Alessandro José Queiroz Sarnaglia

Estimação de séries ARMA periódicas incompletas na presença de observações atípicas

Brasil
Março de 2016

Alessandro José Queiroz Sarnaglia

Estimação de séries ARMA periódicas incompletas na presença de observações atípicas

Universidade Federal de Minas Gerais

Orientador: Valdério Anselmo Reisen

Brasil
Março de 2016

Este trabalho é dedicado à Lucens e Vera.

Agradecimentos

Agradeço aos meus pais, Lucens e Vera, pelos puxões de orelha quando se fizeram necessários, sendo o maior exemplo de caráter para mim. Hoje sou quem eu sou graças a dedicação, carinho, amor e perseverança de vocês. Obrigado por me incentivarem a ter imaginação.

Aos meus irmãos, Thiago e Dayane, por testarem ao extremo a minha paciência, fato que, por sua vez, desenvolveu em mim umas das principais características requeridas de um pesquisador. Amo vocês gordinho e branquela.

À minha namorada, Josi, pelo apoio, amor e carinho que com certeza me deram tranquilidade para desenvolver esta tese e pelas diversas discussões que me fizeram pensar no próprio ser humano que me tornei e virei a ser.

Ao orientador, professor Valdério, pelos conselhos e por, desde a graduação, acreditar em mim e me incentivar a percorrer meu caminho na carreira acadêmica e desenvolver em mim o prazer de fazer ciência.

Ao colaborador, professor Pascal, cuja dedicação ao trabalho desempenhou papel fundamental. Sem essa determinação, o desenvolvimento desta tese teria sido muito mais árduo.

Aos membros da banca examinadora que, através de suas sugestões, aprimoraram substancialmente a qualidade deste trabalho.

Aos amigos de infância e que ainda fazem parte da minha vida. Certamente, os rocks que fizemos ajudaram a aliviar um pouco da pressão que todos sofremos.

Aos amigos de universidade por participar de minha trajetória, desde as discussões, na universidade ou no bar, com respeito a estatística, até o acolhimento em suas casas nos períodos de vai e volta do Espírito Santo. Devo muito a vocês.

Por fim, à todos aqueles que, de uma maneira ou outra, me auxiliam no caminho que percorro nesta vida

Epígrafe...
"A imaginação é mais importante que a ciência,
porque a ciência é limitada,
ao passo que a imaginação abrange o mundo inteiro."
(Albert Einstein)

Resumo

Esta tese propõe estudos teóricos, simulados e aplicados, em processos ARMA Periódicos (PARMA), nos contextos de identificabilidade, robustez e observações faltantes. No que tange ao problema de identificabilidade, suposições são determinadas para garantir unicidade na identificação do processo, teoria que não tem sido ainda explicitamente investigada na literatura de processos PARMA. As condições de identificabilidade sugeridas permitiram o desenvolvimento da teoria assintótica para o estimador de Whittle desse modelo. O estudo de robustez é baseado no domínio da frequência e em regressão M para obter o estimador da matriz de densidade espectral, que é aplicado ao método de Whittle, com o objetivo de propor um estimador robusto do modelo PARMA quando a série é contaminada por *outliers* aditivos ou é gerada por distribuições simétricas de caudas pesadas, como por exemplo, t-student ou exponencial dupla. A Estimação de modelos PARMA com outliers aditivos e observações faltantes é a terceira contribuição desta tese. Nesse contexto, é proposto um estimador que tem propriedades de robustez e pode ser aplicado em series temporais incompletas. A metodologia é baseada no domínio da frequência por meio da convolução entre função espectral e periodograma. Todas as propostas são utilizadas na aplicação à dados reais provenientes da área da poluição do ar. Essas três contribuições científicas estão, respectivamente, apresentadas nesta tese por meio dos seguintes artigos: *Identifiability and Whittle Estimation of Periodic ARMA Models*; *M-regression spectral estimator for periodic ARMA models: a robust method against additive outliers and heavy tail distributions*; e *On the use of classical and M periodograms to fit periodic ARMA models to time series with missing data and additive outliers.*

Palavras-chave: Processos periodicamente estacionários. Modelo PARMA. Estimador de Whittle. Identificabilidade. Robustez. Dados faltantes.

Abstract

This thesis proposes theoretical, simulated and applied studies on Periodic Autoregressive Moving Average (PARMA) processes in the identifiability, robustness and missing data contexts. Regarding the identifiability problem, conditions have been established to ensure the uniqueness of PARMA representation, which has not been explicitly investigated in the literature of PARMA processes yet. The proposed conditions have allowed the development of the asymptotic theory of Whittle estimator of PARMA parameters. The robustness study is based on the frequency domain and M-regression used to obtain an estimator of the spectral density matrix, which is applied in the Whittle's method, aiming to propose a robust estimator of the PARMA model for time series contaminated by additive outliers or generated by symmetric heavy tailed distributions, such as t-student or double exponential. The estimation of PARMA models with additive outliers and missing data is the third contribution of this thesis. In this context, it is proposed a robust estimator which can be applied to fit incomplete time series. The methodology is based on the frequency domain through convolution of spectral density and periodogram. All proposals are used in applications to real air pollution datasets. These three contributions are, respectively, presented in this thesis through the following papers: Identifiability and Whittle Estimation of Periodic ARMA Models; M-regression spectral estimator for periodic ARMA models: a robust method against additive outliers and heavy tail distributions; and On the use of classical and M periodograms to fit periodic ARMA models to time series with missing data and additive outliers.

Keywords: Periodically stationary processes. PARMA model. Whittle estimator. Identifiability. Robustness. Missing data.

Lista de Figuras

Figure 2.1 – Mean computation time in seconds of the MLE and the WLE as a function of N (a); Ratio of the mean computation times of the MLE and the WLE as a function of N (b). 30
Figure 2.2 – Daily mean concentrations of SO_2 in Vitória, ES, Brazil. 30
Figure 2.3 – Sample autocorrelation function of the WLE residuals. 32
Figure 2.4 – Out of sample SO_2 data and their one-step-ahead predictors. 32
Figure 3.1 – Box-plot of the estimates of the MLE, the WLE and the RWLE of Model 3 for $n = 300$. On each subfigure the left and the right boxes represent, respectively, the uncontaminated and the contaminated time series scenarios. 49
Figure 3.2 – Daily mean concentrations of CO in Ibes, Vila Velha, ES, Brazil. . . . 49
Figure 3.3 – Forecast of the discarded PM_{10} concentrations. 50
Figure 4.1 – Box-plot of $\hat{\varphi}$ (Cla) and $\hat{\varphi}_{\boldsymbol{\rho}}$ (Rob) estimates of Model 3 for $n = 300$. Cases: (i) no outlier and complete; (ii) uncontaminated and missing; and (iii) outlier and incomplete time series. 68
Figure 4.2 – Evolution of SO_2 daily mean concentrations in Vitória. 69

Lista de Tabelas

Table 2.1 – PARMA$(1,1)_2$ models. 27
Table 2.2 – Empirical bias of the MLE and the WLE. 28
Table 2.3 – Empirical RMSE of the MLE and the WLE. 29
Table 2.4 – Fitted PARMA model to SO_2 data. 31
Table 2.5 – Empirical RMSE of the one-step-ahead predictor. 33
Table 3.1 – Parameters. 45
Table 3.2 – RMSE of Model 1 with $\vartheta_\phi = 0.1$ and $\vartheta_\theta = 0.0$. 46
Table 3.3 – RMSE of Model 2 with $\vartheta_\phi = 0.1$ and $\vartheta_\theta = 0.1$. 46
Table 3.4 – RMSE of Model 3 with $\vartheta_\phi = 0.5$ and $\vartheta_\theta = 0.0$. 47
Table 3.5 – RMSE of Model 4 with $\vartheta_\phi = 0.5$ and $\vartheta_\theta = 0.1$. 47
Table 3.6 – WLE and RWLE (with a "$*$") estimates. 50
Table 4.1 – Simulated models. 64
Table 4.2 – RMSE of $\hat{\varphi}$ and $\hat{\varphi}_{\boldsymbol{\rho}}$, with "$*$", for Model 1 with $\vartheta_\phi = 0.1$ and $\vartheta_\theta = 0.0$. 65
Table 4.3 – RMSE of $\hat{\varphi}$ and $\hat{\varphi}_{\boldsymbol{\rho}}$, with "$*$", for Model 2 with $\vartheta_\phi = 0.1$ and $\vartheta_\theta = 0.1$. 65
Table 4.4 – RMSE of $\hat{\varphi}$ and $\hat{\varphi}_{\boldsymbol{\rho}}$, with "$*$", for Model 3 with $\vartheta_\phi = 0.5$ and $\vartheta_\theta = 0.0$. 66
Table 4.5 – RMSE of $\hat{\varphi}$ and $\hat{\varphi}_{\boldsymbol{\rho}}$, with "$*$", for Model 4 with $\vartheta_\phi = 0.5$ and $\vartheta_\theta = 0.1$. 66
Table 4.6 – Model estimates for complete and incomplete SO_2 daily mean concentrations. 69

Sumário

1 INTRODUÇÃO 11

2 IDENTIFIABILITY AND WHITTLE ESTIMATION OF PERIODIC ARMA MODELS 15

2.1 Introduction 15

2.2 Model description 16

2.2.1 Identifiability of the PARMA model 20

2.3 Whittle estimation 21

2.4 Monte Carlo study 26

2.4.1 Bias 27

2.4.2 Root mean squared error 28

2.4.3 Computation time 28

2.5 Application 30

2.6 Proofs 32

2.6.1 Proof of Lemma 2.1 32

2.6.2 Proof of Theorem 2.1 36

2.6.3 Proof of Theorem 2.2 36

2.6.4 Proof of Theorem 2.3 37

3 M-REGRESSION SPECTRAL ESTIMATOR FOR PERIODIC ARMA MODELS 38

3.1 Introduction 38

3.2 PARMA model with additive outliers 40

3.3 Robust Whittle M-estimator 42

3.3.1 Robust Fourier Transform 43

3.3.2 Whittle M-estimator of PARMA parameters 44

3.4 Monte Carlo study 45

3.5 Application 48

4 FITTING PARMA MODELS TO TIME SERIES WITH MISSING DATA AND ADDITIVE OUTLIERS 51

4.1 Introduction 51

4.2 PARMA model with additive outliers 53

4.3 APS processes and missing data 55

4.4 PARMA estimation with missing data 59

4.4.1 Asymptotic Whittle likelihood estimator 59

4.4.1.1 Impact of additive outliers 60
4.4.2 M-asymptotic Whittle likelihood estimator 62
4.5 Monte Carlo study 64
4.6 Application 67
4.7 Proofs 70
4.7.1 Proof of Theorem 4.1 70
4.7.2 Proof of Theorem 4.2 72

5 CONCLUSÃO 73

REFERÊNCIAS BIBLIOGRÁFICAS 74

1 Introdução

Dados coletados no tempo usualmente violam a suposição de independência, que é uma das principais condições dos procedimentos estatísticos mais básicos. Nesse contexto, a análise de séries temporais se torna uma metodologia alternativa apropriada para realizar inferência em dados correlacionados no tempo. Formalmente, uma série temporal representa uma realização finita de um processo estocástico que pode ser indexado pelo tempo, espaço ou ambos. No que tange a séries temporais, em geral, a estacionariedade de segunda ordem, ou fraca, é um dos requisitos básicos. Essa condição estabelece que os momentos de primeira e segunda ordens não dependem do tempo. Uma suposição mais forte é a de estacionariedade estrita, a qual impõe que todas as distribuições conjuntas finito-dimensionais do processo sejam invariantes sob translações do tempo. Pode-se mostrar que estacionariedade estrita aliada a momentos de segunda ordem finitos implica a de segunda ordem, enquanto o oposto não necessariamente é válido. Nesta tese, o termo estacionariedade se referirá a estacionariedade de segunda ordem. Ver Brockwell & Davis (2006) e Priestley (1981), para mais informações acerca de processos estacionários.

A suposição de estacionariedade nem sempre é apropriada na prática e pode ser violada de diversas formas, de modo que métodos especiais necessitam ser desenvolvidos para tratar cada fenômeno separadamente. Em particular, processos cuja estrutura de covariância varia no tempo de maneira periódica, ou periodicamente estacionários (PS), têm ganhado atenção especial desde o artigo pioneiro de Gladyshev (1961). Tiao & Grupe (1980) investigaram os efeitos de utilizar modelos estacionários que negligenciam a periodicidade da verdadeira estrutura de covariância, observaram que essa má especificação pode deteriorar o desempenho das predições e mostraram que esse fenômeno não é detectado pelo diagnóstico da qualidade de ajuste padrão. Nesse contexto, alguns autores propuseram métodos para identificar correlação periódica oculta em séries temporais. Veja, por exemplo, Hurd & Gerr (1991) e Vecchia & Ballerini (1991). Evidências da ocorrência de processos PS em situações práticas foram documentadas em Gardner & Franks (1975), Bloomfield, Hurd & Lund (1994), Lund et al. (1995), entre outros.

Em geral, processos estacionários são base para a concepção de modelos para séries temporais PS, onde se admite que os parâmetros variem periodicamente no tempo. O modelo Periódico Autoregressivo (PAR) é um dos mais abordados nesse contexto. A família PAR foi originalmente introduzida por Thomas & Fiering (1962) para o ajuste e a simulação de fluxos de rios, um estudo sistemático das propriedades desses sistemas foi realizado por Troutman (1979) e McLeod (1994) desenvolveu a metodologia de diagnóstico da qualidade de ajuste para esses processos e mostrou várias desvantagens do modelo Periódico Autoregressivo de Médias Móveis (PARMA), tal como a inviabilidade do uso de critérios automáticos de seleção das ordens do modelo, devido, principalmente, à alta

complexidade computacional envolvida. Diante dessa limitação, somente poucos artigos investigaram a família PARMA. Como exemplo, pode-se citar Vecchia (1985) e Li & Hui (1988).

O avanço tecnológico ocorrido a partir do final da década de 90, no entanto, motivou diversos pesquisadores a retomarem o estudo de processos PARMA com função de autocovariância absolutamente somável (memória curta), dedicando especial atenção à métodos de estimação desses modelos. No domínio do tempo, Anderson, Meerschaert & Vecchia (1999) desenvolveram o *innovations algorithm* para a estimação de sistemas PARMA, Anderson & Meerschaert (2005) investigaram as características assintóticas desse estimador, Lund & Basawa (2000) usaram essa metodologia na construção de um procedimento recursivo simples de predições um passo à frente, o qual foi utilizado no desenvolvimento de um algoritmo eficiente para avaliar a verosimilhança exata de séries temporais PARMA gaussianas e Basawa & Lund (2001) estudaram as propriedades assintóticas do estimador de mínimos quadrados ponderados para esses modelos. No domínio da frequência, Sarnaglia, Reisen & Bondon (2015) sugeriram a utilização da metodologia de Whittle para estimar os parâmetros do modelo PARMA. No contexto de memória longa, ou seja, quando a função de autocovariância não é absolutamente somável, Franses & Ooms (1997) e Koopman, Ooms & Carnero (2007) estudaram processos PARMA de Integração Fracionária (PARFIMA).

Em geral, no contexto de estimação, assume-se que a família de modelos considerada é identificável no sentido de Reinsel (1997) ou Deistler, Dunsmuir & Hannan (1978). Em poucas palavras, identificabilidade de uma família de modelos significa que uma estrutura de covariância (ou matriz de densidade espectral) determina um, e somente um, membro dessa família. Portanto, não identificabilidade resulta em uma superfície de verossimilhança com mais de um máximo (BROCKWELL; DAVIS, 2006, página 431). Nesse sentido, a investigação de condições que asseguram a identificabilidade do modelo é um tópico extremamente importante. No que tange a processos ARMA Vetoriais (VARMA), condições suficientes foram introduzidas por Dunsmuir & Hannan (1976) e Deistler, Dunsmuir & Hannan (1978). No entanto, para a família PARMA, esse tópico ainda é inexplorado, sendo que as pesquisas dedicadas a estimação desses processos somente assumem implicitamente que a busca pelas estimativas é restrita à modelos identificáveis. Veja, por exemplo, o último parágrafo da página 652 de Basawa & Lund (2001).

Conjuntos de dados reais frequentemente apresentam observações atípicas, ou *outliers*. Em geral, esses dados aberrantes comprometem os métodos de inferência clássicos. Na literatura, em geral, três tipos de outliers são considerados (DENBY; MARTIN, 1979): outliers de inovação (IO), que afeta todas observações subsequentes; outliers aditivos (AO); e outliers de reposição (RO). Os dois últimos não afetam observações futuras. Ma & Genton (2000) ressaltam que AO e RO têm o mesmo efeito e são muito mais prejudiciais do que IO. Nesse contexto, Fajardo, Reisen & Cribari-Neto (2009) estudaram o efeito do

AO em processos de memória longa e mostraram que, nesse caso, o processo contaminado apresenta a propriedade de perda de memória, isto é, as funções de autocorrelação teórica e amostral convergem para zero com o aumento da magnitude do outlier. No contexto de processos PS, Sarnaglia, Reisen & Lévy-Leduc (2010) chegaram a mesma conclusão. Naturalmente, estimadores clássicos apresentam alta sensibilidade à presença de outliers, devido ao impacto dessas observações atípicas na estrutura de covariância amostral. Por esse motivo, metodologias robustas têm sido sugeridas na literatura.

No domínio do tempo, Ma & Genton (2000) propuseram uma metodologia de estimação da função de autocovariância, baseada no estimador robusto da escala Q_n. Os autores também estudaram a robustez do método proposto contra RO. Baseados nesse estudo, Fajardo, Reisen & Cribari-Neto (2009) propuseram um método robusto para a estimação do parâmetro de diferenciação fracionária de processos ARFIMA. Sarnaglia, Reisen & Lévy-Leduc (2010) estenderam a proposta de Ma & Genton (2000) para realizar a estimação robusta da função de autocovariância periódica e dos parâmetros de modelos PAR. Robustez para modelos PAR também tem sido considerada por Shao (2008).

No domínio da frequência, estimadores robustos da densidade espectral foram introduzidos recentemente como alternativas ao periodograma clássico. O periodograma tem relação com o estimador de mínimos quadrados dos coeficientes do modelo de regressão linear, onde as covariáveis são dadas por senoides (seno e cosseno) avaliados nas frequências harmônicas, veja por exemplo, Priestley (1981). Nesse contexto, diversos autores sugerem, em vez de mínimos quadrados, realizar a estimação do modelo de regressão através do método não linear de estimadores M, o que dá origem aos periodogramas M. Para séries univariadas com memória curta, como, por exemplo, processos ARMA, Li (2008) estuda o caso especial do periodograma de Laplace, onde o método de mínimos quadrados é substituído por mínimo valor absoluto. Extensão dessa metodologia é proposta por Li (2010). No contexto de memória longa, Fajardo et al. (2015) se baseiam nos resultados de Koul (1992) para estudar as propriedades do periodograma M.

No contexto multivariado, o periodograma M ainda permanece inexplorado. Em adição, estimadores robustos para modelos PARMA ainda não foram introduzidas na literatura. Dessa forma, a utilização do periodograma M multivariado em conjunto com a metodologia de Whittle para realizar o ajuste de processos PARMA constitui uma linha de pesquisa interessante.

Em muitas situações a série temporal não pode ser completamente observada. Nesse contexto, os dados faltantes impedem a utilização de técnicas de inferência clássicas. Portanto, vários autores têm desenvolvido metodologias especiais para a análise de séries temporais incompletas. Por exemplo, Metaxoglou & Smith (2007), Drake, Knapik & Leśkow (2014) e Drake, Knapik & Leśkow (2015) sugerem a utilização de algoritmos do tipo *Expectation-Maximization* (EM) para tratar esse problema. Essa abordagem tem a desvantagem de assumir uma distribuição específica (em geral, gaussiana) para os dados.

Alternativamente, uma abordagem promissora é constituída da utilização de processos de amplitude modulada, onde a análise é realizada por meio de uma série temporal alternativa em que as observações faltantes são substituídas por zeros. Por exemplo, para processos de memória curta, Dunsmuir & Robinson (1981a) e Yajima & Nishino (1999) estudaram o comportamento assintótico de diferentes estimadores da função de autocorrelação de processos estacionários com dados faltantes. Baseados na densidade espectral assintótica de processos de amplitude modulada, Dunsmuir & Robinson (1981b) propuseram um estimador de Whittle para coeficientes do modelo ARMA e Dunsmuir & Robinson (1981c) estudaram a distribuição assintótica dessa metodologia. Outras referências que tratam da análise de séries temporais com dados faltantes são Bondon & Bahamonde (2012) que estudam a estimação de modelos autoregressivos condicionalmente heterocedásticos e Efromovich (2014) que propõem uma metodologia não paramétrica de estimação da densidade espectral.

O estudo de séries temporais PS com dados faltantes ainda se encontra em sua infância. Nesse sentido, a literatura é relativamente escassa e as principais contribuições são dadas nos seguintes artigos: Drake, Knapik & Leśkow (2014); Drake, Knapik & Leśkow (2015); e Drake, Leśkow & Garay (2015). Os dois primeiros trabalhos propõem algoritmos do tipo EM para estimar os parâmetros do modelo AR com amplitude modulada por senoides. O último propõe quatro algoritmos para estimação de séries temporais K-dependentes amplitude moduladas por senoides. O caso mais geral, onde os dados são gerados por processos PARMA aparentemente não foi explorado na literatura.

O estudo conduzido nesta tese tem como objetivo apresentar soluções para as lacunas discutidas anteriormente, a saber, identificabilidade, robustez e observações faltantes em processos PARMA. Os resultados obtidos são apresentados em três artigos. O primeiro, intitulado *Identifiability and Whittle Estimation of Periodic ARMA Models* e apresentado no Capítulo 2, introduz condições para identificabilidade de processos PARMA e investiga as propriedades assintóticas do estimador de Whittle para esse modelo. O Capítulo 3 apresenta o artigo *M-regression spectral estimator for periodic ARMA models: a robust method against additive outliers and heavy tail distributions*. Esse artigo aborda o tema de robustez em séries temporais PARMA no domínio da frequência, propondo a utilização do periodograma M multivariado na metodologia de Whittle. A terceira contribuição desta tese é apresentada no Capítulo 4 no formato do artigo *On the use of classical and M periodograms to fit periodic ARMA models to time series with missing data and additive outliers*. Esse artigo investiga de maneira concatenada os temas de robustez e observações faltantes em séries temporais PARMA, propondo dois estimadores no domínio da frequência.

2 Identifiability and Whittle Estimation of Periodic ARMA Models

Abstract

This paper provides verifiable conditions for the identifiability of periodic autoregressive moving average (PARMA) models and proposes the Whittle likelihood estimator (WLE) for the parameters. This estimator is proved to be strongly consistent and asymptotically normal. Monte Carlo simulation results show that the WLE is a very attractive alternative to the gaussian maximum likelihood estimator (MLE) for large data sets since both estimators have similar preciseness while the computational cost of the latter is much larger. The two estimation methods are applied to fit a PARMA model to the sulfur dioxide (SO_2) daily average pollutant concentrations in the city of Vitória (ES), Brazil.

KEYWORDS. Periodic stationarity, PARMA models, identifiability, Whittle estimation, sulfur dioxide.

2.1 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 & 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 & Franks (1975) investigate cyclostationarity in electrical engineering and Bloomfield, Hurd & Lund (1994) study stratospheric ozone data. For recent reviews on PC processes, see e.g. Gardner, Napolitano & Paura (2006) and Hurd & 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 periodic autoregressive model is already well documented in the literature, see e.g. Sarnaglia, Reisen & Lévy-Leduc (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 considering jointly ARMA coefficients, which leads naturally to the PARMA model. However, this model has not been widely used in real applications yet, perhaps, due to the difficulty and high computational cost of the implementation of the standard estimation methods.

The exact Gaussian PARMA likelihood is derived by Li & Hui (1988), but the method requires the Choleski decomposition of a matrix whose dimension is the number of data. This can be a serious handicap for large data sets and Lund & Basawa (2000) propose an efficient algorithm to evaluate the Gaussian likelihood which does not require any matrix inversion.

It is well known that a PARMA model has a vector ARMA (VARMA) representation, see e.g. Basawa & Lund (2001), however a VARMA model needs to satisfy the conditions given by Dunsmuir & Hannan (1976) to be identifiable. These conditions are tacitly assumed in the literature on PARMA models, see e.g. Basawa & Lund (2001). Here, we show that the identifiability conditions of Dunsmuir & 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 & Hannan (1976), Deistler, Dunsmuir & Hannan (1978) and Fox & 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. PC processes and PARMA models are described in Section 2.2 where the identifiability results are also presented. In Section 2.3, the WLE of a PARMA model is introduced and its asymptotic properties are derived. In Section 2.4, 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 2.5. Proofs are deferred to Section 2.6.

