}$. Moreover, these second derivatives are continuous functions of $\omega \in [-\pi, \pi]$. Hence, C2.1 in Dunsmuir (1979) is satisfied. As discussed by Dunsmuir (1979), in the VARMA case $\mathbf{f}(\omega; \varphi)$ and $\partial \mathbf{f}(\omega; \varphi)/\partial \varphi_k$ have elements belonging to the Lipschitz class of degree α, Λ_α as defined by Zygmund (2002, page 42), for $\alpha > 1/2$. Therefore, C2.2 and C2.4 in Dunsmuir (1979) are satisfied for the PARMA model (2.2). Finally, it is easy to show that **(A6)** implies C2.3 in Dunsmuir (1979). Therefore, Theorem 3.2 follows from Corollary 2.2 of Dunsmuir (1979).

References

I. V. Basawa and R. B. Lund. Large sample properties of parameter estimates for periodic ARMA models. *J Time Ser Anal*, 22:651–663, 2001.

P. Bloomfield, H. L. Hurd, and R. B. Lund. Periodic correlation in stratospheric ozone data. *J Time Ser Anal*, 15:127–150, 1994.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer Series in Statistics. Springer Science, New York, NY, 2nd edition, 2006.

M. Deistler, W. Dunsmuir, and E. J. Hannan. Vector linear time series models: Corrections and extensions. *Adv Appl Probab*, 10:360–372, 1978.

W. Dunsmuir. A central limit theorem for parameter estimation in stationary vector ttime series and its application to models for a signal observed with noise. *Ann Stat*, 7(3):490–506, 1979.

W. Dunsmuir and E. J. Hannan. Vector linear time series models. *Adv Appl Probab*, 8(2):339–364, 1976.

R. Fox and M. S. Taqqu. Large-sample properties of parameters estimates for strongly dependent stationary gaussian time series. *Ann Stat*, 14(2): 517–532, 1986.

W. A. Gardner and L. E. Franks. Characterization of cyclostationary random signal processes. *IEEE Trans Signal Process*, 21:4–14, 1975.

W. A. Gardner, A. Napolitano, and L. Paura. Cyclostationarity: Half a century of research. *Signal Process*, 86:639–697, 2006.

E. G. Gladyshev. Periodically correlated random sequences. *Sov. Math.*, 2: 385–388, 1961.

G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, fourth edition, 2013.

M. A. Hauser. Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo study. *J Stat Plan Inference*, 80:229–255, 1999.

M. Heymann. *Structure and realization problems in the theory of dynamical systems*. Springer-Verlag, Vienna-New York, 1975.

H. L. Hurd and A. Miamee. *Periodically correlated random sequences: Spectral theory and practice*. Wiley Series in Probability and Statistics. John Wiley & Sons, Hoboken, NJ, 2007.

W. K. Li and Y. V. Hui. An algorithm for the exact likelihood of periodic autoregressive moving average models. In *Commun Stat Simul Comput*, volume 17, pages 1483–94. Taylor and Francis, 1988.

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

G. C. Reinsel. *Elements of Multivariate Time Series Analysis*. Springer Series in Statistics, second edition, 1997.

A. J. Q. Sarnaglia, V. A. Reisen, and C. Lévy-Leduc. Robust estimation of periodic autoregressive processes in the presence of additive outliers. *J Multivar Anal*, 101:2168–2183, 2010.

G. C. Tiao and M. R. Grupe. Hidden periodic autoregressive-moving average models in time series data. *Biometrika*, 67:365–73, 1980.

A. V. Vecchia. Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *J Am Water Resour Assoc*, 21(5): 721–730, 1985.

P. Whittle. The analysis of multiple stationary time series. *J R Stat Soc Ser B*, 15:125–139, 1953.

A. Zygmund. *Trigonometric Series*. Cambridge University Press, 3rd edition, 2002.