2.2 Model description

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 follow a PARMA model 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 $\mathrm{E}(\varepsilon^2_{n\mathcal{S}+\nu}) = \sigma^2_{n\mathcal{S}+\nu} = \sigma^2_\nu$. 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 $\varphi_\sigma = [\sigma^2_1, \ldots, \sigma^2_\mathcal{S}]'$.

In the following, we set

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

for every $\nu = 1, \ldots, \mathcal{S}$, and we refer to (2.2) as the PARMA$(p, q)_\mathcal{S}$ model. We assume without loss of generality 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 time series defined by $\mathbf{X}_n = [X_{n\mathcal{S}+1}, \ldots, X_{n\mathcal{S}+\mathcal{S}}]'$. It is well known (BASAWA; LUND, 2001) 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 $[\Sigma]_{l,l} = \sigma^2_l$ 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

$$[\boldsymbol{\phi}_0]_{l,m} = \begin{cases} 0 & l < m, \\ 1 & l = m, \\ \phi_{l,l-m} & l > m, \end{cases} \qquad [\boldsymbol{\phi}_k]_{l,m} = \phi_{l,k\mathcal{S}+l-m}, \qquad 1 \leq k \leq P, \tag{2.4}$$

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 $\nu = 1, \ldots, \mathcal{S}$,

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

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 $\eta = [\boldsymbol{\phi}_0, \ldots, \boldsymbol{\phi}_P, \boldsymbol{\theta}_0, \ldots, \boldsymbol{\theta}_Q, \Sigma]$. In other words, $\eta = f_1(\varphi)$ where f_1 is an injective function. Since $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$ are both unit lower triangular matrices and therefore are invertible, (2.3) is called "triangular" VARMA representation of $(\mathbf{X}_n)$.

Note that, the VARMA representation of $(\mathbf{X}_n)$ given by (2.3) 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 form

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

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 a function of parameters φ, say $\eta^* = f_2(\varphi)$. However, f_2 is not necessarily injective 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 the bivariate white noise $(\boldsymbol{\xi}_n)$ with covariance matrix

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

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 the same matrix Σ as (X_t) 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). It is easy to 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 the same matrix Σ as (X_t) 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 implies identifiability problems of PARMA models based on representation (2.6). Finding conditions to ensure that map f_2 be injective is, therefore, 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 f_2 is an injective function.

Proof. The proof is given in Subsection 2.6.1 □

Assumptions **(A1)** and **(A2)** are easy to be verified and give sufficient conditions to guarantee that f_2 be injective. However, these conditions may not be necessary for some subclasses of PARMA models 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.

2.2.1 Identifiability of the PARMA model

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 $[\Gamma(\tau)]_{l,m} = \gamma_l(\tau\mathcal{S} + l - m)$ for every $l, m = 1, \ldots, \mathcal{S}$. The causality and invertibility of (X_t) are equivalent respectively to the causality and invertibility of $(\mathbf{X}_n)$. For more details, we refer to Gladyshev (1961) and Hurd & 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 & Davis (2006, Theorem 11.3.1) whenever $\det \Phi^*(z) \neq 0$ for $|z| \leq 1$. Similar arguments jointly with Brockwell & 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 $|z| \leq 1$.

Causality and invertibility properties do not ensure that 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 $(\mathbf{X}_n)$, see e.g. Brockwell & Davis (2006, page 431) and Reinsel (1997, section 2.3). This identifiability problem results in a likelihood surface with more than one maximum. Further restrictions have to be imposed in order to obtain an identifiable model, and these are discussed as follows.

Following Dunsmuir & 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 again 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, Dunsmuir & Hannan (1978), for every $i = 1, \ldots, \mathcal{S}$, let $g_i(z)$ and $h_i(z)$ be the ith column of $g(z)$ and $h(z)$ (respectively), 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 $h_i(z)$. Let

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

We introduce the following additional assumptions:

(A3) $\det \Phi(z) \neq 0$ and $\det \Theta(z) \neq 0$ for $|z| \leq 1$;

(A4) $\Phi(z)$ and $\Theta(z)$ are left coprime;

(A5) $\operatorname{rank} \mathcal{H}(\Phi, \Theta) = \mathcal{S}$.

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

Proof. The proof is given in Subsection 2.6.2 □

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 $|\theta_{1,1}\theta_{2,1} - \theta_{2,2}| < 1$. Then, under this restriction, the standard VARMA representations (2.6) in Examples 2.1 and 2.2 are identifiable in the sense of Deistler, Dunsmuir & Hannan (1978) whereas, as shown above, the corresponding PARMA models (2.2) are not.

Remark 2.2. Theorem 2.1 holds if **(A1)** and **(A2)** are replaced by any condition which guarantees that map f_2 be injective.

2.3 Whittle estimation

We define the parameter space $\mathcal{P} \subset \mathbb{R}^{(p+q)\mathcal{S}}$ as the set of points $[\varphi'_\phi, \varphi'_\theta]'$ which satisfy assumptions **(A1)** or **(A2)**, **(A3)**, **(A4)** and **(A5)**. In addition, we assume that the true parameters $[\varphi'_{\phi_0}, \varphi'_{\theta_0}]' \in \mathcal{P}$. For simplicity, we suppose 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 $\varphi \in \mathcal{P} \times \mathbb{R}^{\mathcal{S}}_{>0}$, 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, $\Gamma_N(\varphi_0) = \operatorname{Cov}(\mathbf{X}, \mathbf{X})$. 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

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

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 & Basawa (2000). However, as illustrated in Section 2.4, this method

may be inappropriate 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, for all $z \in \mathbb{C}$,

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

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

$$\mathbf{f}(\omega, \varphi) = \frac{1}{2\pi} \Phi^{*-1}(e^{-\mathbf{i}\omega}) \Theta^*(e^{-\mathbf{i}\omega}) \Sigma^* \Theta^{*\prime}(e^{\mathbf{i}\omega}) \Phi^{*\prime-1}(e^{\mathbf{i}\omega}). \tag{2.12}$$

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

$$\tilde{\mathcal{L}}_N(\varphi) = \log\det \Sigma^* + N^{-1} \sum_{j=0}^{N-1} \operatorname{tr}[\mathbf{f}^{-1}(\omega_j, \varphi) I(\omega_j)], \tag{2.13}$$

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),

$$\log\det \Sigma^* = \log\det \Sigma = \sum_{l=1}^{\mathcal{S}} \log \sigma_l^2, \tag{2.14}$$

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

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

Then

$$\operatorname{tr}[\mathbf{f}^{-1}(\omega_j, \varphi) I(\omega_j)] = W'(e^{\mathbf{i}\omega_j}) \mathbf{f}^{-1}(\omega_j, \varphi) W(e^{-\mathbf{i}\omega_j}) = 2\pi \sum_{l=1}^{\mathcal{S}} \sigma_l^{-2} \left| \left[\Theta^{-1}(e^{-\mathbf{i}\omega_j}) \Phi(e^{-\mathbf{i}\omega_j}) W(e^{-\mathbf{i}\omega_j}) \right]_l \right|^2, \tag{2.15}$$

and replacing (2.14) and (2.15) in (2.13), we get that

$$\tilde{\mathcal{L}}_N(\varphi) = \sum_{l=1}^{\mathcal{S}} \left[\log \sigma_l^2 + \frac{2\pi}{N\sigma_l^2} \sum_{j=0}^{N-1} \left| \left[\Theta^{-1}(e^{-\mathbf{i}\omega_j}) \Phi(e^{-\mathbf{i}\omega_j}) W(e^{-\mathbf{i}\omega_j}) \right]_l \right|^2 \right]. \tag{2.16}$$

The WLE of φ_0 is

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

For every $l = 1, \ldots, \mathcal{S}$, the minimum of (2.16) with respect to σ_l^2 is

$$\tilde{\sigma}^2_{l,N}(\varphi_\phi, \varphi_\theta) = \frac{2\pi}{N} \sum_{j=0}^{N-1} \left| \left[\Theta^{-1}(e^{-\mathbf{i}\omega_j}) \Phi(e^{-\mathbf{i}\omega_j}) W(e^{-\mathbf{i}\omega_j}) \right]_l \right|^2. \tag{2.17}$$

Replacing (2.17) in (2.16), 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 2.2. 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.

Proof. The proof is given in Subsection 2.6.3 □

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

(A6) For all $t \in \mathbb{Z}$ and $0 \le r, s \le \mathcal{S}-1$,

a) $\mathrm{E}(\varepsilon_t|\mathcal{F}_{t-1}) = 0$ a.s.,

b) $\mathrm{E}(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2$ a.s.,

c) $\mathrm{E}(\varepsilon_t\varepsilon_{t+r}\varepsilon_{t+s}|\mathcal{F}_{t-1}) = \beta_t(r,s) = \beta_{t+\mathcal{S}}(r,s)$ a.s.,

d) $\mathrm{E}(\varepsilon_t^4) < \infty$,

where $\beta_t(r,s)$ is non random and $\mathcal{F}_t$ is the σ-algebra generated by $\{\varepsilon_s; s \le t\}$.

Theorem 2.3. Under assumption **(A6)**, for any $\varphi_0 \in \mathcal{P} \times \mathbb{R}^{\mathcal{S}}_{>0}$, $N^{1/2}(\tilde{\varphi}_N - \varphi_0)$ converges in law as N tends to infinity to a normal distribution with zero mean vector and covariance matrix

$$\sqrt{N}(\tilde{\varphi} - \varphi_0) \rightsquigarrow \mathcal{N}_d(\mathbf{0}, \Upsilon),$$

where $\rightsquigarrow$ stands for convergence in distribution as $N \to \infty$. The covariance matrix Υ is given by

$$\Upsilon = [\Omega^{-1}(2\Omega + \Pi)\Omega^{-1}],$$

where

$$[\Omega]_{l,m} = \frac{1}{2\pi}\int_0^{2\pi} \operatorname{tr}\left(\mathbf{f}_0^{-1}(\omega)\frac{\partial \mathbf{f}_0(\omega)}{\partial \varphi_l}\mathbf{f}_0^{-1}(\omega)\frac{\partial \mathbf{f}_0(\omega)}{\partial \varphi_m}\right) d\omega,$$

$$[\Pi]_{l,m} = \sum_{a,b,c,d=1}^{\mathcal{S}} \mathcal{C}_{abcd}\left[\frac{\partial \Sigma^{*-1}}{\partial \varphi_l}\right]_{ab}\left[\frac{\partial \Sigma^{*-1}}{\partial \varphi_m}\right]_{cd} = \sum_{a,b,c,d=1}^{\mathcal{S}} \mathcal{C}_{abcd}\left[\Sigma^{*-1}\frac{\partial \Sigma^*}{\partial \varphi_l}\Sigma^{*-1}\right]_{ab}\left[\Sigma^{*-1}\frac{\partial \Sigma^*}{\partial \varphi_m}\Sigma^{*-1}\right]_{cd},$$

$\Sigma^* = \boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0\Sigma\boldsymbol{\theta}_0'\boldsymbol{\phi}_0^{-1\prime}$ and $\mathcal{C}_{abcd} = \mathcal{C}\left[[\boldsymbol{\xi}_n]_a, [\boldsymbol{\xi}_n]_b, [\boldsymbol{\xi}_n]_c, [\boldsymbol{\xi}_n]_d\right]$, $a, b, c, d = 1, \ldots, \mathcal{S}$, stand for the fourth cumulants among the elements a, b, c and d of the vector $\boldsymbol{\xi}_n$.

Proof. The proof is given in Subsection 2.6.4 □

Remark 2.3. If either we strengthen the assumption of non correlated to independent white noise similarly to Basawa & Lund (2001) or restrict to gaussian data, the covariance matrix Υ simplificates to a block diagonal matrix. One block is related to AR and MA parameters and the other one is associated to white noise variances, which means that ϕ's and θ's estimates are asymptotically independent from σ^2's. Since the proof of this fact is quite long, we just indicate it below.

We first show that the Π matrix have a very simple form. Note that the cumulant $\mathcal{C}_{abcd}$ has the more concise formula

$$\mathcal{C}_{abcd} = \sum_{j=1}^{\mathcal{S}} [\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0]_{aj}[\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0]_{bj}[\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0]_{cj}[\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0]_{dj}\,\mathrm{E}(\varepsilon_{n\mathcal{S}+j}^4),$$

which gives

$$[\Pi]_{l,m} = \sum_{j=1}^{\mathcal{S}} \frac{\mathrm{E}(\varepsilon_{n\mathcal{S}+j}^4)}{\sigma_j^8} \left[\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0 \frac{\partial \Sigma^*}{\partial \varphi_l} \boldsymbol{\phi}_0' \boldsymbol{\theta}_0^{-1\prime}\right]_{j,j} \left[\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0 \frac{\partial \Sigma^*}{\partial \varphi_m} \boldsymbol{\phi}_0' \boldsymbol{\theta}_0^{-1\prime}\right]_{j,j}.$$

Obviously, if $k > 0$,

$$\left[\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0 \frac{\partial \Sigma^*}{\partial [\boldsymbol{\phi}_k]_{l,m}} \boldsymbol{\phi}_0' \boldsymbol{\theta}_0^{-1\prime}\right]_{j,j} = \left[\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0 \frac{\partial \Sigma^*}{\partial [\boldsymbol{\theta}_k]_{l,m}} \boldsymbol{\phi}_0' \boldsymbol{\theta}_0^{-1\prime}\right]_{j,j} = 0.$$

For $k = 0$, define the $(\mathcal{S} \times \mathcal{S})$ matrix $\mathbf{1}_{l,m}$ by $[\mathbf{1}_{l,m}]_{a,b} = \mathbb{1}_{\{a=l, b=m\}}$ and note that

$$\begin{aligned}\left[\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0 \frac{\partial \Sigma^*}{\partial [\boldsymbol{\phi}_0]_{l,m}} \boldsymbol{\phi}_0' \boldsymbol{\theta}_0^{-1\prime}\right]_{j,j} &= -\left[\boldsymbol{\theta}_0^{-1}\mathbf{1}_{l,m}\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0\Sigma -\right]_{j,j} - \left[\left(\boldsymbol{\theta}_0^{-1}\mathbf{1}_{l,m}\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0\Sigma -\right)'\right]_{j,j} \\ &= -2\sigma_j^2 \left[\boldsymbol{\theta}_0^{-1}\mathbf{1}_{l,m}\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0\right]_{j,j} = -2\sigma_j^2 \left[\boldsymbol{\theta}_0^{-1}\right]_{j,l}\left[\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0\right]_{m,j}. \quad (2.18)\end{aligned}$$

Observe that, by construction, $k = 0$ implies $l > m$ and, because $\boldsymbol{\theta}_0^{-1}$ and $\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$ are both unit lower triangular, $[\boldsymbol{\theta}_0^{-1}]_{j,l} \neq 0$ and $[\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0]_{m,j} \neq 0$ only if $j \geq l$ and $m \geq j$. Therefore (2.18) equals zero, since otherwise j should satisfy $j \geq l > m \geq j$. In a similar fashion, it can be shown that $[\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0 \frac{\partial \Sigma^*}{\partial [\boldsymbol{\theta}_0]_{l,m}} \boldsymbol{\phi}_0' \boldsymbol{\theta}_0^{-1\prime}]_{j,j} = 0$. Finally, it is easy to see that $[\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0 \frac{\partial \Sigma^*}{\partial \sigma_\nu^2} \boldsymbol{\phi}_0' \boldsymbol{\theta}_0^{-1\prime}]_{j,j} = \mathbb{1}_{\{j=\nu\}}$. Therefore, recalling that $\varphi = \left[\varphi_{\boldsymbol{\phi},\boldsymbol{\theta}}', \varphi_{\boldsymbol{\sigma}}'\right]'$, where $\varphi_{\boldsymbol{\phi},\boldsymbol{\theta}} = \left[\varphi_{\boldsymbol{\phi}}', \varphi_{\boldsymbol{\theta}}'\right]'$ we have

$$\Pi = \begin{bmatrix} \mathbf{0}_{(p+q)\mathcal{S}\times(p+q)\mathcal{S}} & \mathbf{0}_{(p+q)\mathcal{S}\times\mathcal{S}} \\ \mathbf{0}_{\mathcal{S}\times(p+q)\mathcal{S}} & \Pi_{\boldsymbol{\sigma}} \end{bmatrix},$$

where $\Pi_{\boldsymbol{\sigma}}$ is the $(\mathcal{S} \times \mathcal{S})$ diagonal matrix with j diagonal element given by $\mathrm{E}(\varepsilon_{n\mathcal{S}+j}^4)/\sigma_j^8$.

We now turn to the investigation of the Ω matrix. For simplicity, we shall drop the ω term when no confusion arises. Note that

$$[\Omega]_{l,m} = [\Omega]_{m,l} = -\frac{1}{2\pi}\int_0^{2\pi} \mathrm{tr}[(\partial \mathbf{f}_0^{-1}/\partial\varphi_l)(\partial \mathbf{f}_0/\partial\varphi_m)]d\omega.$$

We now turn to evaluation of $\partial \mathbf{f}_0^{-1}/\partial[\boldsymbol{\phi}_r]_{s,t}$ and $\partial \mathbf{f}_0/\partial\sigma_\nu^2$. Observe that

$$\begin{aligned}\frac{\partial \mathbf{f}_0^{-1}}{\partial[\boldsymbol{\phi}_r]_{s,t}} &= \left(\frac{\partial \Phi'}{\partial[\boldsymbol{\phi}_r]_{s,t}}\right)\Theta^{-1\prime}\Sigma^{-1}\Theta^{-1}\Phi + \Phi'\Theta^{-1\prime}\Sigma^{-1}\Theta^{-1}\left(\frac{\partial \Phi}{\partial[\boldsymbol{\phi}_r]_{s,t}}\right)\\ &= \left(\frac{\partial \boldsymbol{\phi}_r}{\partial[\boldsymbol{\phi}_r]_{s,t}}e^{-\mathbf{i}\omega r}\right)'\Theta^{-1\prime}\Sigma^{-1}\Theta^{-1}\Phi + \Phi'\Theta^{-1\prime}\Sigma^{-1}\Theta^{-1}\left(\frac{\partial \boldsymbol{\phi}_r}{\partial[\boldsymbol{\phi}_r]_{s,t}}e^{-\mathbf{i}\omega r}\right)\\ &= \mathbf{1}_{t,s}\Theta^{-1\prime}\Sigma^{-1}\Theta^{-1}\Phi e^{\mathbf{i}\omega r} + \Phi'\Theta^{-1\prime}\Sigma^{-1}\Theta^{-1}\mathbf{1}_{s,t}e^{-\mathbf{i}\omega r}\end{aligned}$$

and

$$\frac{\partial \mathbf{f}_0}{\partial\sigma_\nu^2} = \Phi^{-1}\Theta\left(\frac{\partial \Sigma}{\partial\sigma_\nu^2}\right)\Theta'\Phi^{-1\prime} = \Phi^{-1}\Theta\mathbf{1}_{\nu,\nu}\Theta'\Phi^{-1\prime},$$

where A' stands for the conjugate transpose of the complex matrix A. Now,

$$\begin{aligned}\operatorname{tr}\left[\frac{\partial \mathbf{f}_0^{-1}}{\partial[\boldsymbol{\phi}_r]_{s,t}}\frac{\partial \mathbf{f}_0}{\partial\sigma_\nu^2}\right] &= \sum_{a=1}^{S}\left[\frac{\partial \mathbf{f}_0^{-1}}{\partial[\boldsymbol{\phi}_r]_{s,t}}\frac{\partial \mathbf{f}_0}{\partial\sigma_\nu^2}\right]_{a,a}\\ &= e^{\mathbf{i}\omega r}\sum_{a=1}^{S}\left[\mathbf{1}_{t,s}\Theta^{-1\prime}\Sigma^{-1}\mathbf{1}_{\nu,\nu}\Theta'\Phi^{-1\prime}\right]_{a,a} \qquad (2.19)\\ &\quad + e^{-\mathbf{i}\omega r}\sum_{a=1}^{S}\left[\Phi'\Theta^{-1\prime}\Sigma^{-1}\Theta^{-1}\mathbf{1}_{s,t}\Phi^{-1}\Theta\mathbf{1}_{\nu,\nu}\Theta'\Phi^{-1\prime}\right]_{a,a}. \qquad (2.20)\end{aligned}$$

In one hand, we have that

$$\begin{aligned}(2.20) &= e^{-\mathbf{i}\omega r}\sum_{a=1}^{S}[\Phi'\Theta^{-1\prime}\Sigma^{-1}\Theta^{-1}]_{a,s}[\Phi^{-1}\Theta]_{t,\nu}[\Theta'\Phi^{-1\prime}]_{\nu,a}\\ &= e^{-\mathbf{i}\omega r}[\Phi^{-1}\Theta]_{t,\nu}\sum_{a=1}^{S}[\Theta'\Phi^{-1\prime}]_{\nu,a}[\Phi'\Theta^{-1\prime}\Sigma^{-1}\Theta^{-1}]_{a,s}\\ &= e^{-\mathbf{i}\omega r}[\Phi^{-1}\Theta]_{t,\nu}[\Theta'\Phi^{-1\prime}\Phi'\Theta^{-1\prime}\Sigma^{-1}\Theta^{-1}]_{\nu,s} = e^{-\mathbf{i}\omega r}[\Phi^{-1}\Theta]_{t,\nu}[\Sigma^{-1}\Theta^{-1}]_{\nu,s}\\ &= \frac{e^{-\mathbf{i}\omega r}}{\sigma_\nu^2}[\Theta^{-1}]_{\nu,s}[\Phi^{-1}\Theta]_{t,\nu}\end{aligned}$$

and, on the other hand,

$$(2.19) = e^{\mathbf{i}\omega r}[\Theta^{-1\prime}\Sigma^{-1}]_{s,\nu}[\Theta'\Phi^{-1\prime}]_{\nu,t} = \frac{e^{\mathbf{i}\omega r}}{\sigma_\nu^2}[\Theta^{-1\prime}]_{s,\nu}[\Theta'\Phi^{-1\prime}]_{\nu,t} = \overline{(2.20)}$$

where $\overline{a}$ stands for the complex conjugate of the number a. Observe that, by Assumption **(A3)**, $\Theta^{-1}(z) = \sum_{h=0}^{\infty} C_{\Theta^{-1}}(h)z^h$ and $\Phi^{-1}(z)\Theta(z) = \sum_{h=0}^{\infty} C_{\Phi^{-1}\Theta}(h)z^h$, with $C_{\Theta^{-1}}(0) = \boldsymbol{\theta}_0^{-1}$ and $C_{\Phi^{-1}\Theta}(0) = \boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$. Therefore, $(2.20) = \frac{1}{\sigma_\nu^2}\sum_{h=0}^{\infty} C_{\nu,s,t}(h)e^{-\mathbf{i}\omega(h+r)}$ and $(2.19) + (2.20) = \frac{1}{\sigma_\nu^2}\sum_{h=0}^{\infty} C_{\nu,s,t}(h)\cos[\omega(h+r)]$, where $C_{\nu,s,t}(0) = [\boldsymbol{\theta}_0^{-1}]_{\nu,s}[\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0]_{t,\nu}$. Thus

$$-\frac{1}{2\pi}\int_0^{2\pi}\operatorname{tr}\left(\frac{\partial \mathbf{f}_0^{-1}}{\partial[\boldsymbol{\phi}_r]_{s,t}}\frac{\partial \mathbf{f}_0}{\partial\sigma_\nu^2}\right)d\omega = -\frac{1}{2\pi\sigma_\nu^2}\sum_{h=0}^{\infty} C_{\nu,s,t}(h)\left\{\int_0^{2\pi}\cos[\omega(h+r)]d\omega\right\}.$$

The last integral vanishes if $h + r > 0$, such that for $r > 0$ the asymptotic covariance between the estimates of $[\phi_r]_{s,t}$ and σ_ν^2 equals zero. Therefore, the only non trivial case

is for $r = 0$. Note that in this case the only non zero term in the last sum is obtained for $h = 0$, such that

$$-\frac{1}{2\pi}\int_0^{2\pi} \operatorname{tr}\left(\frac{\partial \mathbf{f}_0^{-1}}{\partial [\boldsymbol{\phi}_0]_{s,t}}\frac{\partial \mathbf{f}_0}{\partial \sigma_\nu^2}\right) d\omega = \frac{1}{2\sigma_\nu^2}[\boldsymbol{\theta}_0^{-1}]_{\nu,s}[\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0]_{t,\nu}.$$

Observe that, by construction, $r = 0$ implies $s > t$ and, because $\boldsymbol{\theta}_0^{-1}$ and $\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$ are both unit lower triangular, $[\boldsymbol{\theta}_0^{-1}]_{\nu,s} \neq 0$ and $[\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0]_{t,\nu} \neq 0$ only if $\nu \geq s$ and $t \geq \nu$. Therefore the above equation equals zero, since otherwise ν should satisfy $\nu \geq s > t \geq \nu$. We conclude that the entries of Ω related to $[\boldsymbol{\phi}_r]_{s,t}$ and σ_ν^2 are zero. Similar arguments can be used to prove that the entries relating $[\boldsymbol{\theta}_r]_{s,t}$ and σ_ν^2 are also zero. Hence, writing $\varphi = \left[\varphi'_{\boldsymbol{\phi},\boldsymbol{\theta}}, \varphi'_{\boldsymbol{\sigma}}\right]'$ again, gives

$$\Omega = \begin{bmatrix} \Omega_{\boldsymbol{\phi},\boldsymbol{\theta}} & 0 \\ 0 & \Omega_{\boldsymbol{\sigma}} \end{bmatrix},$$

where the $((p+q)\mathcal{S} \times (p+q)\mathcal{S})$ matrix $\Omega_{\boldsymbol{\phi},\boldsymbol{\theta}}$ has the similar definition such as in Ω in Dunsmuir & Hannan (1976) and $\Omega_{\boldsymbol{\sigma}}$ is $(\mathcal{S} \times \mathcal{S})$. Finally, the asymptotic covariance matrix is given by

$$\Upsilon = \begin{bmatrix} 2\Omega_{\boldsymbol{\phi},\boldsymbol{\theta}}^{-1} & 0 \\ 0 & \Omega_{\boldsymbol{\sigma}}^{-1}(2\Omega_{\boldsymbol{\sigma}} + \Pi_{\boldsymbol{\sigma}})\Omega_{\boldsymbol{\sigma}}^{-1} \end{bmatrix}.$$

2.4 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 & Basawa (2000). In each of the $M = 1000$ replications, a PARMA series with $\mathcal{S} = 2$ and $N = 50,200$ full periods is generated. Intermediate sample sizes were considered, however the results do not change the conclusions, so that they are not displayed here to save space. The bias, the root mean squared error (RMSE) and the computation time of the WLE and the MLE are analyzed.

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

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

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 $\phi_{2,1} \neq \theta_{2,1}$,

$[\Phi(z), \Theta(z)]$ has rank 2 for all $z \in \mathbb{C}$. We have

$$\mathcal{H}(\Phi, \Theta) = \begin{bmatrix} 1 & \theta_{1,1} & 1 & \phi_{1,1} \\ \theta_{2,1} & 0 & \phi_{2,1} & 0 \end{bmatrix},$$

and then, $\operatorname{rank} \mathcal{H}(\Phi, \Theta) = 2$ since necessarily $\phi_{1,1}$ or $\theta_{1,1}$ is nonzero and $\phi_{2,1}$ or $\theta_{2,1}$ is nonzero.

The Monte Carlo experiments are made with the four PARMA$(1,1)_2$ models whose parameters are given in Table 2.1 and whose innovation process (ε_t) is gaussian. These models satisfy **(A1)**, **(A2)**, **(A3)**, **(A4)**, **(A5)**, **(A6)**, and are chosen in order to evaluate the effect caused by closeness of the parameters to noncausality and noninvertibility regions. Model 1 is far from both 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 2.1 – 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

2.4.1 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,

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

Table 2.2 displays the empirical bias of the MLE and the WLE, for models 1, 2, 3 and 4. This table shows that the bias decreases as the sample size increases for both estimators. Furthermore, both estimators overestimate the AR parameters. However, the MLE and the WLE behave differently in the estimation of the MA parameters and the white noise variances. The MLE overestimates the MA parameters, while the WLE underestimates them. The MLE underestimates the white noise variances, and the WLE overestimates them. Closeness to noncausality or noninvertibility regions seems to have no significant effect in the bias of the MLE. However, mainly for the estimation of the MA parameters and the white noise variances, this seems to increase substantially the bias of the WLE and, as expected, the worse results are obtained for Model 4. Although the MLE has the smallest bias, the bias of the WLE is also small, especially for large sample sizes.

Table 2.2 – Empirical bias of the MLE and the WLE.

			Bias					
			$\nu = 1$			$\nu = 2$		
		N	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
Model 1	MLE	50	0.014	0.027	-0.023	0.017	0.036	-0.051
		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
		200	0.005	-0.006	0.018	0.003	-0.011	0.006
Model 2	MLE	50	0.015	0.031	-0.036	0.015	0.035	-0.057
		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
		200	0.009	-0.015	0.082	0.004	-0.051	0.035
Model 3	MLE	50	0.012	0.030	-0.038	0.019	0.040	-0.062
		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
		200	0.007	-0.026	0.039	0.005	-0.043	0.035
Model 4	MLE	50	0.012	0.040	-0.037	0.015	0.049	-0.056
		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
		200	0.006	-0.062	0.142	0.001	-0.120	0.121

2.4.2 Root mean squared error

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

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

Tables 2.3 displays the empirical RMSE of the MLE and the WLE, for models 1, 2, 3 and 4. This table shows that the RMSE decreases as the sample size increases for both estimators. Again, closeness to noncausality or noninvertibility regions seems to have no significant effect in the RMSE of the MLE. In fact, 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. Now, for the estimation of the MA parameters and the white noise variances, the distance to noncausality or noninvertibility regions increases significantly the RMSE of the $\theta_{\nu,j}$ and σ_ν^2 parts of the WLE, and the worse results are obtained for Model 4. However, for large sample sizes, the RMSE of the MLE and the WLE are of the same order of magnitude.

2.4.3 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, the computation

Table 2.3 – Empirical RMSE of the MLE and the WLE.

			RMSE					
			$\nu = 1$			$\nu = 2$		
		N	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
Model 1	MLE	50	0.098	0.156	0.210	0.120	0.178	0.200
		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
		200	0.047	0.072	0.110	0.059	0.090	0.101
Model 2	MLE	50	0.058	0.139	0.205	0.064	0.151	0.202
		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
		200	0.032	0.066	0.190	0.031	0.109	0.123
Model 3	MLE	50	0.080	0.129	0.205	0.111	0.169	0.208
		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
		200	0.041	0.072	0.129	0.054	0.108	0.130
Model 4	MLE	50	0.056	0.112	0.210	0.065	0.140	0.202
		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
		200	0.028	0.100	0.259	0.030	0.185	0.220

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

Figures figure 2.1a and figure 2.1b 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 largest computation time being obtained for Model 4. The computation time is larger 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. Now, for small sample sizes where the computation time of the MLE is reasonable, this should be the preferable estimation method, 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

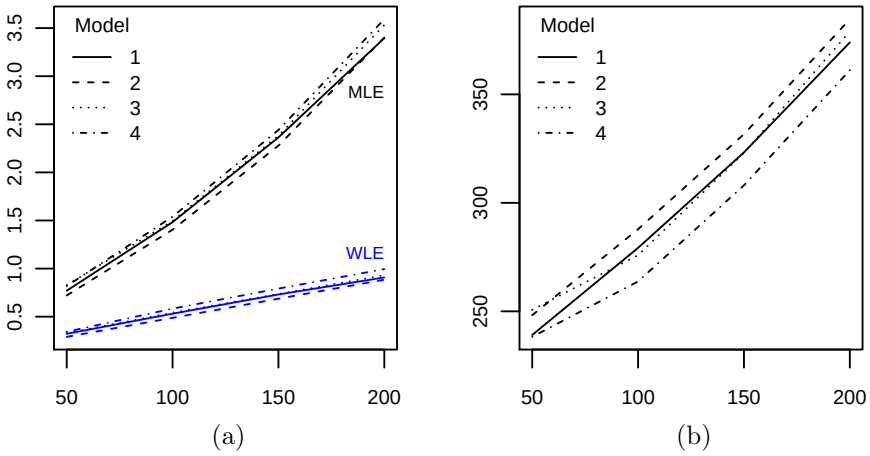


Figure 2.1 – Mean computation time in seconds of the MLE and the WLE as a function of N (a); Ratio of the mean computation times of the MLE and the WLE as a function of N (b).

model selection through information criteria like Akaike and Schwarz criteria.

2.5 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 2.2 displays the data.

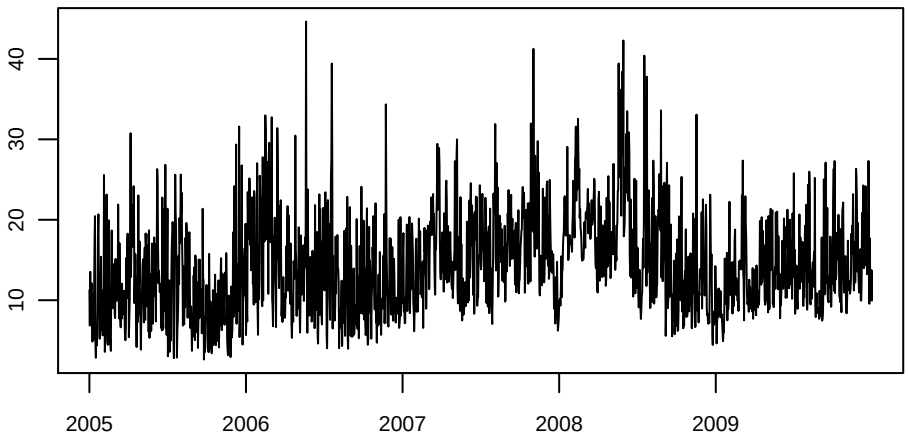


Figure 2.2 – Daily mean concentrations of SO_2 in Vitória, ES, Brazil.

Since one data per day is collected, 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 first $N\mathcal{S} = 1603$ observations are used to fit the model and the last $T = 223$ observations are considered for the out-of-

sample forecast study. The sample periodic autocorrelation and partial autocorrelation functions indicate 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. As it was seen in Section 2.4, the initial values for the white noise variances $(\sigma_1^2, \ldots, \sigma_\mathcal{S}^2)$ have an impact on the computation time of the MLE which is not the case for the WLE. Indeed, for the MLE, taking as initial values $(1, \ldots, 1)$, $(\hat{\sigma}_X^2, \ldots, \hat{\sigma}_X^2)$ and $(\hat{\sigma}_{X,1}^2, \ldots, \hat{\sigma}_{X,\mathcal{S}}^2)$, 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 2.4 and are almost the same. Finally, the sample autocorrelation function of the WLE residuals for each season is plotted in Figure 2.3 and confirms they are uncorrelated. This result is also corroborated by the periodic extension of the Ljung-Box test proposed by McLeod (1994) which presents p-value smaller than 0.05. The MLE presents the same results.

Table 2.4 – 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

We now turn to the forecasting performance. The empirical RMSE is defined by

$$\text{RMSE} = \left(T^{-1} \sum_{t=N\mathcal{S}+1}^{N\mathcal{S}+T} (X_t - \widehat{X}_t)^2 \right)^{1/2}$$

where $\widehat{X}_t$ is the one-step head predictor of X_t. As we see in Table 2.5, the RMSE is the same when $\hat{X}_t$ is calculated from the model fitted by MLE or WLE. Hence, both models have the same predictive performance. Figure 2.4 plots the remaining 233 data and their one-step-ahead forecasts obtained from the model fitted by WLE. Similar results are obtained with the MLE. Visual inspection of this figure shows that the forecasts follow satisfactorily the actual data.

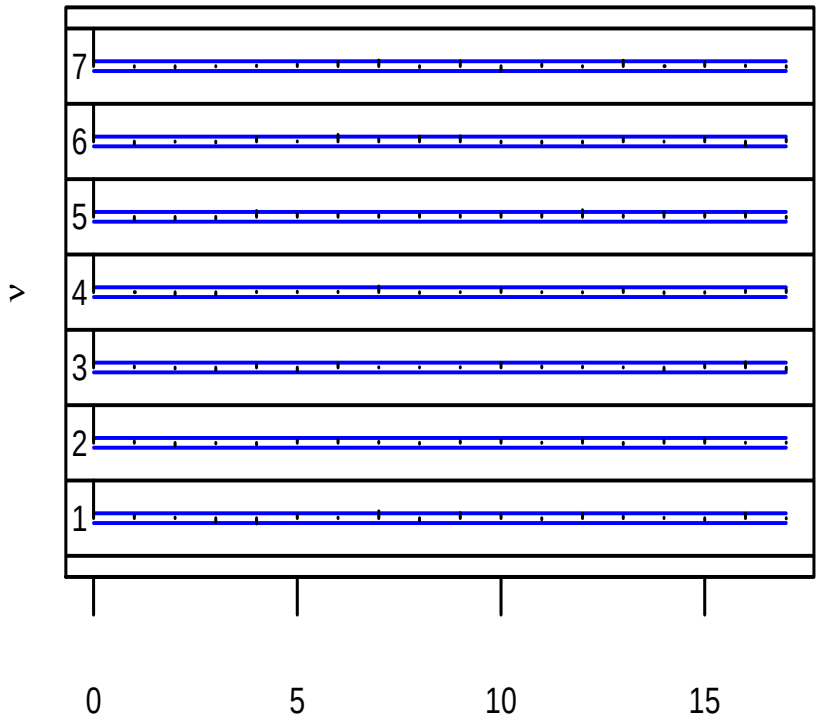


Figure 2.3 – Sample autocorrelation function of the WLE residuals.

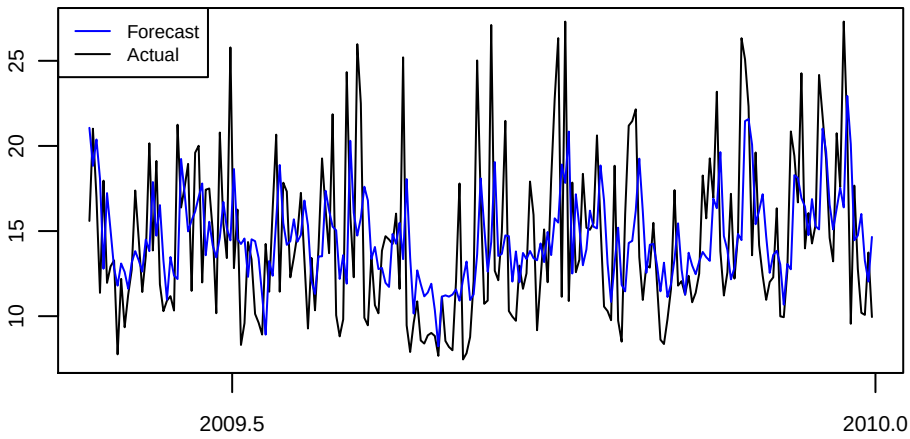


Figure 2.4 – Out of sample SO_2 data and their one-step-ahead predictors.

2.6 Proofs

2.6.1 Proof of Lemma 2.1

Since $\eta = f_1(\varphi)$ where f_1 is injective, 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

Table 2.5 – Empirical RMSE of the one-step-ahead predictor.

	MLE	WLE
In-sample	5.23	5.23
Out-of-sample	4.55	4.55

$\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$ 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,

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

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 & Loan (2012, 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 we define

$$\mathbf{U} = \boldsymbol{\phi}_P = \begin{bmatrix} \phi_{1,p} & \phi_{1,p-1} & \cdots & \phi_{1,p-\mathcal{S}+1} \\ 0 & \phi_{2,p} & \cdots & \phi_{2,p-\mathcal{S}+2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \phi_{\mathcal{S},p} \end{bmatrix}. \tag{2.21}$$

If $p/\mathcal{S}$ is not an integer, we have $P - 1 < p/\mathcal{S} < P$. Setting $\kappa = (P-1)\mathcal{S}$, we have

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

and

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

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 (2.21). 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,

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

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 $\mathbf{LF} = \mathbf{I}_{\mathcal{S}}$ where $\mathbf{I}_{\mathcal{S}}$ is the $\mathcal{S} \times \mathcal{S}$ identity matrix, 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

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

where $[\mathbf{I}_{11}, \mathbf{I}_{12}] = \mathbf{I}_p$, $\mathbf{I}_{23} = \mathbf{I}_{\mathcal{S}-p}$, $[\mathbf{F}_{11}, \mathbf{F}_{12}]$ 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 $\mathbf{LF} = \mathbf{I}_{\mathcal{S}}$, $[\mathbf{F}_{11}, \mathbf{F}_{12}] = \mathbf{L}_{11}^{-1}$ 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,

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

where $\mathbf{I}_{11} = \mathbf{I}_p$, $\mathbf{L}_{11}$, $\mathbf{L}_{22}$, $\mathbf{L}_{33}$ 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 $\mathbf{LF} = \mathbf{I}_{\mathcal{S}}$, 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 $\boldsymbol{\phi}_1^*$.

2.6.2 Proof of Theorem 2.1

We first prove that it is equivalent Φ, Θ or Φ^*, Θ^* satisfying **(A3)**, **(A4)** and **(A5)**. By relations in (2.11), we note that, if **(A3)** holds for Φ, Θ, then

$$\det \Phi^*(z) \det \Theta^*(z) = \frac{\det \Phi(z) \det \Theta(z)}{\det \boldsymbol{\phi}_0 \det \boldsymbol{\theta}_0} \overset{(1)}{\neq} 0, \quad |z| \leq 1,$$

where inequality (1) follows by **(A3)** and invertibility of $\boldsymbol{\phi}_0$ and $\boldsymbol{\theta}_0$. The converse is proven similarly. Now, suppose that Φ, Θ satisfy **(A4)** and that there exists some polynomial matrix $e^*(z)$ such that $\Phi^*(z) = e^*(z)\Phi_1^*(z)$ and $\Theta^*(z) = e^*(z)\Theta_1^*(z)$. We have to show that $\det e^*(z) = c^*$ constant. In fact, by relations in (2.11), we have that

$$\Phi^*(z) = \boldsymbol{\phi}_0^{-1}\Phi(z) = e^*(z)\Phi_1^*(z) \Rightarrow \Phi(z) = e(z)\Phi_1(z)$$

and

$$\Theta^*(z) = \boldsymbol{\phi}_0^{-1}\Theta(z)\boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0 = e^*(z)\Theta_1^*(z) \Rightarrow \Theta(z) = e(z)\Theta_1(z)$$

where $e(z) = \boldsymbol{\phi}_0 e^*(z)$, $\Phi_1(z) = \Phi_1^*(z)$ and $\Theta_1(z) = \Theta_1^*(z)\boldsymbol{\phi}_0^{-1}\boldsymbol{\theta}_0$. Now, by **(A4)**, $c = \det e(z) = \det \boldsymbol{\phi}_0 \det e^*(z)$, which in turn implies that $\det e^*(z) = c^* = c/\det \boldsymbol{\phi}_0$ constant. Therefore, Φ^*, Θ^* are left prime. The converse is shown in the same manner. Finally, assume that Φ, Θ satisfy **(A5)** and let $\mathbf{A}$ be the $2\mathcal{S} \times 2\mathcal{S}$ block diagonal matrix with $\mathcal{S} \times \mathcal{S}$ block entries $\mathbf{A}_{11} = \boldsymbol{\theta}_0^{-1}\boldsymbol{\phi}_0$, $\mathbf{A}_{12} = \mathbf{A}_{21} = \mathbf{0}$ and $\mathbf{A}_{22} = \mathbf{I}$. In addition, from the relations in (2.11), it can be shown that $\mathcal{H}(\Theta^*, \Phi^*) = \boldsymbol{\phi}_0^{-1}\mathcal{H}(\Theta, \Phi)\mathbf{A}$. Therefore,

$$\begin{aligned} \operatorname{rank} \mathcal{H}(\Theta^*, \Phi^*) = \operatorname{rank} \left[\left(\boldsymbol{\phi}_0^{-1}\mathcal{H}(\Theta, \Phi) \right) \mathbf{A} \right] &\overset{(1)}{=} \operatorname{rank} \left(\boldsymbol{\phi}_0^{-1}\mathcal{H}(\Theta, \Phi) \right) \\ &\overset{(2)}{=} \operatorname{rank} \mathcal{H}(\Theta, \Phi) \overset{(3)}{=} \mathcal{S}, \end{aligned}$$

where equalities (1) and (2) hold, respectively, by the following rank properties: if $\operatorname{rank} \mathbf{C}_{n\times k} = n$, then $\operatorname{rank}(\mathbf{BC}) = \operatorname{rank} \mathbf{B}$; and if $\operatorname{rank} \mathbf{D}_{l\times m} = m$, then $\operatorname{rank}(\mathbf{DB}) = \operatorname{rank} \mathbf{B}$. Equality (3) follows by **(A5)**.

Now, by the above equivalencies and the results in Dunsmuir & Hannan (1976) and Deistler, Dunsmuir & Hannan (1978), **(A3)**, **(A4)** and **(A5)** ensure identifiability of the standard VARMA form in (2.6). See also pages 36 and 37 of Reinsel (1997). In addition, by Lemma 2.1, either **(A1)** or **(A2)** guarantee that the standard VARMA representation can be generated by just one PARMA model. Therefore, we conclude that the PARMA model is identifiable.

2.6.3 Proof of Theorem 2.2

Define $\varphi_1^* = \operatorname{vec}[\boldsymbol{\phi}_1^*, \ldots, \boldsymbol{\phi}_P^*, \boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_Q^*]$, φ_2^* as the vector of the elements of and below the diagonal of Σ^* and $\varphi^* = [\varphi_1^{*\prime}, \varphi_2^{*\prime}]'$. By construction of φ^*, relations (2.7) and (2.8) induce continuous constraints in φ^*. In addition, by definition of h_2, Lemma 2.1 ensures that, under **(A1)** and/or **(A2)**, there is a one-to-one continuous function h such that

$\varphi^* = h(\varphi)$, with continuous inverse h^{-1}. Define $\tilde{\mathcal{L}}_N^*(\varphi^*, \mathbf{X}) = \tilde{\mathcal{L}}_N(h^{-1}(\varphi^*), \mathbf{X})$ and $\mathcal{P}^* = h(\mathcal{P} \times \mathbb{R}_{>0}^{\mathcal{S}})$ and let

$$\tilde{\varphi}_N^* = \underset{\varphi^* \in \mathcal{P}^*}{\operatorname{argmin}}\, \tilde{\mathcal{L}}_N^*(\varphi^*, \mathbf{X})$$

be the WLE of $\varphi_0^* \in \mathcal{P}^*$. By Theorem 2.1, the additional Assumptions **(A3)**, **(A4)** and **(A5)** ensure identifiability of the PARMA process. Therefore, in light of Theorems 4 of Dunsmuir & Hannan (1976) and 4' of Deistler, Dunsmuir & Hannan (1978) and their respective remarks, $\tilde{\varphi}_N^*$ is a strongly consistent estimator of φ_0^*. It is not hard to see that, for one-to-one functions, WLE has the so-called invariance property, which ensures that the WLE of φ_0 is given by $\tilde{\varphi}_N = h^{-1}(\tilde{\varphi}_N^*)$. Finally, the continuous map theorem guarantees that

$$\tilde{\varphi}_N = h^{-1}(\tilde{\varphi}_N^*) \xrightarrow{a.s.} h^{-1}(\varphi_0^*) = \varphi_0.$$

2.6.4 Proof of Theorem 2.3

Theorem 3 of Deistler, Dunsmuir & Hannan (1978) can be changed to show that $\mathcal{P} \times \mathbb{R}_{>0}^{\mathcal{S}}$ is open in $\mathbb{R}^{\mathcal{S}(p+q+1)}$. Note that the elements $[\mathbf{f}(\omega, \varphi)]_{l,m}$, $l, m = 1, \ldots, \mathcal{S}$, of the spectral matrix $\mathbf{f}(\omega, \varphi)$ are division of polynomials with respect to the elements of φ and, therefore, are twice continuously differentiable functions of $\varphi \in \mathcal{P} \times \mathbb{R}_{>0}^{\mathcal{S}}$. These second order derivatives being continuous in $\omega \in [-\pi, \pi]$. Hence, C2.1. of Dunsmuir (1979) is satisfied. As discussed in Dunsmuir (1979), in this VARMA case $\mathbf{f}(\omega; \varphi)$ and $\partial \mathbf{f}(\omega; \varphi)/\partial \varphi_j$ have elements belonging to the Lipschitz class of degree α, Λ_α (see page 42 of Zygmund (2002) for the definition), for $\alpha > 1/2$, such that C2.2. and C2.4. of Dunsmuir (1979) are satisfied for the PARMA model. Finally, it can be shown that **(A6)** implies C2.3. of Dunsmuir (1979). Therefore, Corollary 2.2. of Dunsmuir (1979) applies directly to provide the CLT for $N^{1/2}(\tilde{\varphi}_N - \varphi_0)$.

3 M-regression spectral estimator for periodic ARMA models: a robust method against additive outliers and heavy tail distributions

Abstract

This paper proposes a robust approach based on the M-regression method to estimate periodic autoregressive moving average (PARMA) processes. The estimator is based on the frequency domain approach and makes use of the standard Whittle estimator adapted for PARMA models. Empirical studies are addressed to analyse the finite sample size performance of the proposed estimator under the scenarios of contaminated and uncontaminated PARMA processes with additive outliers (AO). The maximum gaussian and Whittle likelihood estimators are also considered in the simulation aiming to show that, under the non-contaminated scenario, the three methods present comparable estimates which indicates that they have similar convergence properties. However, in the case of PARMA series with AO, the latter methods give estimates dramatically biased, which is an unsurprising empirical evidence, while the proposed methodology presents almost unchanged estimates. An application to Carbon Monoxide (CO) concentrations is considered in order to show the usefulness of the proposed method in a real scenario.

KEYWORDS. Periodic stationarity, PARMA models, robust estimation, outliers, Whittle estimation.

3.1 Introduction

Stochastic processes exhibiting Periodic Correlation (PC) are frequently named as periodically correlated, Periodically Stationary (PS) or cyclostationary. Tiao & Grupe (1980) point out that PC may be neglected and misspecified as stationary seasonality if the standard time series tools are used. Since the introduction of PS processes in the literature by Gladyshev (1961), many authors have identified the PC phenomenon in time series of different areas, see e.g. Gardner & Franks (1975) and Bloomfield, Hurd & Lund (1994). Recent reviews on PS processes can be found, for instance, in Gardner, Napolitano & Paura (2006) and Hurd & Miamee (2007).

The standard stationary models, such as, the Autoregressive Moving Average (ARMA) processes, are, in general, the base of the cyclostationary counterparts in which the pa-

rameters vary periodically in time. In this context, the Periodic ARMA (PARMA) framework represents a natural candidate for parsimoniously fitting PS time series. Estimation methods for PARMA models have been investigated in the literature. For example, Lund & Basawa (2000) have considered the Gaussian Maximum Likelihood Estimator (MLE), Basawa & Lund (2001) have studied the least square method and Sarnaglia, Reisen & Bondon (2015) have proposed a Whittle Likelihood Estimator (WLE). All these papers assume tacitly that the process to be estimated is identifiable in the sense of Dunsmuir & Hannan (1976). Conditions to ensure PARMA identifiability have been recently the motivation of the work by Sarnaglia, Reisen & Bondon (2016a). In that paper, the authors have shown empirically that the MLE and WLE have similar good finite sample performance for PARMA time series. However, the behaviour of these methods can be dramatically changed in a scenario wherein atypical observations, or outliers, may occur.

There are several types of outliers which cause different effects on the estimates. However, in general, the following three types are usually considered (DENBY; MARTIN, 1979): innovation outliers (IO), which affect all subsequent observations; additive outliers (AO) or replacement outliers (RO), which have no effect on subsequent observations. AO affect the parameter estimates more than IO, and they have the same effect as RO (MA; GENTON, 2000). In the case of PS processes, the effect of AO in the theoretical and sample autocorrelation functions has been discussed in Sarnaglia, Reisen & Lévy-Leduc (2010). These authors have proposed a robust autocovariance function for PS processes which is used in the periodic Yule-Walker equations to provide robust estimates for Periodic Autoregressive (PAR) models. Shao (2008) has also suggested a robust estimation method for PAR models.

In the frequency domain, robust estimators of the spectral density have been recently introduced as alternatives to the classical periodogram. It is well-known that the periodogram is related to the least square estimator of the coefficients of a linear regression model with sine and cosine regressors, see, for example, Priestley (1981). Alternatively, several authors have defined M-periodogram by using the non-linear method of M-regression, see e.g. Li (2008) and Li (2010). In Fajardo et al. (2015), the authors have studied the M-periodogram for long-memory processes based on the M-regression approach discussed in Koul (1992).

In this paper, we have extended the method proposed by Fajardo et al. (2015) to PARMA models by introducing a multivariate M-periodogram spectral estimator on the Whittle likelihood function given in Sarnaglia, Reisen & Bondon (2016a). The empirical performance of the proposed methodology is evaluated through an extensive Monte Carlo simulation study. The results show very similar behavior of the proposed methodology compared to the MLE and WLE in the uncontaminated scenario. On the other hand, in the contaminated time series with AO scheme, the empirical results show that both estimation methods MLE and WLE are destroyed and, also, the superiority of the robust

method over these two approaches.

The rest of the paper is structured as follows: Section 3.2 describes the PARMA model with AO; Section 3.3 introduces the robust Whittle estimation method; The finite sample performance of the robust estimator is investigated through a Monte Carlo study and the results are discussed in Section 3.4; An application of the methodology to CO daily mean concentrations is the motivation of Section 3.5.

3.2 PARMA model with additive outliers

Let $\mathbb{Z}$ be the set of integer numbers and $(Z_t)_{t\in\mathbb{Z}}$ be a real valued stochastic process satisfying $\mathrm{E}(Z_t^2) < \infty$ for all $t \in \mathbb{Z}$. Let $\mu_{Z,t} = \mathrm{E}(Z_t)$ and $\gamma_{Z,t}(\tau) = \mathrm{Cov}(Z_t, Z_{t-\tau})$. We say that (Z_t) is PS with period $\mathcal{S}$ ($\mathrm{PS}_{\mathcal{S}}$) if, for every $(t,\tau) \in \mathbb{Z}^2$,

$$\mu_{Z,t+\mathcal{S}} = \mu_{Z,t} \quad \text{and} \quad \gamma_{Z,t+\mathcal{S}}(\tau) = \gamma_{Z,t}(\tau), \tag{3.1}$$

and there are no smaller values of $\mathcal{S} > 0$ for which (3.1) holds. This definition implies that $\mu_{Z,t}$ and $\gamma_{Z,t}(\tau)$ are periodic functions in t and need to be known only for $t = 1, \ldots, \mathcal{S}$. If (Z_t) is PS_1 then it is weakly stationary in the usual sense. In the following, we assume without loss of generality that $\mu_{Z,t} = 0$ for all $t \in \mathbb{Z}$, and we use the notation $t = (r-1)\mathcal{S}+\nu$ where $r \in \mathbb{Z}$ and the season $\nu = 1, \ldots, \mathcal{S}$.

One of the most popular $\mathrm{PS}_{\mathcal{S}}$ process is the PARMA model which generalizes the ARMA model, see e.g. Vecchia (1985). (Z_t) is said to be a PARMA model if it satisfies the difference equation

$$\sum_{j=0}^{p_\nu} \phi_{\nu,j} Z_{(r-1)\mathcal{S}+\nu-j} = \sum_{k=0}^{q_\nu} \theta_{\nu,k} \varepsilon_{(r-1)\mathcal{S}+\nu-k}, \tag{3.2}$$

where, for each season ν, p_ν and q_ν are the AR and MA orders, respectively, $\phi_{\nu,1}, \ldots, \phi_{\nu,p_\nu}$ and $\theta_{\nu,1}, \ldots, \theta_{\nu,q_\nu}$ are the AR and MA coefficients, respectively, and $\phi_{\nu,0} = \theta_{\nu,0} = 1$. The sequence (ε_t) is zero-mean and uncorrelated, and has periodic variances with period $\mathcal{S}$, i.e. $\mathrm{E}(\varepsilon^2_{(r-1)\mathcal{S}+\nu}) = \sigma^2_\nu$ for $\nu = 1, \ldots, \mathcal{S}$. In the following, we set $p = \max_\nu p_\nu$, $q = \max_\nu q_\nu$, $\phi_{\nu,j} = 0$ for $j > p_\nu$, $\theta_{\nu,k} = 0$ for $k > q_\nu$, and we refer to (3.2) as the $\mathrm{PARMA}(p,q)_{\mathcal{S}}$ model.

Let $(\mathbf{Z}_r)_{r\in\mathbb{Z}}$ be the $\mathcal{S}$-variate time series defined by $\mathbf{Z}_r' = [Z_{(r-1)\mathcal{S}+1}, \ldots, Z_{(r-1)\mathcal{S}+\mathcal{S}}]$, where $\mathbf{Z}_r'$ denotes the transpose of $\mathbf{Z}_r$. It is well known that (Z_t) is $\mathrm{PS}_{\mathcal{S}}$ if and only if $(\mathbf{Z}_r)$ is weakly stationary. The covariance matrix function of $(\mathbf{Z}_r)$ is $\Gamma_{\mathbf{Z}}(\tau) = \mathrm{Cov}(\mathbf{Z}_r, \mathbf{Z}_{r-\tau})$ and is related to $\gamma_{Z,t}(\tau)$ by $[\Gamma_{\mathbf{Z}}(\tau)]_{l,m} = \gamma_{Z,l}(\tau\mathcal{S} + l - m)$ for every $l, m = 1, \ldots, \mathcal{S}$. Now (3.2) is equivalent to the vector ARMA (VARMA) difference equation

$$\sum_{j=0}^{P} \Phi_j \mathbf{Z}_{r-j} = \sum_{k=0}^{Q} \Theta_k \boldsymbol{\varepsilon}_{r-k},$$

where $P = \lceil p/\mathcal{S} \rceil$, $Q = \lceil q/\mathcal{S} \rceil$ and $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. The entries of matrix Φ_j are $[\Phi_j]_{l,m} = \phi_{l,j\mathcal{S}+l-m}$ with the convention that $[\Phi_0]_{l,m} = 0$

when $l < m$. The definition of Θ_k is similar. The white noise vector process $(\boldsymbol{\varepsilon}_r)$ is defined by $\boldsymbol{\varepsilon}_r = [\varepsilon_{(r-1)\mathcal{S}+1}, \ldots, \varepsilon_{(r-1)\mathcal{S}+\mathcal{S}}]'$ and has the covariance matrix $\Sigma_{\boldsymbol{\varepsilon}} = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_{\mathcal{S}}^2)$.

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

$$\Phi(z) = \sum_{j=0}^{P} \Phi_j z^j \quad \text{and} \quad \Theta(z) = \sum_{k=0}^{Q} \Theta_k z^k,$$

and assume that $\det \Phi(z)\Theta(z) \neq 0$ for $|z| \leq 1$. Therefore, $(\mathbf{Z}_r)$ is causal and invertible and the spectral density matrix of $(\mathbf{Z}_r)$ is

$$\mathbf{f}_{\mathbf{Z}}(\omega) = \frac{1}{2\pi}\Phi^{-1}(e^{-\mathbf{i}\omega})\Theta(e^{-\mathbf{i}\omega})\Sigma_{\boldsymbol{\varepsilon}}\Theta'(e^{\mathbf{i}\omega})\Phi'^{-1}(e^{\mathbf{i}\omega}), \quad \omega \in (-\pi, \pi].$$

The causality and invertibility do not ensure that $\Sigma_{\boldsymbol{\varepsilon}}$, $\Phi(z)$ and $\Theta(z)$ are uniquely determined by the covariance matrix function $\Gamma_{\mathbf{Z}}(\tau)$, or equivalently the spectral matrix $\mathbf{f}_{\mathbf{Z}}(\omega)$, see e.g. Brockwell & Davis (2006, page 431). Additional restrictions have to be imposed in order to obtain identifiable models, see Dunsmuir & Hannan (1976) and Deistler, Dunsmuir & Hannan (1978). In the following, we assume that model (3.2) is identifiable.

Example 3.1. When $p_\nu = q_\nu = 1$ for all $\nu = 1, \ldots, \mathcal{S}$, we have $p = q = 1$, $P = Q = 1$,

$$\Phi_0 = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \phi_{2,1} & 1 & \cdots & \cdots & 0 \\ 0 & \phi_{3,1} & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \phi_{\mathcal{S},1} & 1 \end{bmatrix}, \quad \Phi_1 = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & \phi_{1,1} \\ 0 & \cdots & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \cdots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix} \quad \text{and}$$

$$\det \Phi(z) = \det(\Phi_0 + \Phi_1 z) = 1 + (-1)^{\mathcal{S}+1} \left(\prod_{\nu=1}^{\mathcal{S}} \phi_{\nu,1} \right) z.$$

The results for Θ_0, Θ_1 and $\det \Theta(z)$ are similar. The causality condition $\det \Phi(z) \neq 0$ for $|z| \leq 1$, and the invertibility condition $\det \Theta(z) \neq 0$ for $|z| \leq 1$ are equivalent respectively to

$$\vartheta_\phi = \prod_{\nu=1}^{\mathcal{S}} |\phi_{\nu,1}| < 1 \quad \text{and} \quad \vartheta_\theta = \prod_{\nu=1}^{\mathcal{S}} |\theta_{\nu,1}| < 1. \tag{3.3}$$

As discussed in Section 3.1, we shall consider AO since they cause the more deleterious effect in the inference of time series. Sarnaglia, Reisen & Lévy-Leduc (2010) have shown, for periodic processes, that this contamination can induce a spurious memory loss by increasing the variance of the process, both theoretically and empirically. Let (Y_t) be defined as

$$Y_t = Z_t + \varrho B_t, \tag{3.4}$$

where $\varrho > 0$ is the magnitude of the outlier and $(B_t)_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed random variables assuming $-1, 0, 1$ values with probabilities

$\mathsf{P}(B_t = 0) = 1 - \zeta$ and $\mathsf{P}(B_t = -1) = \mathsf{P}(B_t = 1) = \frac{\zeta}{2}$, where $\zeta \in (0, 1)$ is the probability of occurrence of an outlier. We assume that (B_t) and (Z_t) are independent processes. Note that the definition of the process in (3.4) ensures equal probabilities for positive and negative outliers. Observe that if either $\zeta = 0$ or $\varrho = 0$, then $Y_t = Z_t$, such that (Y_t) is uncontaminated. It is worth to point out that, for the AO process defined in (3.4), the location of the outliers is chosen at random, which seems to be more appropriate in real applications, since the position or even the occurrence of the outliers in the sample is usually unknown. There are other ways to describe atypical observations. For example, one can consider heavy-tailed distributions for the white noise process (ε_t), see e.g. Katkovnik (1998). However, in this kind of outlier generating mechanism, there is no explicit definition for the magnitude of the outliers and their investigation can not be performed directly.

The effects of outliers in the spectral density and the classical periodogram have been investigated by Fajardo, Reisen & Cribari-Neto (2009) for long memory processes. The autocovariance function $\gamma_{Y,t}(\tau) = \mathrm{Cov}(Y_t, Y_{t-\tau})$ of the contaminated process (Y_t) in (3.4) is given by

$$\gamma_{Y,t}(\tau) = \begin{cases} \gamma_{Z,t}(0) + \varrho^2\zeta, & \tau = 0, \\ \gamma_{Z,t}(\tau), & \tau \neq 0, \end{cases}$$

while $\mathrm{E}(Y_t) = \mu_{Y,t} = \mu_{Z,t}$. Therefore, (Y_t) is also a $\mathrm{PS}_{\mathcal{S}}$ process but with larger variance than (Z_t). Let $\mathbf{Y}_r' = [Y_{(r-1)\mathcal{S}+1}, \ldots, Y_{(r-1)\mathcal{S}+\mathcal{S}}]$ and $\Gamma_{\mathbf{Y}}(\tau) = \mathrm{Cov}(\mathbf{Y}_r, \mathbf{Y}_{r-\tau})$. Then

$$\Gamma_{\mathbf{Y}}(\tau) = \Gamma_{\mathbf{Z}}(\tau) + \mathbf{D}\,\mathbb{1}_{\{\tau=0\}}, \tag{3.5}$$

where $\mathbf{D}$ is the $(\mathcal{S} \times \mathcal{S})$ diagonal matrix with diagonal entries $(\mathbf{D})_{l,l} = \varrho^2\zeta$, $l = 1, \ldots, \mathcal{S}$. Therefore, the spectral density matrix of the contaminated vector process $(\mathbf{Y}_r)$ is given by

$$\mathbf{f}_{\mathbf{Y}}(\omega) = \mathbf{f}_{\mathbf{Z}}(\omega) + \frac{1}{2\pi}\mathbf{D}, \quad \omega \in (-\pi, \pi].$$

Note that, letting $\varrho \to \infty$ makes the diagonal matrix $\mathbf{D}$ dominate $\mathbf{f}_{\mathbf{Y}}$, which becomes close to the spectral density of a vector white noise process. This is the frequency domain counterpart of the memory loss property of AO processes.

3.3 Robust Whittle M-estimator

We now introduce the estimation method proposed in this paper. Firstly, we define a robust alternative to the Fourier transform based on the non-linear M-regression approach. Next, the robust Whittle-type method is presented.

3.3.1 Robust Fourier Transform

Let $(\mathbf{X}_r)_{r\in\mathbb{Z}}$ be any $\mathcal{S}$-dimensional vector process and $\mathbf{X} = [\mathbf{X}_1', \ldots, \mathbf{X}_N']'$ be a sample of size N observed from $(\mathbf{X}_r)$. The Fourier transform of $\mathbf{X}$, at the frequency $\omega \in (-\pi, \pi]$, is defined as

$$W_{\mathbf{X}}(\omega) = (2\pi N)^{-1/2} \sum_{r=1}^{N} \mathbf{X}_r e^{-\mathbf{i}r\omega}. \tag{3.6}$$

Equation (3.6) can be rapidly obtained at the Fourier frequencies $\omega_j = 2\pi j/N$, $j = 1, \ldots, N'$, where $N' = \lfloor (N-1)/2 \rfloor$ is the greatest integer smaller than or equal to $(N-1)/2$.

For any fixed frequency ω_j, the $\mathcal{S}$-dimensional vector $W_{\mathbf{X}}(\omega_j)$ can be viewed as a linear regression vector as follows. Let $X_{r,\nu}$ be the νth component of vector $\mathbf{X}_r$, $r = 1, \ldots, N$, $\nu = 1, \ldots, \mathcal{S}$. Define the vector of covariates $C_{r,j} = [\cos(r\omega_j), \sin(r\omega_j)]'$ and consider the linear model

$$X_{r,\nu} = C_{r,j}'\beta_\nu(\omega_j) + \xi_{r,\nu},$$

where $\xi_{r,\nu}$ is a random error term and the coefficient vector $\beta_\nu(\omega_j) = [\beta_{\nu,1}(\omega_j), \beta_{\nu,2}(\omega_j)]'$ can be seen as describing the impact of the jth harmonic in the time series $X_{1,\nu}, \ldots, X_{N,\nu}$. The classical least square estimator of the vector $\beta_\nu(\omega_j)$ is given by

$$\hat{\beta}_\nu(\omega_j) = \underset{\beta_\nu(\omega_j)\in\mathbb{R}^2}{\operatorname{argmin}} \left[\sum_{r=1}^{N} \big(X_{r,\nu} - C_{r,j}'\beta_\nu(\omega_j) \big)^2 \right]. \tag{3.7}$$

Now, define the vector $\mathbf{d}(\omega_j) = [d_1(\omega_j), \ldots, d_{\mathcal{S}}(\omega_j)]'$, where $d_\nu(\omega_j) = \hat{\beta}_{\nu,1}(\omega_j) - \mathbf{i}\hat{\beta}_{\nu,2}(\omega_j)$. Similar arguments as those in Fajardo et al. (2015) can be used in order to show that

$$W_{\mathbf{X}}(\omega_j) = \sqrt{N/8\pi}\, \mathbf{d}(\omega_j), \quad j = 1, \ldots, N'.$$

As well known, $\hat{\beta}_\nu(\omega_j)$ does not have the necessary robustness to withstand the effect of neither atypical observations nor heavy-tailed distributions. For improving robustness, one idea is to replace $\hat{\beta}_\nu(\omega_j)$ by a non-linear M-regression estimator in $W_{\mathbf{X}}(\omega_j)$. This will lead to the robust periodogram for PS processes proposed here. The key idea is to replace the quadratic loss function in (3.7) by an alternative function $\rho(\cdot)$, which gives

$$\underset{\beta_\nu(\omega_j)\in\mathbb{R}^2}{\operatorname{argmin}} \left[\sum_{r=1}^{N} \rho(X_{r,\nu} - C_{r,j}'\beta_\nu(\omega_j)) \right].$$

Equivalently, one can define the M-estimator $\hat{\beta}_{\nu,\psi}(\omega_j)$ of $\beta_\nu(\omega_j)$ as the solution of

$$\sum_{r=1}^{N} C_{r,j}\, \psi\big(X_{r,\nu} - C_{r,j}'\hat{\beta}_{\nu,\psi}(\omega_j) \big) = 0,$$

where $\psi(\cdot)$ is the derivative of $\rho(\cdot)$. In this paper, we use the Huber (1964) function,

$$\rho(x) = \begin{cases} x^2/2, & |x| \le \delta, \\ \delta(|x| - \delta/2), & |x| > \delta. \end{cases} \tag{3.8}$$

The choice of the tunning parameter $\delta > 0$ is quite important and provides the compromise between robustness and efficiency of the M-estimators.

Finally, by defining $\mathbf{d}_\psi(\omega_j)$ similarly to $\mathbf{d}(\omega_j)$ with $\hat{\beta}_\nu(\omega_j)$ replaced by $\hat{\beta}_{\nu,\psi}(\omega_j)$, the robust alternative to $W_{\mathbf{X}}(\omega_j)$ proposed here is given by

$$W_{\mathbf{X},\psi}(\omega_j) = \sqrt{N/8\pi}\,\mathbf{d}_\psi(\omega_j), \quad j = 1, \ldots, N'. \tag{3.9}$$

3.3.2 Whittle M-estimator of PARMA parameters

Assume that $Z_1, \ldots, Z_n$ is a sample from a PARMA process with known orders p and q. For simplicity, suppose that $n = N\mathcal{S}$, such that every season $\nu = 1, \ldots, \mathcal{S}$ is observed N times. As previously, define the vector $\mathbf{Z}_r = [Z_{(r-1)\mathcal{S}+1}, \ldots, Z_{(r-1)\mathcal{S}+\mathcal{S}}]'$ corresponding to the rth cycle, the full sample being given by $\mathbf{Z} = [\mathbf{Z}_1', \ldots, \mathbf{Z}_N']'$. The parameter vector of model (3.2) is $\varphi = [\varphi_\phi', \varphi_\theta', \varphi_\sigma']'$ where φ_ϕ and φ_θ contain all the AR and MA parameters, respectively, and $\varphi_\sigma = [\sigma_1^2, \ldots, \sigma_{\mathcal{S}}^2]'$.

We define the parameter space $\mathcal{P} \subset \mathbb{R}^{(p+q)\mathcal{S}}$ as the set of points $[\varphi_\phi', \varphi_\theta']'$ for which model (3.2) is identifiable in the sense of Deistler, Dunsmuir & Hannan (1978). 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 $\Gamma_{\mathbf{Z}}(m-l)$ in the (l,m)th block of $\mathcal{S} \times \mathcal{S}$ elements, $1 \leq l, m \leq N$. The Gaussian log likelihood with the scaling factor $-2N^{-1}$ is

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

We denote by φ_0 the true parameter vector φ from which the sample $Z_1, \ldots, Z_n$ is generated. We assume that $[\varphi_{\phi_0}', \varphi_{\theta_0}']' \in \mathcal{P}$, and we have $\Gamma_N(\varphi_0) = \mathrm{Cov}(\mathbf{Z}, \mathbf{Z})$. The Gaussian maximum likelihood estimator (MLE) of φ_0 is $\hat{\varphi}_N = \mathrm{argmin}_{\varphi \in \mathcal{P} \times \mathbb{R}_{>0}^{\mathcal{S}}} \hat{\mathcal{L}}_N(\varphi)$.

To obtain $\hat{\varphi}_N$, an optimization algorithm is used and can demand high computational effort due to the fact that $\Gamma_N(\varphi)$ has to be inverted. To circumvent this difficulty, we use the multivariate version of Whittle's methodology proposed by Dunsmuir & Hannan (1976) to approximate $\hat{\mathcal{L}}_N(\varphi)$. For a PARMA process, it was shown by Sarnaglia, Reisen & Bondon (2015) that the corresponding Whittle likelihood estimator (WLE) of φ_0 is $\tilde{\varphi}_N = [\tilde{\varphi}_{\phi_N}', \tilde{\varphi}_{\theta_N}', \tilde{\varphi}_{\sigma_N}']'$ where

$$\begin{aligned}
[\tilde{\varphi}_{\phi_N}', \tilde{\varphi}_{\theta_N}']' &= \underset{(\varphi_\phi', \varphi_\theta')' \in \mathcal{P}}{\mathrm{argmin}} \sum_{\nu=1}^{\mathcal{S}} \log \tilde{\sigma}_{N,\nu}^2(\varphi_\phi, \varphi_\theta), \\
\tilde{\sigma}_{N,\nu}^2(\varphi_\phi, \varphi_\theta) &= 2\pi N'^{-1} \sum_{j=1}^{N'} \left| \left[\Theta^{-1}(e^{-\mathbf{i}\omega_j})\Phi(e^{-\mathbf{i}\omega_j}) W_{\mathbf{Z}}(\omega_j)\right]_\nu \right|^2, \quad \nu = 1, \ldots, \mathcal{S},
\end{aligned} \tag{3.10}$$

$W_{\mathbf{Z}}(\omega_j)$ is given by (3.6) in which $\mathbf{X}_r$ is replaced by $\mathbf{Z}_r$, and the νth component of $\tilde{\varphi}_{\sigma_N}$ is $\tilde{\sigma}_{N,\nu}^2(\tilde{\varphi}_{\phi_N}, \tilde{\varphi}_{\theta_N})$ for $\nu = 1, \ldots, \mathcal{S}$.

Now, we define the robust WLE (RWLE) $\tilde{\varphi}_{N,\psi}$ of φ_0 similarly as $\tilde{\varphi}_N$ by replacing $W_{\mathbf{Z}}(\omega_j)$ in (3.10) by $W_{\mathbf{Z},\psi}(\omega_j)$ defined in (3.9) where $\mathbf{X}_r$ is replaced by $\mathbf{Z}_r$.

It was pointed out by Sarnaglia, Reisen & Bondon (2015) that $[\varphi'_\phi, \varphi'_\theta]$ involves $(p+q)\mathcal{S}$ parameters whereas the dimension of φ is $(p+q+1)\mathcal{S}$. Then $\tilde{\varphi}_N$ is easier to calculate and is obtained faster than $\hat{\varphi}_N$. The same remark applies to the calculation of $\tilde{\varphi}_{N,\psi}$. However, the computation time of $\tilde{\varphi}_{N,\psi}$ may be larger than $\tilde{\varphi}_N$ because a numerical optimization method is needed to obtain $W_{\mathbf{Z},\psi}(\omega_j)$, since this function does not have a closed form expression.

3.4 Monte Carlo study

In this section we investigate the finite sample behaviour of the proposed estimator. Additionally, the MLE and WLE approaches are also considered in the study for comparison purposes. We consider uncontaminated and contaminated data with AO with occurrence probability $\zeta = 0.01$ and magnitude $\varrho = 10$. Less detrimental contamination parameters were also considered, however we prefer to show just the worst scenario in order to highlight the advantages of the proposed methodology. We generate $M = 1000$ replicates of the PARMA$(1,1)_{\mathcal{S}}$ process (Z_t) in (3.2), with $\mathcal{S} = 2$ and parameters given in Table 3.1. Other PARMA models were also considered and the results shown similar conclusions. They are not presented here to save space, but are available upon request.

Table 3.1 – Parameters.

	$\nu = 1$			$\nu = 2$			Eq. (3.3)	
Model	$\phi_{1,1_0}$	$\theta_{1,1_0}$	$\sigma^2_{1_0}$	$\phi_{2,1_0}$	$\theta_{2,1_0}$	$\sigma^2_{2_0}$	ϑ_ϕ	ϑ_θ
1	-0.2	0.0	1.0	-0.5	0.0	1.0	0.1	0.0
2	-0.2	-0.5	1.0	-0.5	-0.2	1.0	0.1	0.1
3	-1.0	0.0	1.0	-0.5	0.0	1.0	0.5	0.0
4	-1.0	-0.5	1.0	-0.5	-0.2	1.0	0.5	0.1

The sample sizes are $n = N\mathcal{S} = 300$, 800 ($N = 150$, 400, respectively) and the Huber (1964) function (3.8) is used with $\delta = 1.345$, which ensure that the M-estimator is 95% as efficient as the least squares estimator for univariate multiple linear models with independent and identically distributed Gaussian white noise.

We evaluate the finite sample performance of the estimators by computing the sample root mean square error (RMSE) and the results are displayed in Tables 3.2, 3.3, 3.4 and 3.5. The values with "$*$" refer to the RMSE for the contaminated series.

Table 3.2 – RMSE of Model 1 with $\vartheta_\phi = 0.1$ and $\vartheta_\theta = 0.0$.

Method	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	300	**0.067**; 0.121*	—	**0.117**; 1.366*	**0.079**; 0.252*	—	**0.111**; 1.363*
	800	**0.048**; 0.101*	—	**0.079**; 1.122*	**0.046**; 0.239*	—	**0.074**; 1.253*
WLE	300	**0.068**; 0.121*	—	**0.117**; 1.368*	**0.079**; 0.252*	—	**0.111**; 1.364*
	800	**0.048**; 0.101*	—	**0.079**; 1.122*	**0.046**; 0.239*	—	**0.074**; 1.253*
RWLE	300	**0.067**; **0.067***	—	0.147; **0.179***	**0.083**; **0.089***	—	0.147; **0.189***
	800	**0.051**; **0.054***	—	0.118; **0.149***	**0.051**; **0.058***	—	0.108; **0.152***

Table 3.3 – RMSE of Model 2 with $\vartheta_\phi = 0.1$ and $\vartheta_\theta = 0.1$.

Method	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	300	**0.364**; 1.393*	**0.371**; 1.398*	**0.120**; 1.433*	0.638; 2.219*	0.649; 2.249*	**0.114**; 1.257*
	800	**0.171**; 0.492*	**0.184**; 0.500*	**0.065**; 1.118*	**0.167**; 1.007*	**0.171**; 1.031*	**0.064**; 1.102*
WLE	300	**0.369**; 0.981*	**0.377**; 0.985*	**0.119**; 1.433*	**0.545**; 1.309*	**0.559**; 1.340*	**0.114**; 1.255*
	800	**0.171**; 0.439*	**0.184**; 0.448*	**0.065**; 1.119*	**0.167**; 0.466*	**0.171**; 0.504*	**0.064**; 1.102*
RWLE	300	**0.357**; **0.344***	**0.371**; **0.363***	0.135; **0.164***	0.747; **0.704***	0.760; **0.714***	0.150; **0.180***
	800	**0.178**; **0.186***	**0.193**; **0.201***	0.101; **0.132***	0.189; **0.200***	0.194; **0.206***	0.101; **0.131***

Table 3.4 – RMSE of Model 3 with $\vartheta_\phi = 0.5$ and $\vartheta_\theta = 0.0$.

Method	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	300	**0.064**; 0.394*	—	**0.107**; 1.770*	**0.047**; 0.159*	—	**0.107**; 1.327*
	800	**0.036**; 0.373*	—	**0.069**; 1.642*	**0.030**; 0.142*	—	**0.070**; 1.279*
WLE	300	**0.066**; 0.397*	—	**0.116**; 1.780*	**0.047**; 0.159*	—	**0.107**; 1.328*
	800	**0.037**; 0.374*	—	**0.073**; 1.646*	**0.030**; 0.142*	—	**0.070**; 1.279*
RWLE	300	**0.080**; **0.107***	—	0.193; **0.334***	**0.058**; **0.067***	—	0.156; **0.227***
	800	**0.053**; **0.077***	—	0.168; **0.317***	**0.038**; **0.047***	—	0.134; **0.215***

Table 3.5 – RMSE of Model 4 with $\vartheta_\phi = 0.5$ and $\vartheta_\theta = 0.1$.

Method	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
MLE	300	**0.233**; 1.539*	**0.252**; 1.594*	**0.118**; 1.307*	**0.124**; 0.319*	**0.150**; 0.350*	**0.107**; 1.337*
	800	**0.140**; 0.278*	**0.146**; 0.376*	**0.068**; 1.230*	**0.077**; 0.134*	**0.089**; 0.181*	**0.071**; 1.141*
WLE	300	**0.236**; 0.792*	**0.255**; 0.864*	**0.121**; 1.313*	**0.125**; 0.238*	**0.150**; 0.280*	**0.107**; 1.336*
	800	**0.141**; 0.276*	**0.147**; 0.372*	**0.068**; 1.233*	**0.077**; 0.133*	**0.089**; 0.180*	**0.071**; 1.141*
RWLE	300	0.272; **0.296***	0.288; **0.311***	0.153; **0.207***	0.134; **0.140***	**0.148**; **0.152***	0.141; **0.183***
	800	0.150; **0.149***	0.155; **0.154***	0.112; **0.169***	0.082; **0.085***	**0.094**; **0.097***	0.107; **0.145***

In the uncontaminated data scenario, in general, all estimators present similar behaviour in the AR and MA counterparts. Relating to the estimation of the variance of the innovations, the MLE and WLE seems to be more precise which is an expected result since the data is Gaussian with zero-mean and these two methods are asymptotically equivalents. The RMSE of the estimators decreases as the sample size increases. In addition, increasing the model order will also affect the estimates. However, the conclusion by comparing the tree methods are similar.

Now, the discussion is related to the case where the process is contaminated with additive outliers. As expected, the MLE and WLE estimates are totally corrupted by the atypical observations since their sample RMSEs increases substantially. Therefore, these methods should be avoided when the series contains additive outliers. The robust estimator (RWLE) presents generally accurate estimates even for the largest proportion of contamination we have considered. Its superiority over the MLE and WLE methods is clearly shown in the above tables.

We also display in Figure 3.1 the empirical distributions of the estimates provided by the MLE, the WLE and the RWLE for Model 3. It is clear the AO effect in the MLE and the WLE, while the RWLE remains almost unchanged.

In addition to the above empirical investigation, the robust method was also considered when dealing with PARMA model with heavy tail distributions, such as t-student and double exponential. The results have leaded to similar conclusions with AO ones. These are not presented here to save space but are available upon request.

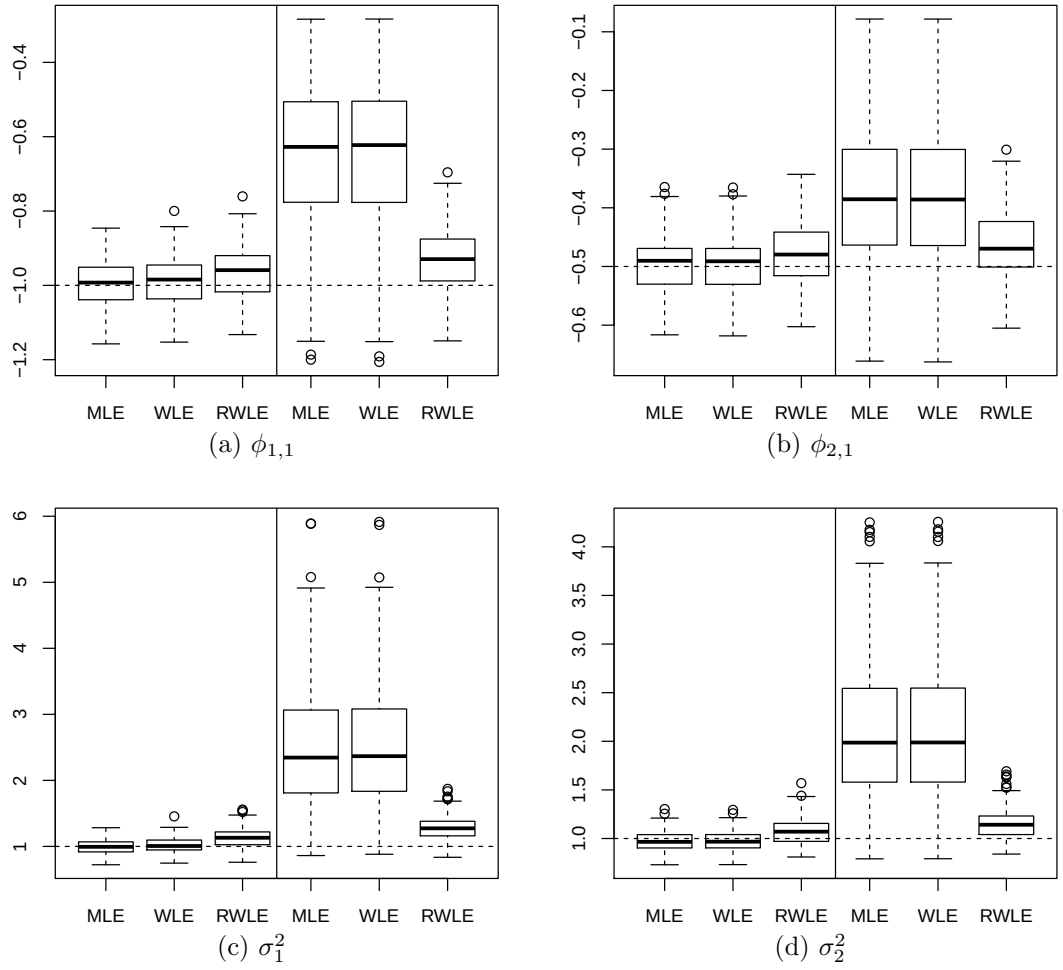
3.5 Application

We analyze the daily mean concentrations of Carbon Monoxide (CO) in Ibes, Vila Velha, ES, Brazil, observed from January 1, 2005 to March 31, 2007 at the monitoring station of environment and water resources state institute. Since the time series is daily collected, a PARMA$_{\mathcal{S}}$ model with $\mathcal{S} = 7$ seems to be appropriate. The first $n = N\mathcal{S} = 728$ observations (almost two years) are used to fit the model. The last 92 records are kept to perform a forecast study, see Figure 3.2.

Since the MLE takes a large amount of time to provide the estimates (SARNAGLIA; REISEN; BONDON, 2016a), we consider in this application only the WLE and the RWLE in order to be able to use the Schwarz Information Criterion (BIC) for model identification. The RWLE is obtained using the Huber discrepancy function (see Equation 3.8) such as in the simulation study. For WLE, the BIC is given by

$$BIC = N \sum_{\nu=1}^{\mathcal{S}} \log(\tilde{\sigma}^2_{N,\nu}) + \log(N) \sum_{\nu=1}^{\mathcal{S}} (p_\nu + q_\nu),$$

while, for RWLE, one just have to replace $\tilde{\sigma}^2_{N,\nu}$ by $\tilde{\sigma}^2_{N,\nu,\psi}$ in the above equation. For WLE, the original data are centralized by subtracting the sample periodic means, while



(a) $\phi_{1,1}$ (b) $\phi_{2,1}$

(c) σ_1^2 (d) σ_2^2

Figure 3.1 – Box-plot of the estimates of the MLE, the WLE and the RWLE of Model 3 for $n = 300$. On each subfigure the left and the right boxes represent, respectively, the uncontaminated and the contaminated time series scenarios.

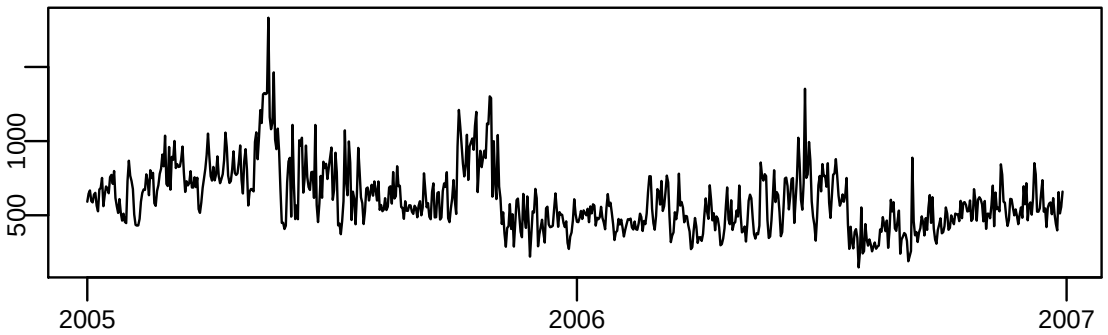


Figure 3.2 – Daily mean concentrations of CO in Ibes, Vila Velha, ES, Brazil.

for RWLE they are corrected by subtracting the huber location M-estimator evaluated periodically. The initial values for AR and MA estimates were set as zero, and for white noise variances, as $S^2_{X,\nu}$, which is the data sample variance of the period ν. We compare the BIC just of identifiable models by imposing the restrictions of Sarnaglia, Reisen & Bondon (2016a). The BIC has selected the constant AR order $p_\nu = 1$ for both estimators

and the MA orders $[q_1, \ldots, q_{\mathbb{S}}]' = [0, 0, 1, 0, 1, 1, 1]'$ and $[q_1, \ldots, q_{\mathbb{S}}]' = [0, 0, 0, 1, 0, 0, 1]'$ for WLE and RWLE, respectively. The BIC values were 7126.90 and 7041.92 and the estimation times were 32.61 and 38.05 seconds for WLE and RWLE, respectively. These estimates were obtained in an intel core i7-2630QM computer with 8 GB of RAM. Table 3.6 shows WLE and RWLE estimates for the series.

Table 3.6 – WLE and RWLE (with a "$*$") estimates.

ν	$\phi_{\nu,1}$		$\theta_{\nu,1}$		σ^2_ν
1	-0.73;	-0.70^*	—	—	118.59; 119.86*
2	-0.74;	-0.73^*	—	—	115.50; 109.73*
3	-1.18;	-0.89^*	-0.37	—	123.08; 116.98*
4	-0.65;	-0.78^*	—	-0.04^*	153.92; 131.06*
5	-0.77;	-0.71^*	-0.24	—	129.12; 109.07*
6	-0.94;	-0.77^*	-0.26	—	129.26; 135.31*
7	-0.97;	-0.95^*	-0.41;	-0.26^*	137.22; 138.77*

We first observe in Figure 3.2 large spikes incompatible with the overall dynamics of the times series. Also, BIC has selected different models for WLE and RWLE. In addition, we observe from Table 3.6 that WLE and RWLE estimates are quite different for some periods. See for example the reduction of the white noise variance estimates of the RWLE compared to the WLE. This may be the indication of the presence of outliers. The Median Absolute Deviation (MAD) for MLE and RWLE are, respectively, 75.51 and 74.45 for the residuals and 58.59 and 56.15 for the 92 discarded values. The RMSE was also calculated and presents for the residuals smaller value for WLE (130.09 versus 132.45), which is expected since WLE is asymptotically equivalent to Least Square Estimator. The out-of-sample RMSE is almost similar for both estimators: 107.9 for WLE; and 107.17 for RWLE. We observe that RWLE provide better fit and forecast performance with respect to MAD. Figure 3.3 displays the one-step-ahead forecasts and 95% forecast intervals. One may note the difference in the intervals due to the contrast in the white noise variance estimates provided by the WLE and the RWLE.

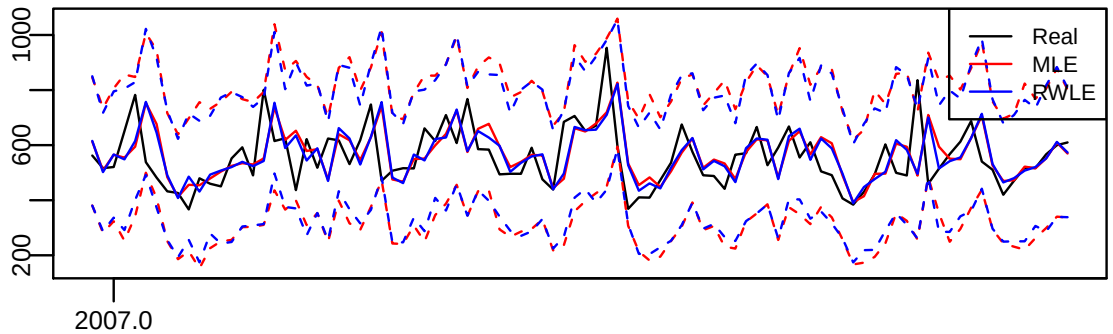


Figure 3.3 – Forecast of the discarded PM_{10} concentrations.

4 On the use of classical and M periodograms to fit periodic ARMA models to time series with missing data and additive outliers

Abstract

This paper proposes two estimators for Periodic Autoregressive Moving Average models with missing data. The first one is based on the frequency domain and uses the asymptotic spectrum of an amplitude modulated process. In order to improve robustness, the second estimator is built from the first one by replacing the classical periodogram by the M-periodogram of Sarnaglia, Reisen & Bondon (2016b). The finite sample properties of the proposed methodologies are investigated through an extensive Monte Carlo study. The results show that, under absence of additive outliers, both methods behave satisfactorily well compared to the complete sample estimates. However, under the presence of outliers the first method becomes corrupted, whilst the robust alternative remains reliable. In order to illustrate the usefulness in applications, an air pollution time series is fitted using the proposed methodologies.

KEYWORDS. Periodic stationarity, PARMA models, Missing data, Robust estimation, Additive outliers, Whittle estimation.

4.1 Introduction

Processes with periodically varying covariances have been introduced in the seminal paper by Gladyshev (1961) and are usually denominated as periodically correlated, Periodically Stationary (PS) or cyclostationary. Tiao & Grupe (1980) have shown the effects of misspecification of PS processes. Their importance has been corroborated by real applications in many areas. For example, Gardner & Franks (1975) have investigated cyclostationarity in electrical engineering and Bloomfield, Hurd & Lund (1994) have studied periodic correlation in stratospheric ozone data. For recent reviews on PS processes, see e.g., Gardner, Napolitano & Paura (2006) and Hurd & Miamee (2007).

The Periodic Autoregressive Moving Average (PARMA) model has been considerably investigated in the literature. In special, estimation methodologies for PARMA parameters. For example, Lund & Basawa (2000), have proposed an efficient algorithm to

evaluate the exact gaussian likelihood of the PARMA models, Basawa & Lund (2001) have studied the least square estimation of these models and Sarnaglia, Reisen & Bondon (2015) have proposed a Whittle type estimator and Sarnaglia, Reisen & Bondon (2016a) have investigated its asymptotics.

The good performance of the previous estimation methods for PARMA time series has been evidenced by theoretical and simulation results. See e.g. Sarnaglia, Reisen & Bondon (2016a). However, in the presence of Additive Outliers (AO), these estimators are generally deteriorated. This kind of outlier is usually investigated due to its more deleterious effect in parameter estimates. In this context, Sarnaglia, Reisen & Bondon (2016b) have suggested a robust estimation methodology based on a alternative periodogram obtained through M-regression, which generalizes the idea of Fajardo et al. (2015) to the periodic scenario. They have shown that this method presents similar performance to the classical alternatives in the absence of outliers and it is still robust for AO contaminated PARMA time series.

Frequently, in real applications, the data set is not completely observable. Obviously, the classical tools can not be used in the missing observations scenario. On one hand, one can deal with this problem by Expectation-Maximization type algorithms. This approach has the disadvantage of assuming an specific distribution (gaussian in general) for the data set. On the other hand, many papers extend standard tools to the incomplete sample perspective, for example Yajima & Nishino (1999) have studied the asymptotics of three different estimators of the sample autocorrelation function of stationary time series, Dunsmuir & Robinson (1981b) propose a Whittle type estimation method in the presence of missing data and Dunsmuir & Robinson (1981c) studied its asymptotics. Other recent examples of extentions of classical tools to the missing data situation are Bondon & Bahamonde (2012) and Efromovich (2014).

To our best knowledge, fitting PARMA models to incomplete time series is still a little exploited subject and methods for PARMA time series with missing data and additive outliers have never been investigated. This paper aim to deal with these problems. More specifically, we shall propose two estimation methods. The first one is a Whittle type estimator based on the asymptotic spectrum of the amplitude modulated model generated by the missing data which generalizes the method of Dunsmuir & Robinson (1981b), Dunsmuir & Robinson (1981c). In order to improve robustness, we shall suggest another estimator based on the first one by replacing the classical periodogram by the M-periodogram of Sarnaglia, Reisen & Bondon (2016b), so that it can be used in situations with additive outliers and missing data simultaneously.

The rest of the paper is structured as follows: Section 4.2 defines the PARMA model with AO; Section 4.3 introduces the Asymptotically Periodically Stationary processes, studies some of their properties and investigates the amplitude modulated model. Section 4.4 proposes two estimation methods and investigates some effects of additive outliers;

The finite sample performances of these estimators are investigated and compared through an extensive Monte Carlo study presented in Section 4.5; An application to air pollution is presented in Section 4.6.

4.2 PARMA model with additive outliers

We first introduce some notation. A stochastic process observed over the integer set $\mathbb{Z}$, say $(Z_t)_{t\in\mathbb{Z}}$, will be denoted by the cleaner notation (Z_t). The scalar $[\mathbf{A}]_{l,m}$ will refer to the l, m entry of the matrix $\mathbf{A}$, that is the element in the lth row and mth column $\mathbf{A}$. The lth row and mth column of the matrix $\mathbf{A}$ are denoted, respectively, by $[\mathbf{A}]_{l\bullet}$ and $[\mathbf{A}]_{\bullet m}$. The transpose of the matrix $\mathbf{A}$ is denoted by $\mathbf{A}'$. The conjugate transpose of a complex matrix $\mathbf{A}$ is denoted by $\mathbf{A}^{\mathcal{H}}$. Vectors will be treated as one column matrices, i.e., a n-dimensional vector x has ith entry given by $[x]_i = [x]_{i,1}$, $i = 1, \ldots, n$. In this paper, we follow the usual practice and will not differ from notations of a process and a time series, the difference will be made clear from the context.

Let (Z_t) be a stochastic process with expectation and autocovariance functions given, respectively, by $\mu_{t,Z} = \mathrm{E}(Z_t)$ and $\gamma_{t,Z}(\tau) = \mathrm{Cov}(Z_t, Z_{t-\tau})$. One says that the process (Z_t) is Periodically Stationary (or cyclostationary) with season length $\mathcal{S}$ ($\mathrm{PS}_{\mathcal{S}}$) if $\mu_{t,Z}$ and $\gamma_{t,Z} = \gamma_{t,Z}(\tau)$ are periodic functions in t with period $\mathcal{S}$, that is,

$$\mu_{t,Z} = \mu_{t+\mathcal{S},Z} \quad \text{and} \quad \gamma_{t,Z} = \gamma_{t+\mathcal{S},Z}, \quad \forall\, t \in \mathbb{Z}.$$

These processes have been first considered in the seminal paper of Gladyshev (1961) and a broad collection of results in this area is given in Hurd & Miamee (2007). From now on, we shall consider the case where $\mu_{t,Z} = 0$, $t = 1, \ldots, \mathcal{S}$.

The Periodic Autoregressive Moving-Average model with season length $\mathcal{S}$ ($\mathrm{PARMA}_{\mathcal{S}}$) plays an important role in the analysis of $\mathrm{PS}_{\mathcal{S}}$ processes. It is a generalization of the so-called ARMA models to the periodic scenario, see e.g. Box, Jenkins & Reinsel (2008). In the definition of PARMA models, in order to emphasize which period is being considered, we shall use the periodic time index notation $t = r\mathcal{S} + \nu$, $r \in \mathbb{Z}$, $\nu = 1, \ldots, \mathcal{S}$. One says that (Z_t) follows a $\mathrm{PARMA}_{\mathcal{S}}$ model if it satisfies the difference equations

$$Z_{r\mathcal{S}+\nu} + \sum_{j=1}^{p_\nu} \phi_{\nu,j} Z_{r\mathcal{S}+\nu-j} = \varepsilon_{r\mathcal{S}+\nu} + \sum_{j=1}^{q_\nu} \theta_{\nu,j} \varepsilon_{r\mathcal{S}+\nu-j}, \tag{4.1}$$

where $\phi_{\nu,1}, \ldots, \phi_{\nu,p_\nu}$ and $\theta_{\nu,1}, \ldots, \theta_{\nu,q_\nu}$ and p_ν and q_ν are, respectively, the AR and MA coefficients and orders of the period ν. The process (ε_t) is a zero mean white noise sequence with periodic variances $\sigma_t^2 = \sigma_{t+\mathcal{S}}^2$. We can always consider the AR orders to be constant, otherwise we can take $p = \max_\nu p_\nu$ and make the additional AR coefficients to be zero. The same can be done for MA orders. Therefore, from now on, we shall assume $p_\nu = p$ and $q_\nu = q$ constant in $\nu = 1, \ldots, \mathcal{S}$.

It is well-known (see e.g. Gladyshev (1961)) that (Z_t) is a $\mathrm{PS}_{\mathcal{S}}$ process if, and only if, the vector process $(\mathbf{Z}_r)$ defined as $\mathbf{Z}_r = [Z_{r\mathcal{S}+1}, \ldots, Z_{r\mathcal{S}+\mathcal{S}}]'$ is weakly stationary in the vector sense. In addition, it is well-known that if (Z_t) follows the $\mathrm{PARMA}_{\mathcal{S}}$ model in Equation 4.1, its vector counterpart $(\mathbf{Z}_r)$ follows the VARMA model

$$\sum_{j=0}^{P} \Phi_j \mathbf{Z}_{r-j} = \sum_{j=0}^{Q} \Theta_j \boldsymbol{\varepsilon}_{r-j}, \tag{4.2}$$

where the orders $P = \lceil \frac{p}{\mathcal{S}} \rceil$ and $Q = \lceil \frac{q}{\mathcal{S}} \rceil$, wherein $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. The AR matricial coefficients are given by $[\Phi_j]_{l,m} = \phi_{l,j\mathcal{S}+l-m}$, wherein the conventions $\phi_{\nu,0} = 1$ and $\phi_{\nu,k} = 0$, $k < 0, k > p$, have been made. The definition of the Θ_j's is similar. The white noise vector process $(\boldsymbol{\varepsilon}_r)$ is defined by $\boldsymbol{\varepsilon}_r = [\varepsilon_{r\mathcal{S}+1}, \ldots, \varepsilon_{r\mathcal{S}+\mathcal{S}}]'$ and has variance matrix Σ with entries given by $[\Sigma]_{l,m} = \sigma_l^2 \mathbb{1}_{\{l=m\}}$, $l, m = 1, \ldots, \mathcal{S}$, where $\mathbb{1}$ is the indicator function. Define the AR and MA matricial polynomials, respectively, by

$$\Phi(z) = \sum_{j=0}^{P} \Phi_j z^j \quad \text{and} \quad \Theta(z) = \sum_{j=0}^{P} \Theta_j z^j.$$

We shall assume that causality and invertibility of $(\mathbf{Z}_r)$ which are given, respectively, by

$$\det \Phi(z) \neq 0 \quad \text{and} \quad \det \Theta(z) \neq 0, \quad |z| \leq 1,$$

hold. These assumptions ensure the existence of the spectral density matrix of $(\mathbf{Z}_r)$ which is given by

$$\mathbf{f}_{\mathbf{Z}}(\omega) = \frac{1}{2\pi} \Phi^{-1}(e^{-\mathbf{i}\omega}) \Theta(e^{-\mathbf{i}\omega}) \Sigma [\Phi^{-1}(e^{-\mathbf{i}\omega}) \Theta(e^{-\mathbf{i}\omega})]^{\mathcal{H}}, \tag{4.3}$$

where $e^{\mathbf{i}x} = \cos x + \mathbf{i} \sin x$ is the Euler's formula, or complex exponential. It can be seen that, for $\mathrm{PARMA}(1,1)_{\mathcal{S}}$ models, causality and invertibility conditions simplify to

$$\vartheta_\phi := \left| \prod_{\nu=1}^{\mathcal{S}} \phi_{\nu,1} \right| < 1 \quad \text{and} \quad \vartheta_\theta := \left| \prod_{\nu=1}^{\mathcal{S}} \theta_{\nu,1} \right| < 1, \tag{4.4}$$

respectively. Hereafter, besides these conditions, we also assume that the process $(\mathbf{Z}_r)$ is identifiable in the sense of Dunsmuir & Hannan (1976). Identifiability is particularly important to avoid a likelihood surface with more than one maxima. See conditions on PARMA parameters for identifiability in the paper of Sarnaglia, Reisen & Bondon (2016a).

As previously discussed, we shall consider AO since it causes the more deleterious effect in the inference of time series. Let (Z_t) be a scalar stochastic process and define the AO process (X_t) by

$$X_t = \mathcal{B}_t \mathcal{O}_t + Z_t, \tag{4.5}$$

where $(\mathcal{B}_t)$ is an independent Bernoulli sequence with success (outlier) probability $\mathsf{P}(\mathcal{B}_t = 1) = \zeta$, the $(\mathcal{O}_t)$ is an i.i.d. sequence of some contamination zero mean symmetric random

variable (r.v.), $\mathcal{B}_u, \mathcal{O}_s, Z_t$, $u, s, t \in \mathbb{Z}$, are independent r.v. and, in this paper, (Z_t) will be a PARMA process. The symmetry of the marginal distribution of $\mathcal{O}_t$ is imposed to ensure equal probabilities of positive and negative outliers. In general, the r.v. $\mathcal{O}_t$ may have a heavy-tailed distribution, or even the same marginal distribution of (Z_t) with larger variance. The contamination process considered by Sarnaglia, Reisen & Bondon (2016b) is a special case of (4.5) wherein the r.v. $\mathcal{O}_t$ obeys the following marginal law:

$$\mathsf{P}(\mathcal{O}_t = \varrho) = \mathsf{P}(\mathcal{O}_t = -\varrho) = \frac{1}{2}, \tag{4.6}$$

where in this case ϱ represents the outlier's magnitude.

Note that in Equation 4.5, if $\zeta = 0$, then $X_t = Z_t$, such that (X_t) is uncontaminated. It is worth to point out that, for the AO process defined in Equation 4.5, the location of the outliers is chosen at random, which seems to be more appropriate in real applications, since the position or even the occurrence of the outliers in the sample is usually unknown. There are other ways to describe atypical observations. For example, one can consider heavy-tailed distributions to the data. See e.g. Katkovnik (1998).

The effects of outliers in the spectral density and the classical periodogram have been investigated in Fajardo, Reisen & Cribari-Neto (2009) for scalar long memory processes. For periodic processes, when the contamination r.v. $\mathcal{O}_t$ is defined as in (4.6), the effect on correlation structure is presented by Sarnaglia, Reisen & Lévy-Leduc (2010) in the time domain, while, in the frequency domain, Sarnaglia, Reisen & Bondon (2016b) have shown that the spectral density of the $\mathcal{S}$-dimensional vector process $(\mathbf{X}_r)$, $r \in \mathbf{Z}$, wherein $[\mathbf{X}_r]_\nu = X_{r\mathcal{S}+\nu}$, $\nu = 1, \ldots, \mathcal{S}$, is given by

$$\mathbf{f}_{\mathbf{X}}(\omega) = \mathbf{f}_{\mathbf{Z}}(\omega) + \frac{1}{2\pi}\mathbf{D},$$

where $\mathbf{D}$ is a $(\mathcal{S} \times \mathcal{S})$ diagonal matrix with diagonal entries $[\mathbf{D}]_{l,l} = \varrho^2\zeta$, $l = 1, \ldots, \mathcal{S}$ and $\mathbf{f}_{\mathbf{Z}}$ is the spectral density matrix of the uncontaminated vector process $(\mathbf{Z}_r)$ defined similarly to $(\mathbf{X}_r)$. See Sarnaglia, Reisen & Bondon (2016b) for more details. We shall extend these results in Section 4.4 for the more general model in Equation 4.5 with missing values.

4.3 APS processes and missing data

We now extend some of the results of Dunsmuir & Robinson (1981a) to periodic processes. Consider the (random or deterministic) sequence $(\eta_t)_{t\geq 1}$. For each $\mathcal{S} \in \{1, 2, \ldots\}$, define the partial sums

$$\bar{\eta}_{\nu,\mathcal{S},N} := \frac{1}{N}\sum_{r=0}^{N-1} \eta_{r\mathcal{S}+\nu} \quad \text{and} \quad c_{\nu,\eta,\mathcal{S},N}(\tau) := \frac{1}{N}\sum_{r=r^*}^{N-1} \eta_{r\mathcal{S}+\nu}\eta_{r\mathcal{S}+\nu-\tau}, \quad 0 \leq \tau < N\mathcal{S} + \nu - \mathcal{S},$$

where r^* is the smallest $0 \leq r < N$ such that $r^*\mathcal{S} + \nu - \tau > 0$. The (η_t) is said to be Asymptotically $\mathrm{PS}_{\mathcal{S}}$ ($\mathrm{APS}_{\mathcal{S}}$) if, and only if, $\mathcal{S}$ is the smallest value in $\{1, 2, \ldots\}$ such that

$$\mu_{\nu,\eta} = \lim_{N\to\infty} \bar{\eta}_{\nu,\mathcal{S},N} \quad \text{and} \quad \varsigma_{\nu,\eta}(\tau) = \lim_{N\to\infty} c_{\nu,\eta,\mathcal{S},N}$$

exist almost surely (a.s.) for fixed τ. In this context, $\mu_{\nu,\eta}$ and $\gamma_{\nu,\eta}(\tau) := \varsigma_{\nu,\eta}(\tau) - \mu_{\nu,\eta}\mu_{\nu-\tau,\eta}$ are referred to as the asymptotic periodic mean and autocovariance functions of (η_t), respectively. One can show that these functions satisfy the periodic relations: $\mu_{\nu+\mathcal{S},\eta} = \mu_{\nu,\eta}$ and $\gamma_{\nu+\mathcal{S},\eta}(\tau) = \gamma_{\nu,\eta}(\tau)$. In addition, $\gamma_{\nu,\eta}(\tau) = \gamma_{\nu+\tau,\eta}(-\tau)$.

It can be shown that the sample autocovariance

$$\hat{\gamma}_{\nu,\eta,N}(\tau) = \frac{1}{N}\sum_{r=r^*}^{N-1}(\eta_{r\mathcal{S}+\nu} - \bar{\eta}_{\nu,N})(\eta_{r\mathcal{S}+\nu-\tau} - \bar{\eta}_{\nu-\tau,N}) \tag{4.7}$$

of the $\text{APS}_{\mathcal{S}}$ process (η_t) satisfies $\lim_{N\to\infty}\hat{\gamma}_{\nu,\eta,N}(\tau) = \gamma_{\nu,\eta}(\tau)$ a.s., for fixed τ.

Asymptotically periodically stationary processes naturally arise through amplitude modulation of PS processes in a similar manner as described in Parzen (1963). Let (X_t) be the AO process in (4.5) and (U_t) be a sequence which can be random or deterministic. We define (Y_t) as the process (X_t) amplitude modulated by (U_t), that is

$$Y_t = U_t \cdot X_t, \quad t \in \mathbb{Z}. \tag{4.8}$$

We now introduce necessary conditions on (U_t, X_t) to ensure that (Y_t) in (4.8) is an $\text{APS}_{\mathcal{S}}$ process.

(A1) U_s, X_t, $s, t \in \mathbb{Z}$, are independent;

(A2) (U_t) is an $\text{APS}_{\mathcal{S}}$ process with asymptotic mean and autocovariance functions given by $\mu_{\nu,U}$ and $\gamma_{\nu,U}(\tau)$, respectively, and satisfies $\text{E}(|U_t|^{2+\delta_U}) \leq K < \infty$, $t \geq 1$, for some $\delta_U > 0$. In this case, we define $\varsigma_{\nu,U}(\tau) = \gamma_{\nu,U}(\tau) + \mu_{\nu,U}\mu_{\nu-\tau,U}$;

(A3) (Z_t) is a $\text{PS}_{\mathcal{S}}$ process satisfying:

a) $Z_{r\mathcal{S}+\nu} = \sum_{j\geq 0}\psi_{\nu,j}\varepsilon_{r\mathcal{S}+\nu-j}$, where $\psi_{\nu,j} = \psi_{\nu+\mathcal{S},j}$ with $\sup_\nu\sum_{j\geq 0}|\psi_{\nu,j}| < \infty$ and (ε_t) is a zero mean white noise sequence with periodic variances $\sigma_t^2 = \text{Var}(\varepsilon_t) = \sigma_{t+\mathcal{S}}^2$;

b) $\text{E}(|\varepsilon_t|^{2+\delta_\varepsilon}) \leq K < \infty$, $t \geq 1$, for some $\delta_\varepsilon > 0$;

(A4) The sequence $(\mathcal{O}_t)$ satisfies $\text{E}(|\mathcal{O}_t|^{2+\delta_\mathcal{O}}) \leq K < \infty$, $t \geq 1$.

Condition **(A1)** is standard in literature and, as reinforced by Dunsmuir & Robinson (1981c), the more general case allowing dependence between (U_t) and (X_t) seems to be much harder to deal with. Moreover, the case wherein the actual process (X_t) and the missing mechanism do not affect each other seems to be natural in most (but not all) cases, such that **(A1)** does not seem to be very restrictive. Conditions **(A2)** and **(A3)** extend Assumptions A and B4 of Dunsmuir & Robinson (1981a) to the periodic case. The causal PARMA process satisfies **(A3**a**)**. **(A4)** bounds the magnitudes of the outliers. One can consider $\delta_U = \delta_\varepsilon = \delta_\mathcal{O} = \delta$ in **(A2)**, **(A3**b**)** and **(A4)** without any loss of generality, otherwise we can choose $\delta = \min(\delta_U, \delta_\varepsilon, \delta_\mathcal{O})$, so that these conditions remain valid.

Theorem 4.1. Let $(Y_t) \equiv (U_t \cdot X_t)$ be the amplitude modulated process in (4.8). Assume **(A1)**, **(A2)**, **(A3)** and **(A4)** hold. Then (Y_t) is an $\text{APS}_{\mathcal{S}}$ process with asymptotic periodic mean and autocovariance functions given, respectively, by $\mu_{\nu,Y} = 0$ and $\gamma_{\nu,Y}(\tau) = \varsigma_{\nu,U}(\tau)\gamma_{\nu,X}(\tau)$, $\nu = 1, \ldots, \mathcal{S}$.

Proof. The proof is given in Subsection 4.7.1 □

One can take $\mathcal{S} = 1$ in either **(A2)** or **(A3)** so that the following result still holds, i.e., amplitude modulating either stationary process with APS or PS with asymptotically stationary produces an APS process. Although it is not investigated here, we believe that alternative conditions to **(A3)** extending B1, B2, B3 or B5 of Dunsmuir & Robinson (1981a) can be formulated such that Theorem 4.1 still holds.

Note that, Theorem 4.1 and Equation 4.7 ensures that $\hat{\gamma}_{\nu,Y,N}(\tau)$ is a strongly consistent estimator of $\varsigma_{\nu,U}(\tau)\gamma_{\nu,X}(\tau)$. Since, by definition, **(A2)** imposes that $\lim_{N\to\infty} c_{\nu,U,N}(\tau) = \varsigma_{\nu,U}(\tau)$ a.s., Continuous Map Theorem ensures that $\hat{\gamma}_{\nu,X,N}(\tau) := \hat{\gamma}_{\nu,Y,N}(\tau)/c_{\nu,U,N}(\tau)$ is a strongly consistent estimator of $\gamma_{\nu,X}(\tau)$, when $\varsigma_{\nu,U}(\tau) \neq 0$. This conclusion is summarized in the following

Corollary 4.1. Suppose **(A1)**, **(A2)**, **(A3)** and **(A4)** hold and $\varsigma_{U,\nu}(\tau) \neq 0$. Define

$$\hat{\gamma}_{\nu,X,N}(\tau) = \frac{\hat{\gamma}_{\nu,Y,N}(\tau)}{c_{\nu,U,N}(\tau)}.$$

Then $\lim_{N\to\infty} \hat{\gamma}_{\nu,X,N}(\tau) = \gamma_{\nu,X}(\tau)$, a.s., for fixed τ.

Corollary 4.1 motivates the use of $\hat{\gamma}_{\nu,X,N}(\tau)$ to evaluate the covariance structure of the original process (X_t). However, we point out that $\hat{\gamma}_{\nu,X,N}(\tau)$ is not a positive semidefinite function, so that it is not an actual autocovariance function. Therefore, for example, its use to estimate autocorrelation coefficients could result in values with absolute value greater than 1, at least in small sample sizes.

Parzen (1961) has shown that asymptotically stationary time series possesses a generalized harmonic analysis. One can also introduce a similar treatment for $\text{APS}_{\mathcal{S}}$ processes in the sense of Hurd & Miamee (2007). We will not pursue this topic here. In turn, we shall consider the generalized harmonic analysis of the process $(\mathbf{Y}_r)$ defined by $\mathbf{Y}_r = [Y_{r\mathcal{S}+1}, \ldots, Y_{r\mathcal{S}+\mathcal{S}}]'$, $r \in \mathbb{Z}$, $\nu = 1, \ldots, \mathcal{S}$. The vector processes $(\mathbf{X}_r)$, $(\mathbf{U}_r)$ and $(\mathbf{Z}_r)$ are defined similarly to $(\mathbf{Y}_r)$. It can be shown that $(\mathbf{Y}_r)$ has an asymptotic (in a similar sense to the definition of $\text{APS}_{\mathcal{S}}$ processes) autocovariance matrix $\Gamma_{\mathbf{Y}}(\tau)$ with entries given by

$$\begin{aligned}
[\Gamma_{\mathbf{Y}}(\tau)]_{l,m} &= \gamma_{l,Y}(\tau\mathcal{S} + l - m) = \varsigma_{l,U}(\tau\mathcal{S} + l - m)\gamma_{l,X}(\tau\mathcal{S} + l - m) \\
&= (\gamma_{l,U}(\tau\mathcal{S} + l - m) + \mu_{l,U}\mu_{m,U})\gamma_{l,X}(\tau\mathcal{S} + l - m) \\
&= ([\Gamma_{\mathbf{U}}(\tau)]_{l,m} + [\boldsymbol{\mu}_{\mathbf{U}}\boldsymbol{\mu}'_{\mathbf{U}}]_{l,m})[\Gamma_{\mathbf{X}}(\tau)]_{l,m},
\end{aligned}$$

where $[\Gamma_{\mathbf{X}}(\tau)]_{l,m}$ is the l, m entry of the autocovariance matrix $\Gamma_{\mathbf{X}}(\tau)$ of $(\mathbf{X}_r)$ and $[\boldsymbol{\mu}_{\mathbf{U}}]_l$ and $[\Gamma_{\mathbf{U}}(\tau)]_{l,m}$ are, respectively, the lth component of the asymptotic mean vector $\boldsymbol{\mu}_{\mathbf{U}}$ and the l, m entry of the asymptotic autocovariance matrix $\Gamma_{\mathbf{U}}(\tau)$ of $(\mathbf{U}_r)$. In matrix notation the above equation simplifies to

$$\Gamma_{\mathbf{Y}}(\tau) = (\Gamma_{\mathbf{U}}(\tau) + \boldsymbol{\mu}_{\mathbf{U}}\boldsymbol{\mu}_{\mathbf{U}}') \odot \Gamma_{\mathbf{X}}(\tau), \tag{4.9}$$

where $\odot$ stands for the so-called Hadamard (entrywise or Schur) product of matrices. Now, the following matrix version of the Bochner representations

$$\Gamma_{\mathbf{U}}(\tau) = \int_0^{2\pi} e^{\mathbf{i}\tau\omega}\mathbf{F}_{\mathbf{U}}(\mathrm{d}\,\omega) \quad \text{and} \quad \Gamma_{\mathbf{X}}(\tau) = \int_0^{2\pi} e^{\mathbf{i}\tau\omega}\mathbf{f}_{\mathbf{X}}(\omega)\,\mathrm{d}\,\omega,$$

where $\mathbf{F}_{\mathbf{U}}$ and $\mathbf{f}_{\mathbf{X}}$ are the spectral distribution and the spectral density matrices of $(\mathbf{U}_r)$ and $(\mathbf{X}_r)$, respectively, can be combined with Equation 4.9 to obtain the following

Corollary 4.2. The asymptotic matrix spectrum of $(\mathbf{Y}_r)$ is given by

$$\mathbf{f}_{\mathbf{Y}}(\omega) = \mathbf{f}_{\mathbf{X}}(\omega) \odot \boldsymbol{\mu}_{\mathbf{U}}\boldsymbol{\mu}_{\mathbf{U}}' + \int_0^{2\pi} \mathbf{f}_{\mathbf{X}}(\omega - \varpi) \odot \mathbf{F}_{\mathbf{U}}(\mathrm{d}\,\varpi).$$

In addition, supposing that $\mathbf{F}_{\mathbf{U}}(\varpi)$ is absolutely continuous gives

$$\mathbf{f}_{\mathbf{Y}}(\omega) = \mathbf{f}_{\mathbf{X}}(\omega) \odot \boldsymbol{\mu}_{\mathbf{U}}\boldsymbol{\mu}_{\mathbf{U}}' + \int_0^{2\pi} \mathbf{f}_{\mathbf{X}}(\omega - \varpi) \odot \mathbf{f}_{\mathbf{U}}(\varpi)\,\mathrm{d}\,\varpi,$$

where $\mathbf{f}_{\mathbf{U}}$ is the spectral density matrix associated to $\mathbf{F}_{\mathbf{U}}(\varpi)$.

Corollary 4.2 extends to the vector case the convolution property of the scalar spectral density of amplitude modulated models. See for example Equation 2.1 in Dunsmuir & Robinson (1981c).

The analysis of time series with missing values is a natural example of application of amplitude modulate process. In the stationary case, if the original process (X_t) is not fully observed, the covariance matrix of the sample is not necessarily Toeplitz (DUNSMUIR; ROBINSON, 1981c). In the periodic scenario it is not block Toeplitz. These features prevent the straightforward use of Whittle methodologies in both cases, so that the Whittle methods of Sarnaglia, Reisen & Bondon (2015), Sarnaglia, Reisen & Bondon (2016a), Sarnaglia, Reisen & Bondon (2016b) can not be used directly on the observations of (X_t). However, one can still use the nice block Toeplitz properties if we replace the original time series, with its missing values, by a related sequence equally spaced which can be handled in a similar way as periodically stationary processes. This particular sequence is given by the amplitude modulated process (Y_t) in (4.8), wherein the (U_t) denotes an indicator sequence with $U_t = 1$ or 0 if the observation X_t is observed or missing, respectively. In this framework (U_t) will be referred to as the not missing indicator process. Roughly speaking, the (U_t) inputs zeroes in place of missing observations in order to produce an

artificial "complete" time series observed from (Y_t). This approach can be used in the cases where the missing data is produced in a deterministic fashion, such as the regular $A-B$ sampling (DUNSMUIR; ROBINSON, 1981a), or driven by some random schematic, as long as they satisfy Assumption **(A2)**.

4.4 PARMA estimation with missing data

From now on, $(X_t)_{t=1}^n$ is a (possibly contaminated) time series generated by the process (X_t) in (4.5). This paper is concerned with the case where the sample $(X_t)_{t=1}^n$ is not completely observable. Therefore, we consider the amplitude modulated model in (4.8), by defining the not missing indicators as $(U_t)_{t=1}^n$, where $U_t = 1$ or 0 if the observation X_t is observed or missing, respectively. The amplitude modulated time series is defined as $(Y_t)_{t=1}^n$ and given by $Y_t = U_t \cdot X_t$, $t = 1, \ldots, n$. In the case where the generating process (X_t) in (4.5) is contaminated, the (hidden) uncontaminated time series is denoted by $(Z_t)_{t=1}^n$. For simplicity, we assume the sample size $n = N\mathcal{S}$, $N \in \{1, 2, \ldots\}$, such that N full cycles will be analysed. The $\mathcal{S}$-dimensional vector time series $(\mathbf{X}_r)_{r=0}^{N-1}$ is defined by $[\mathbf{X}_r]_\nu = X_{r\mathcal{S}+\nu}$, $\nu = 1, \ldots, \mathcal{S}$, $r = 0, 1, \ldots, N-1$. The $(\mathbf{U}_r)_{r=0}^{N-1}$, $(\mathbf{Y}_r)_{r=0}^{N-1}$ and $(\mathbf{Z}_r)_{r=0}^{N-1}$ are defined similarly. The parameters of the PARMA model are grouped in the vector φ. We assume that the true parameter vector φ_0 lies in the parameter space $\mathcal{P}$ whose points satisfy the identifiability assumptions of Sarnaglia, Reisen & Bondon (2016a).

4.4.1 Asymptotic Whittle likelihood estimator

We first consider the uncontaminated case, where $(X_t) \equiv (Z_t)$ and $\mathbf{f}_\mathbf{X}(\omega) = \mathbf{f}_\mathbf{Y}(\omega)$, $\omega \in [0, 2\pi)$. Consider the complete vector time series $(\mathbf{Y}_r)_{r=0}^{N-1}$. In light of Corollary 4.2 and in analogy with the multivariate version of Whittle's approximation of the gaussian likelihood, one could estimate φ_0 by choosing φ to minimize

$$\ell(\varphi) = \frac{1}{N'} \sum_{j=1}^{N'} \left\{ \operatorname{tr}\left[\mathbf{f}_\mathbf{Y}^{-1}(\omega_j; \varphi) I_\mathbf{Y}(\omega_j) \right] + \log\det \mathbf{f}_\mathbf{Y}(\omega_j; \varphi) \right\}, \tag{4.10}$$

where $\mathbf{f}_\mathbf{Y}(\omega_j; \varphi)$ is the asymptotic matrix spectrum of $(\mathbf{Y}_r)$ in Corollary 4.2 evaluated at the Fourier frequencies $\omega_j = \frac{2\pi j}{N}$, $j = 1, \ldots, N'$, $N' = \lfloor (N-1)/2 \rfloor$ is the greatest integer smaller than or equal to $(N-1)/2$, the term φ is now being used in $\mathbf{f}_\mathbf{Y}$ to reinforce their intrinsic dependency, and

$$W_\mathbf{Y}(\omega) = \frac{1}{\sqrt{N}} \sum_{r=0}^{N-1} \mathbf{Y}_r e^{-\mathbf{i}r\omega} \quad \text{and} \quad I_\mathbf{Y}(\omega) = W_\mathbf{Y}(\omega) W_\mathbf{Y}(\omega)^{\mathcal{H}}$$

are, respectively, the discrete Fourier transform and the periodogram of $(\mathbf{Y}_r)_{r=0}^{N-1}$ at ω. At the Fourier frequency ω_j, $W_\mathbf{Y}(\omega_j)$ can be efficiently obtained through the multivariate version of the fast Fourier transform algorithm.

Unfortunately, the expression in Equation 4.10 may be inaccessible since $\mathbf{F_U}$ is usually unknown. We follow Dunsmuir & Robinson (1981c) and circumvent this problem by approximating the spectral density matrix $\mathbf{f_Y}(\omega_j;\varphi)$ by

$$\begin{aligned}\hat{\mathbf{f}}_{\mathbf{Y}}(\omega_j;\varphi) &= \mathbf{f_X}(\omega_j;\varphi)\odot\bar{\mathbf{U}}\bar{\mathbf{U}}' + \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{f_X}(\omega_j-\omega_k;\varphi)\odot I_{\mathbf{U}^*}(\omega_k)\\ &= \frac{1}{N}\sum_{k=0}^{N-1}\mathbf{f_X}(\omega_j-\omega_k;\varphi)\odot I_{\mathbf{U}}(\omega_k),\end{aligned} \tag{4.11}$$

where $I_{\mathbf{U}}(\omega_k)$ and $I_{\mathbf{U}^*}(\omega_k)$ are, respectively, the periodograms of the vector time series $(\mathbf{U}_r)_{r=0}^{N-1}$ and its centralized counterpart $(\mathbf{U}_r^*)_{r=0}^{N-1}$, wherein $\mathbf{U}_r^* = \mathbf{U}_r - \bar{\mathbf{U}}$, $r = 0, 1, \ldots, N-1$, and $\bar{\mathbf{U}} = \frac{1}{N}\sum_r \mathbf{U}_r$. Now, $\mathbf{f_Y}$ can be replaced by $\hat{\mathbf{f}}_{\mathbf{Y}}$ in $\ell(\varphi)$ (Equation 4.10) which gives

$$\hat{\ell}(\varphi) = \frac{1}{N'}\sum_{j=1}^{N'}\left\{\operatorname{tr}\left[\hat{\mathbf{f}}_{\mathbf{Y}}^{-1}(\omega_j;\varphi)I_{\mathbf{Y}}(\omega_j)\right] + \log\det\hat{\mathbf{f}}_{\mathbf{Y}}(\omega_j;\varphi)\right\}. \tag{4.12}$$

We propose to estimate the parameter vector φ by

$$\hat{\varphi} = \underset{\varphi\in\mathcal{P}}{\operatorname{argmin}}\,\hat{\ell}(\varphi).$$

This particular approximation $\hat{\ell}$ is prefered since it requires no prior knowledge of $\mathbf{F_U}$. The computation of the "Hadarmard convolution" required in (4.11) can be done efficiently through the fast Fourier transform algorithm applied entrywise. Henceforth, $\hat{\varphi}$ will be called Asymptotic Whittle Likelihood Estimator (AWLE). Asymptotic properties of $\hat{\varphi}$ are not easy to be obtained and deserve a paper uniquely devoted to this subject. We intend to approach this in a forthcoming work.

4.4.1.1 Impact of additive outliers

We shall see in Section 4.5 that the AWLE has a small loss of efficiency for incomplete time series observed from the AO process (X_t) when the probability $\zeta = 0$, that is when $(X_t) \equiv (Z_t)$. We now investigate some theoretical aspects which discourage the use of AWLE when $\zeta > 0$.

Observe that

$$\gamma_{t,X}(\tau) = \operatorname{Cov}(X_t, X_{t-\tau}) = \operatorname{Cov}(\mathcal{B}_t\mathcal{O}_t + Z_t, \mathcal{B}_{t-\tau}\mathcal{O}_{t-\tau} + Z_{t-\tau}) = \sigma_{\mathcal{O}}^2\zeta\,\mathbb{1}_{\{\tau=0\}} + \gamma_{t,Z}(\tau),$$

where $\sigma_{\mathcal{O}}^2 = \operatorname{Var}(\mathcal{O}_t) = \mathrm{E}(\mathcal{O}_t^2)$, which exists from **(A4)** and is constant by definition of the sequence $(\mathcal{O}_t)$. The above equation generalizes the result in Sarnaglia, Reisen & Lévy-Leduc (2010) for the general AO process in (4.5). Now we have that

$$\Gamma_{\mathbf{X}}(\tau) = \mathbf{D}_{\mathcal{O}}\,\mathbb{1}_{\{\tau=0\}} + \Gamma_{\mathbf{Z}}(\tau),$$

where $\mathbf{D}_\mho$ is diagonal matrix with entries given by $[\mathbf{D}_\mho]_{l,m} = \zeta\sigma_\mho^2 \mathbb{1}_{\{l=m\}}$. Therefore, the spectral matrix of $(\mathbf{X}_r)$ is given by

$$\mathbf{f}_\mathbf{X}(\omega) = \mathbf{f}_\mathbf{Z}(\omega) + \frac{1}{2\pi}\mathbf{D}_\mho. \tag{4.13}$$

Replacing Equation 4.13 in $\mathbf{f}_\mathbf{Y}$ given in Corollary 4.2 gives

$$\begin{aligned}
\mathbf{f}_\mathbf{Y}(\omega) &= \left[\frac{1}{2\pi}\mathbf{D}_\mho + \mathbf{f}_\mathbf{Z}(\omega)\right] \odot \boldsymbol{\mu}_\mathbf{U}\boldsymbol{\mu}_\mathbf{U}' + \int_0^{2\pi} \left[\frac{1}{2\pi}\mathbf{D}_\mho + \mathbf{f}_\mathbf{Z}(\omega-\varpi)\right] \odot \mathbf{F}_\mathbf{U}(\mathrm{d}\,\varpi) \\
&= \frac{1}{2\pi}\mathbf{D}_\mho \odot \left[\boldsymbol{\mu}_\mathbf{U}\boldsymbol{\mu}_\mathbf{U}' + \int_0^{2\pi} \mathbf{F}_\mathbf{U}(\mathrm{d}\,\varpi)\right] + \mathbf{f}_\mathbf{Z}(\omega) \odot \boldsymbol{\mu}_\mathbf{U}\boldsymbol{\mu}_\mathbf{U}' + \int_0^{2\pi} \mathbf{f}_\mathbf{Z}(\omega-\varpi) \odot \mathbf{F}_\mathbf{U}(\mathrm{d}\,\varpi) \\
&= \frac{1}{2\pi}\mathbf{D}_\mho \odot \mathbf{C}_\mathbf{U}(0) + \mathbf{f}_\mathbf{Z}(\omega) \odot \boldsymbol{\mu}_\mathbf{U}\boldsymbol{\mu}_\mathbf{U}' + \int_0^{2\pi} \mathbf{f}_\mathbf{Z}(\omega-\varpi) \odot \mathbf{F}_\mathbf{U}(\mathrm{d}\,\varpi),
\end{aligned}$$

where $\mathbf{C}_\mathbf{U}(\tau)$ is the asymptotic autocovariance matrix of $(\mathbf{U}_r)$ which has entries given by $[\mathbf{C}_\mathbf{U}(\tau)]_{l,m} = \varsigma_{l,U}(\tau\mathcal{S} + l - m)$, $l, m = 1, \ldots, \mathcal{S}$, with $\varsigma_{l,U}(\tau)$ given in **(A2)**. In addition,

$$\begin{aligned}
\hat{\mathbf{f}}_\mathbf{Y}(\omega_j;\varphi) &= \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{f}_\mathbf{X}(\omega_j - \omega_k;\varphi) \odot I_\mathbf{U}(\omega_k) \\
&= \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{f}_\mathbf{Z}(\omega_j - \omega_k;\varphi) \odot I_\mathbf{U}(\omega_k) + \frac{1}{2\pi}\mathbf{D}_\mho \odot \left(\frac{1}{N}\sum_{k=0}^{N-1} I_\mathbf{U}(\omega_k)\right) \\
&= \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{f}_\mathbf{Z}(\omega_j - \omega_k;\varphi) \odot I_\mathbf{U}(\omega_k) + \frac{1}{2\pi}\mathbf{D}_\mho \odot \hat{\mathbf{C}}_\mathbf{U}(0),
\end{aligned}$$

where $\hat{\mathbf{C}}_\mathbf{U}(\tau) = \frac{1}{N}\sum_r \mathbf{U}_r\mathbf{U}_{r-\tau}'$. These results are summarized in the following

Corollary 4.3. The asymptotic matrix spectrum of $(\mathbf{Y}_r)$ is given by

$$\mathbf{f}_\mathbf{Y}(\omega) = \frac{1}{2\pi}\mathbf{D}_\mho \odot \mathbf{C}_\mathbf{U}(0) + \mathbf{f}_\mathbf{Z}(\omega) \odot \boldsymbol{\mu}_\mathbf{U}\boldsymbol{\mu}_\mathbf{U}' + \int_0^{2\pi} \mathbf{f}_\mathbf{Z}(\omega-\varpi) \odot \mathbf{F}_\mathbf{U}(\mathrm{d}\,\varpi)$$

and

$$\hat{\mathbf{f}}_\mathbf{Y}(\omega_j;\varphi) = \frac{1}{2\pi}\mathbf{D}_\mho \odot \hat{\mathbf{C}}_\mathbf{U}(0) + \frac{1}{N}\sum_{k=0}^{N-1} \mathbf{f}_\mathbf{Z}(\omega_j - \omega_k;\varphi) \odot I_\mathbf{U}(\omega_k). \tag{4.14}$$

Sarnaglia, Reisen & Bondon (2016b) have shown that, for the AO model with the contamination given in (4.6), the matritial classical periodogram $I_\mathbf{X}$ is a r.v. whose expected value is impacted by the outliers in the same amount as the spectral density. One can easily extend this result for the more general model in (4.5). Observe that, for $\omega \in (0, \pi)$,

$$\begin{aligned}
\mathrm{E}(I_\mathbf{X}(\omega)) = \mathrm{E}(W_X(\omega)W_X(\omega)^{\mathcal{H}}) &= \frac{1}{N}\sum_{r=0}^{N-1}\sum_{s=0}^{N-1} \mathrm{E}(\mathbf{X}_r\mathbf{X}_s)e^{-\mathbf{i}\omega(r-s)} \\
&= \sum_{|\tau|<N} \left(1 - \frac{|\tau|}{N}\right) \Gamma_\mathbf{X}(\tau)e^{-\mathbf{i}\omega\tau}.
\end{aligned}$$

Therefore, $\lim_{N\to\infty} \mathrm{E}(I_{\mathbf{X}}(\omega)) = \mathbf{f}_{\mathbf{X}}(\omega) = \mathbf{f}_{\mathbf{Z}}(\omega) + \frac{1}{2\pi}\mathbf{D}_{\mathcal{O}}$, such that it is expected that the periodogram $I_{\mathbf{X}}$ will be impacted by AO in the same amount as the spectral density $\mathbf{f}_{\mathbf{X}}$.

In order to extend this result to the amplitude modulated process, we have to impose the following additional assumption, which is satisfied for the not missing indicator process (U_t) described at the end of Section 4.3:

(A2') $|U_t| \leq U$, $t \in \mathbb{Z}$, and $\mathrm{E}(U) < \infty$

Theorem 4.2. Let $I_{\mathbf{Y}}(\omega)$ be the periodogram of $(\mathbf{Y}_r)_{r=0}^{N-1}$. Then, under **(A1)**, **(A2)** and **(A2')**, **(A3)** and **(A4)**, we have that, for $\omega \in (0, \pi)$,

$$\mathrm{E}(I_{\mathbf{Y}}(\omega)) \to \mathbf{f}_{\mathbf{Y}}(\omega) = \frac{1}{2\pi}\mathbf{D}_{\mathcal{O}} \odot \mathbf{C}_{\mathbf{U}}(0) + \mathbf{f}_{\mathbf{Z}}(\omega) \odot \boldsymbol{\mu}_{\mathbf{U}}\boldsymbol{\mu}_{\mathbf{U}}' + \int_0^{2\pi} \mathbf{f}_{\mathbf{Z}}(\omega - \varpi) \odot \mathbf{F}_{\mathbf{U}}(\mathrm{d}\,\varpi), \quad (4.15)$$

as $N \to \infty$.

Proof. The proof is given in Subsection 4.7.2 □

In Equation 4.14, the term $\frac{1}{2\pi}\mathbf{D}_{\mathcal{O}} \odot \hat{\mathbf{C}}_{\mathbf{U}}(0) \to \frac{1}{2\pi}\mathbf{D}_{\mathcal{O}} \odot \mathbf{C}_{\mathbf{U}}(0)$, $N \to \infty$, a.s., and quantify the AO contribution to $\hat{\mathbf{f}}_{\mathbf{Y}}$. At first, one may model the outliers influence directly by specifying this additional term in $\hat{\ell}(\varphi)$. However, the presence of outliers is unknown in practical situations and the main reason in specifying them in the modelling is to remove its effect, not necessarily to estimate it. Therefore, the inclusion of this additional quantity could result in an unnecessary overparametrization, so that we neglect it in $\hat{\ell}(\varphi)$. Nevertheless, from Equation 4.15, it is expected that, in a contaminated time series scenario, the periodogram $I_{\mathbf{Y}}(\omega)$ presents the AO effect, so that maximizing $\hat{\ell}(\varphi)$ disregarding the term $\frac{1}{2\pi}\mathbf{D}_{\mathcal{O}} \odot \hat{\mathbf{C}}_{\mathbf{U}}(0)$ will increase bias of the estimates. This motivates us to replace $I_{\mathbf{Y}}(\omega)$ in $\hat{\ell}(\varphi)$ by a robust estimator of

$$\mathbf{f}_{\mathbf{Z}}(\omega) \odot \boldsymbol{\mu}_{\mathbf{U}}\boldsymbol{\mu}_{\mathbf{U}}' + \int_0^{2\pi} \mathbf{f}_{\mathbf{Z}}(\omega - \varpi) \odot \mathbf{F}_{\mathbf{U}}(\mathrm{d}\,\varpi), \quad (4.16)$$

which is the proposal of the next subsection.

4.4.2 M-asymptotic Whittle likelihood estimator

We propose the use of the multivariate M-periodogram of Sarnaglia, Reisen & Bondon (2016b) to estimate the spectral density in Equation 4.16. Let $C_r(\omega) = [\cos(r\omega), \sin(r\omega)]'$, $\omega \in (0, \pi)$, $r = 0, 1, \ldots, N-1$. At the elementary frequencies, the multivariate periodogram $I_{\mathbf{Y}}(\omega_j)$, $j = 1, \ldots, N'$, is related to the least square estimator of the multivariate multiple linear model

$$\mathbf{Y}_r = \boldsymbol{\beta}(\omega_j)' C_r(\omega_j) + \boldsymbol{\xi}_r, \quad r = 0, 1, \ldots, N-1, \quad (4.17)$$

where $\boldsymbol{\beta}(\omega_j) = [\beta_1(\omega_j), \ldots, \beta_{\mathcal{S}}(\omega_j)]$ is a $(2 \times \mathcal{S})$ coefficient matrix and $\boldsymbol{\xi}_r = [\xi_{r,1}, \ldots, \xi_{r,\mathcal{S}}]'$, $r = 0, 1, \ldots, N-1$, is a random error term. In this framework, the 2-dimensional vector

$\beta_\nu(\omega_j)$ can be seen as the impact of the jth harmonic in the νth time series $([\mathbf{Y}_r]_\nu)_{r=0}^{N-1}$, $\nu = 1, \ldots, \mathcal{S}$. See Sarnaglia, Reisen & Bondon (2016b) for more details.

In order to produce a robust estimator of the spectral density, Sarnaglia, Reisen & Bondon (2016b) propose the use of M-regression techniques to estimate the model in (4.17). More specifically, let $\boldsymbol{\rho} : \mathbb{R}^{\mathcal{S}} \mapsto \mathbb{R}$ be a suitable discrepancy function, one defines the M-estimator of $\boldsymbol{\beta}(\omega_j)$ by

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{\rho}}(\omega_j) = \underset{\boldsymbol{\beta}\in\mathbb{R}^{2\times\mathcal{S}}}{\operatorname{argmin}} \left\{ \sum_{r=0}^{N-1} \boldsymbol{\rho}\left(\mathbf{Y}_r - \boldsymbol{\beta}' C_r(\omega_j)\right) \right\}, \quad j = 1, \ldots, N'. \tag{4.18}$$

Equivalently, one can obtain $\hat{\boldsymbol{\beta}}_{\boldsymbol{\rho}}(\omega_j)$ as the zeroes of

$$\sum_{r=0}^{N-1} \boldsymbol{\psi}\left(\mathbf{Y}_r - \hat{\boldsymbol{\beta}}_{\boldsymbol{\rho}}(\omega_j)' C_r(\omega_j)\right) C_r(\omega_j)' = \mathbf{0}_{\mathcal{S}\times 2}, \tag{4.19}$$

where, in this context, $\boldsymbol{\psi} = \boldsymbol{\psi}(x_1, \ldots, x_{\mathcal{S}})$ is the $\mathcal{S}$-dimensional vector gradient function related to $\boldsymbol{\rho}$ with elements given by $[\boldsymbol{\psi}(x_1, \ldots, x_{\mathcal{S}})]_i = \frac{\partial}{\partial x_i}\boldsymbol{\rho}(x_1, \ldots, x_{\mathcal{S}})$. Observe that the classical least squares estimator is a particular case of M-estimator where $\boldsymbol{\rho}(x_1, \ldots, x_{\mathcal{S}}) = \sum_{i=1}^{\mathcal{S}} x_i^2$ and $\boldsymbol{\psi}(x_1, \ldots, x_{\mathcal{S}}) = 2[x_1, \ldots, x_{\mathcal{S}}]'$. Finally, at the elementary frequencies, the M-transform and M-periodogram of $(\mathbf{Y}_r)_{r=0}^{N-1}$ relative to $\boldsymbol{\rho}$ are, respectively, defined as

$$W_{\mathbf{Y},\boldsymbol{\rho}}(\omega_j) = \sqrt{N/8\pi}\left([\hat{\boldsymbol{\beta}}_{\boldsymbol{\rho}}(\omega_j)]_{1\bullet} - \mathbf{i}[\hat{\boldsymbol{\beta}}_{\boldsymbol{\rho}}(\omega_j)]_{2\bullet}\right) \tag{4.20}$$

and

$$I_{\mathbf{Y},\boldsymbol{\rho}}(\omega_j) = W_{\mathbf{Y},\boldsymbol{\rho}}(\omega_j) W_{\mathbf{Y},\boldsymbol{\rho}}(\omega_j)^{\mathcal{H}}, \tag{4.21}$$

where $j = 1, \ldots, N'$.

Now, we define the M-Asymptotic Whittle Likelihood Estimator (M-AWLE) of PARMA parameters as

$$\hat{\varphi}_{\boldsymbol{\rho}} = \underset{\varphi\in\mathcal{P}}{\operatorname{argmin}}\, \hat{\ell}_{\boldsymbol{\rho}}(\varphi),$$

where

$$\hat{\ell}_{\boldsymbol{\rho}}(\varphi) = \frac{1}{N'}\sum_{j=1}^{N'} \left\{ \operatorname{tr}\left[\hat{\mathbf{f}}_{\mathbf{Y}}^{-1}(\omega_j;\varphi) I_{\mathbf{Y},\boldsymbol{\rho}}(\omega_j)\right] + \log\det \hat{\mathbf{f}}_{\mathbf{Y}}(\omega_j;\varphi) \right\}.$$

As we shall see in the next section, $\hat{\varphi}_{\boldsymbol{\rho}}$ keeps the robust features of M-estimators and has small loss of efficiency compared to $\hat{\varphi}$ in the uncontaminated time series scenario. Obtaining asymptotics for $\hat{\varphi}_{\boldsymbol{\rho}}$ seem to be very challenging and will be considered in a forthcoming paper.

A natural choice for the $\mathcal{S}$-variate discrepancy function is given by $\boldsymbol{\rho}(x_1, \ldots, x_{\mathcal{S}}) = \sum_{i=1}^{\mathcal{S}} \rho(x_i)$, where ρ is a convenient function, such as $\rho(x) = \rho_\kappa(x) = |x|^\kappa$ or the well-known Huber function (HUBER, 1964) given by

$$\rho(x) = \rho_\delta(x) = \begin{cases} \frac{1}{2}z^2 & , |z| \le \delta, \\ \delta(|z| - \frac{\delta}{2}) & , |z| > \delta. \end{cases} \tag{4.22}$$

The choice of the tunning parameters such as κ, δ is quite important and provides a interchange between robustness and efficiency of the M-estimators.

4.5 Monte Carlo study

In this section we investigate the finite sample behaviour of the proposed estimator. We consider three cases: (i) complete uncontaminated which will be considered as a benchmark; (ii) incomplete and uncontaminated time series; and (iii) incomplete and contaminated data.

We generate $M = 1000$ replicates of each scenario, where the data will be generated from the AO process in (4.5) with contamination sequence $(\mathcal{O}_t)$ given in (4.6), where $\zeta \in \{0, 0.01\}$ and $\varrho = 10$. The hidden process (Z_t) is a PARMA$(1,1)_{\mathcal{S}}$ model, with $\mathcal{S} = 2$ and parameters given in Table 4.1, which are set in order to evaluate the influence of closeness of non causality and/or non invertibility regions on the estimators. Other coefficient configurations were also considered and presented similar results. They are not displayed here to save space, but they are available upon request. The missing data are chosen through independent Bernoulli trials with success (the data be not missing) probability 0.95. We consider the sample sizes $n = N\mathcal{S} = 300, 800$ ($N = 150, 400$, respectively).

We use the $\mathcal{S}$-variate discrepancy function $\boldsymbol{\rho}(x_1, \ldots, x_{\mathcal{S}}) = \sum_{i=1}^{\mathcal{S}} \rho_\delta(x_i)$, where $\rho_\delta(x)$ is the Huber function defined in (4.22). As usual in literature, the tuning parameter is set as $\delta = 1.345$, which ensure that the M-estimator is 95% as efficient as the least squares estimator for univariate multiple linear models with i.i.d. gaussian white noise. It can be even more efficient in many other cases, as the simulation results will show.

Table 4.1 – Simulated models.

	Period 1			Period 2			Roots in (4.4)	
Model	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2	ϑ_ϕ	ϑ_θ
1	−0.2	0.0	1.0	−0.5	0.0	1.0	0.1	0.0
2	−0.2	−0.5	1.0	−0.5	−0.2	1.0	0.1	0.1
3	−1.0	0.0	1.0	−0.5	0.0	1.0	0.5	0.0
4	−1.0	−0.5	1.0	−0.5	−0.2	1.0	0.5	0.1

We evaluate the finite sample performance of the estimators through empirical Root Mean Square Error (RMSE). The BIAS and the Standard Error were also calculated, however they are not shown here to save space, but they are available upon request. The RMSE of the AWLE and M-AWLE are presented in Tables 4.2, 4.3, 4.4 and 4.5. Values with "$*$" refer to the empirical RMSE of the M-AWLE estimates. Boldface numbers represent comparable small RMSE.

Table 4.2 – RMSE of $\hat{\varphi}$ and $\hat{\varphi}_{\boldsymbol{\rho}}$, with "$*$", for Model 1 with $\vartheta_\phi = 0.1$ and $\vartheta_\theta = 0.0$.

Case	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
	300	**0.070**; **0.075***	—	**0.128**; 0.152*	**0.081**; **0.089***	—	**0.114**; 0.166*
(i)	800	**0.047**; **0.049***	—	**0.072**; 0.097*	**0.051**; **0.058***	—	**0.064**; 0.101*
	300	**0.074**; **0.079***	—	**0.129**; 0.155*	**0.084**; **0.090***	—	**0.118**; 0.162*
(ii)	800	**0.050**; **0.052***	—	**0.075**; 0.092*	**0.055**; **0.063***	—	**0.067**; 0.098*
	300	0.148; **0.081***	—	1.309; **0.165***	0.272; **0.095***	—	1.384; **0.209***
(iii)	800	0.127; **0.056***	—	1.136; **0.116***	0.276; **0.070***	—	1.253; **0.138***

Table 4.3 – RMSE of $\hat{\varphi}$ and $\hat{\varphi}_{\boldsymbol{\rho}}$, with "$*$", for Model 2 with $\vartheta_\phi = 0.1$ and $\vartheta_\theta = 0.1$.

Case	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
	300	**0.328**; 0.418*	**0.334**; 0.432*	**0.112**; 0.138*	**0.400**; 0.551*	**0.415**; 0.569*	**0.122**; 0.143*
(i)	800	**0.167**; **0.174***	**0.169**; **0.177***	**0.071**; **0.087***	**0.173**; **0.186***	**0.183**; **0.196***	**0.065**; **0.101***
	300	**0.372**; 0.423*	**0.380**; 0.438*	**0.117**; 0.140*	**0.464**; 0.633*	**0.478**; 0.649*	**0.124**; 0.139*
(ii)	800	**0.176**; **0.185***	**0.178**; **0.191***	**0.071**; **0.082***	**0.181**; **0.201***	**0.189**; **0.213***	**0.069**; **0.097***
	300	1.122; **0.461***	1.119; **0.479***	1.332; **0.163***	1.195; **0.609***	1.240; **0.630***	1.312; **0.158***
(iii)	800	0.659; **0.198***	0.662; **0.201***	1.117; **0.103***	0.592; **0.206***	0.630; **0.221***	1.209; **0.130***

Table 4.4 – RMSE of $\hat{\varphi}$ and $\hat{\varphi}_{\boldsymbol{\rho}}$, with "$*$", for Model 3 with $\vartheta_\phi = 0.5$ and $\vartheta_\theta = 0.0$.

Case	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
	300	**0.070**; **0.089***	—	**0.126**; 0.207*	**0.054**; **0.063***	—	**0.118**; 0.164*
(i)	800	**0.044**; **0.054***	—	**0.070**; 0.161*	**0.031**; **0.044***	—	**0.063**; 0.134*
	300	**0.077**; **0.100***	—	**0.144**; 0.200*	**0.054**; **0.063***	—	**0.134**; 0.167*
(ii)	800	**0.044**; **0.054***	—	**0.094**; 0.173*	**0.031**; **0.044***	—	**0.070**; 0.130*
	300	0.425; **0.118***	—	1.632; **0.329***	0.170; **0.070***	—	1.392; **0.237***
(iii)	800	0.434; **0.077***	—	1.648; **0.307***	0.154; **0.054***	—	1.243; **0.202***

Table 4.5 – RMSE of $\hat{\varphi}$ and $\hat{\varphi}_{\boldsymbol{\rho}}$, with "$*$", for Model 4 with $\vartheta_\phi = 0.5$ and $\vartheta_\theta = 0.1$.

Case	n	$\phi_{1,1}$	$\theta_{1,1}$	σ_1^2	$\phi_{2,1}$	$\theta_{2,1}$	σ_2^2
	300	**0.212**; **0.228***	**0.228**; **0.239***	**0.104**; 0.152*	**0.126**; **0.134***	**0.141**; **0.148***	**0.114**; 0.144*
(i)	800	**0.118**; **0.130***	**0.130**; **0.141***	**0.070**; 0.126*	**0.077**; **0.083***	**0.083**; **0.089***	**0.063**; 0.094*
	300	**0.235**; **0.241***	**0.253**; **0.255***	**0.109**; 0.148*	**0.130**; **0.137***	**0.148**; **0.151***	**0.118**; 0.141*
(ii)	800	**0.126**; **0.148***	**0.141**; **0.164***	**0.070**; 0.118*	**0.083**; **0.094***	**0.089**; **0.094***	**0.063**; 0.089*
	300	1.142; **0.251***	1.214; **0.263***	1.347; **0.190***	0.282; **0.144***	0.296; **0.154***	1.386; **0.173***
(iii)	800	0.341; **0.158***	0.409; **0.176***	1.237; **0.170***	0.161; **0.094***	0.204; **0.100***	1.080; **0.122***

For complete and uncontaminated time series, Case (i), in general, both estimators present similar performance for the AR and MA coefficients. Regarding the estimation of white noise variances, AWLE seems to be more accurate. For both estimators, the RMSE seems to present an overall decrease with sample size. The presence of MA structure increase the RMSE of AR coefficients for both estimators, while the white noise variances seem to be unaffected.

In the scenario where the time series has missing values and no outliers, Case (ii), all conclusions in the previous paragraph still hold. However, we observe an overall increase of RMSE which can be due to the fact that the original sample is not used in this case.

The results of Case (iii), where the data set is incomplete and contaminated by additive outliers, are incontestable and show the remarkable superiority of M-AWLE over the AWLE, while the latter is totally corrupted by the atypical observations, the former presents RMSE almost unchanged compared to the uncontaminated data.

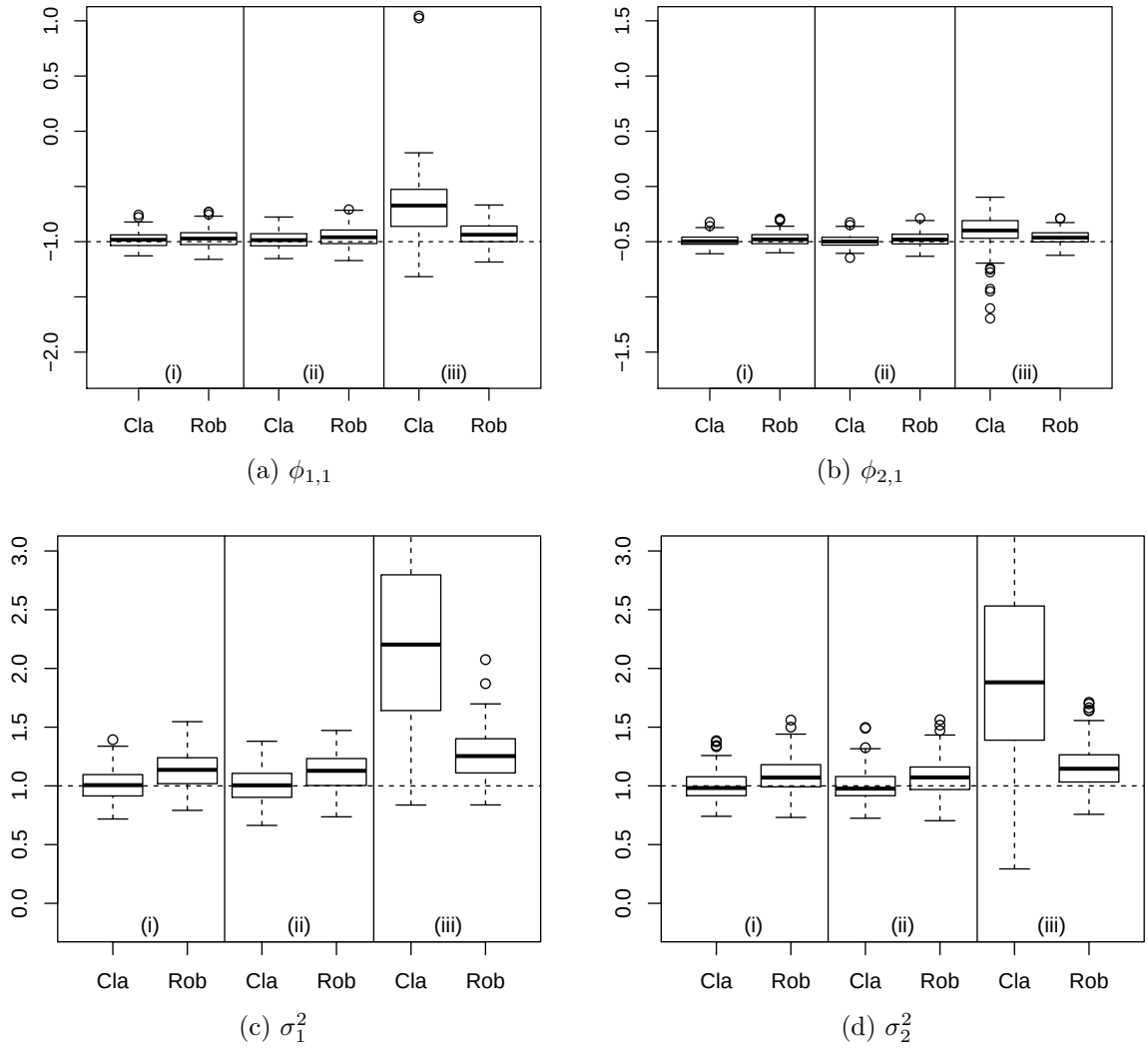
We also display in Figure 4.1 the empirical distributions of the estimates provided by the MLE, the WLE and the RWLE for Model 3. It is clear the AO effect in the MLE and the WLE, while the RWLE remains almost unchanged.

Therefore, this simulation study shows that AWLE should be used in the case where the user is sure about the absence of outliers. In contrast, in case of suspicion of the presence of the aberrant data, the user should rather use the M-AWLE.

4.6 Application

In this section we analyze the daily mean concentrations of sulfur dioxide (SO_2). The raw series was observed from 1st January 2005 to 31st December 2009 at the monitoring station of Environment and Water Resources State Institute located in Vitória downtown, Brazil. We prefer this data set since it is completely available and order identification can be easily performed. For time series with missing data, identification can be handled by using the autocovariance estimator in Corollary 4.1. Following Sarnaglia, Reisen & Bondon (2016a), the first 1603 observations were considered in the model learning (estimation) stage and the remaining 223 observations were reserved for the out-of-sample forecast study. Figure 4.2a displays the SO_2 time series and the learning and prediction values. Because data is completely observed, the missing data scenario is artificially generated by a binomial random sample with size 1603 and missing probability = 0.05, see Figure 4.2b for the stretch between 2005 and 2006. The observed percentual of missing in the whole sample is 5.05%.

Since data is daily collected, the PARMA model with period $\mathcal{S} = 7$ seems to be appropriated to fit the series. As shown in Sarnaglia, Reisen & Bondon (2016a), the sample Periodic Autocorrelation and Partial Autocorrelation functions indicate a PARMA model with orders $p_\nu = 1$, $\forall \nu$, and $q_\nu = 1, 1, 1, 1, 0, 0, 1$. Note that, this model satisfies



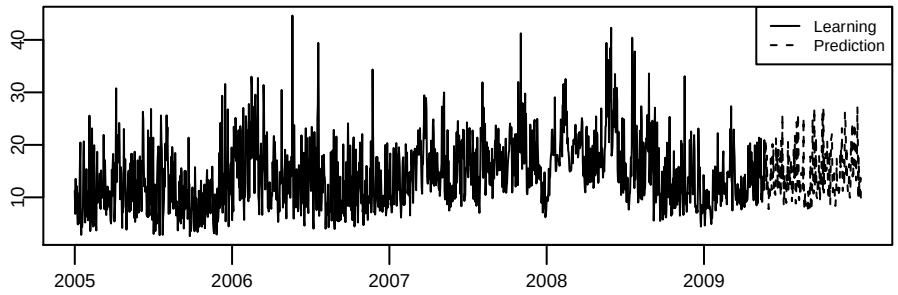
(a) $\phi_{1,1}$ (b) $\phi_{2,1}$

(c) σ_1^2 (d) σ_2^2

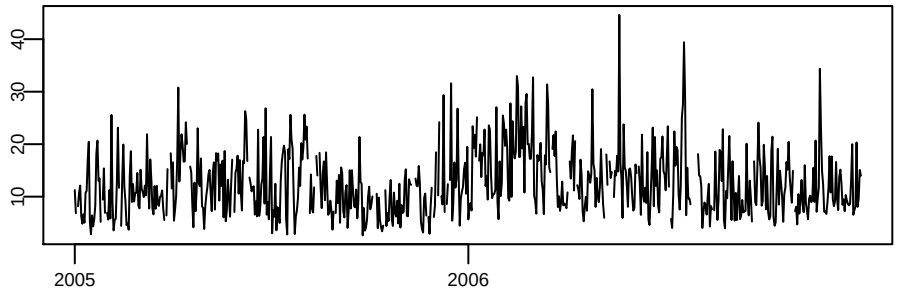
Figure 4.1 – Box-plot of $\hat{\varphi}$ (Cla) and $\hat{\varphi}_{\boldsymbol{\rho}}$ (Rob) estimates of Model 3 for $n = 300$. Cases: (i) no outlier and complete; (ii) uncontaminated and missing; and (iii) outlier and incomplete time series.

the identifiability condition **(A1)** of Sarnaglia, Reisen & Bondon (2016a), while the other restrictions are imposed as constraints in the optimization algorithm.

Due to the fact that the proposed estimation methods are implemented for zero mean processes, a prior centralization by the periodic sample means has to be performed. For the full sample, this is carried out by subtracting the usual periodic sample means. On the other hand, for the time series with missing values we estimate the average by $\hat{\mu}_{\nu,X} = \hat{\mu}_{\nu,Y}/\hat{\mu}_{\nu,U}$ which is a natural generalization of the sample mean for time series with missing data, as used in Dunsmuir & Robinson (1981b). The initial guess for the AR and MA parts of the estimators was set as zero, that is a periodic white noise. The σ_ν^2 parts were set, in the full sample case, as the usual sample variance $s_X^2 = 39.4$ and, in the incomplete sample scenario, as the variance estimator $\hat{\sigma}_X^2 = s_Y^2/\hat{\mu}_U = 39.6$, where $\hat{\mu}_U = \sum_t U_t/1603$ is the sample mean of the no missing indicator sample. See Dunsmuir &



(a) Full sample from January 1, 2005 to December 31, 2009.



(b) Stretch from January 1, 2005 to December 31, 2006 with the generated missing data.

Figure 4.2 – Evolution of SO_2 daily mean concentrations in Vitória.

Robinson (1981b), Dunsmuir & Robinson (1981a) for more details. The model estimates obtained are presented in Table 4.6.

Table 4.6 – Model estimates for complete and incomplete SO_2 daily mean concentrations.

			ν						
Param.	Scenario	Method	1	2	3	4	5	6	7
		AWLE	−0.72	−1.14	−0.80	−0.89	−0.58	−0.61	−0.70
	Complete	M-AWLE	−0.79	−1.16	−0.82	−0.88	−0.58	−0.58	−0.64
		AWLE	−0.72	−1.13	−0.79	−0.83	−0.60	−0.57	−0.74
$\phi_{\nu,1}$	Missing	M-AWLE	−0.80	−1.12	−0.81	−0.82	−0.58	−0.56	−0.64
		AWLE	−0.49	−0.74	−0.54	−0.50	—	—	−0.36
	Complete	M-AWLE	−0.57	−0.79	−0.52	−0.46	—	—	−0.31
		AWLE	−0.46	−0.71	−0.53	−0.41	—	—	−0.41
$\theta_{\nu,1}$	Missing	M-AWLE	−0.56	−0.73	−0.50	−0.36	—	—	−0.30
		AWLE	29.2	28.7	23.6	19.5	26.2	33.1	32.6
	Complete	M-AWLE	29.8	25.8	22.1	20.8	24.8	32.4	27.9
		AWLE	29.1	29.2	23.4	20.2	26.3	32.2	33.5
σ^2_ν	Missing	M-AWLE	30.5	26.4	21.8	21.2	25.8	31.8	28.0

From the results in Table 4.6, we observe that the estimates in the full sample case using the classical periodogram are almost the same as in Sarnaglia, Reisen & Bondon (2016a). This is expected since the difference in their likelihood and the one considered here is negligible for large sample sizes such as in this case. The other results in the above table reflects the conclusion obtained in the simulation study. For some periods there is no substantial difference in AR, MA and variance estimates from AWLE and M-AWLE. However, there are periods on which the AWLE and the M-AWLE provide notably different estimates, wherein the AWLE of AR and MA parts are smaller and the

white noise variance parts are greater than the M-AWLE counterparts, e.g. for $\nu = 1, 2$ and 7. This can be an indicative of the presence of outliers. The occurrence of missing data does not change significantly the estimates.

4.7 Proofs

4.7.1 Proof of Theorem 4.1

From now on, K represents a constant which may vary throughout the text. Let $\alpha = 2 + \delta$, then Markov inequality gives

$$\begin{aligned}\mathsf{P}(|\bar{Y}_{\nu,N}| > \epsilon) \leq \frac{1}{\epsilon^\alpha}\,\mathrm{E}(|\bar{Y}_{\nu,N}|^\alpha) &\overset{(1)}{\leq} \frac{1}{N^\alpha \epsilon^\alpha}\sum_{r=0}^{N-1}\mathrm{E}(|Y_{r\mathcal{S}+\nu}|^\alpha) \overset{(2)}{=} \frac{1}{N^\alpha \epsilon^\alpha}\sum_{r=0}^{N-1}\mathrm{E}(|U_{r\mathcal{S}+\nu}|^\alpha)\,\mathrm{E}(|X_{r\mathcal{S}+\nu}|^\alpha) \\ &\overset{(3)}{\leq} \frac{K}{N^{1+\delta}\epsilon^{2+\delta}},\end{aligned}$$

where (1) follows from Jensen inequality, (2) from **(A1)** and (3) from **(A2)**, **(A3)** and **(A4)**. From the integral test,

$$\sum_{N=1}^{\infty}\mathsf{P}(|\bar{Y}_{\nu,N}| > \epsilon) \leq \sum_{N=1}^{\infty}\frac{K}{N^{1+\delta}\epsilon^{2+\delta}} < \infty.$$

Therefore, the Borel-Cantelli Lemma ensures that $\bar{Y}_{\nu,N} \to 0$, $N \to \infty$, a.s..

In order to prove a.s. convergence of $c_{\nu,Y,N}(\tau)$, we first note that

$$\begin{aligned}c_{\nu,Y,N}(\tau) = &\frac{1}{N}\sum_r U_{r\mathcal{S}+\nu}X_{r\mathcal{S}+\nu}U_{r\mathcal{S}+\nu-\tau}X_{r\mathcal{S}+\nu-\tau} \\ &- \frac{1}{N}\sum_r (U_{r\mathcal{S}+\nu}U_{r\mathcal{S}+\nu-\tau} - \varsigma_{\nu,U}(\tau))\varsigma_{\nu,X}(\tau) + \frac{1}{N}\sum_r (U_{r\mathcal{S}+\nu}U_{r\mathcal{S}+\nu-\tau} - \varsigma_{\nu,U}(\tau))\varsigma_{\nu,X}(\tau),\end{aligned}$$

which implies

$$c_{\nu,Y,N}(\tau) - \varsigma_{\nu,U}(\tau)\varsigma_{\nu,X}(\tau) = \frac{1}{N}\sum_r U_{r\mathcal{S}+\nu}U_{r\mathcal{S}+\nu-\tau}(X_{r\mathcal{S}+\nu}X_{r\mathcal{S}+\nu-\tau} - \varsigma_{\nu,X}(\tau)) \tag{4.23}$$

$$+ \frac{1}{N}\sum_r (U_{r\mathcal{S}+\nu}U_{r\mathcal{S}+\nu-\tau} - \varsigma_{\nu,U}(\tau))\varsigma_{\nu,X}(\tau) + \frac{r^*}{N}\varsigma_{\nu,X}(\tau). \tag{4.24}$$

From **(A2)**, (4.24) $\to 0$, $N \to \infty$, a.s., so that we just need to consider (4.23). Now, let $\alpha = 1 + \frac{\delta}{2}$, $X^*_{r,\nu,\tau} = X_{r\mathcal{S}+\nu}X_{r\mathcal{S}+\nu-\tau} - \varsigma_{\nu,X}(\tau)$, $U^*_{r,\nu,\tau} = U_{r\mathcal{S}+\nu}U_{r\mathcal{S}+\nu-\tau}$ and $S_{\nu,N} = \sum_r U^*_{r,\nu,\tau}X^*_{r,\nu,\tau}$. Therefore,

$$\begin{aligned}\mathrm{E}\left(\left|\tfrac{1}{N}S_{\nu,N}\right|^\alpha\right) &= N^{-\alpha}\,\mathrm{E}[\mathrm{E}(|\textstyle\sum_r U^*_{r,\nu,\tau}X^*_{r,\nu,\tau}|^\alpha | U_t, t \in \mathbb{Z})] \\ &\leq N^{-\alpha}\,\mathrm{E}\{\mathrm{E}[(\textstyle\sum_r |U^*_{r,\nu,\tau}||X^*_{r,\nu,\tau}|)^\alpha | U_t, t \in \mathbb{Z}]\},\end{aligned}$$

by triangle inequality. Then

$$\begin{aligned}|S_{\nu,N}| &\leq \textstyle\sum_r [(|U^*_{r,\nu,\tau}|)^{1-1/\alpha}][|(U^*_{r,\nu,\tau})^{1/\alpha} X^*_{r,\nu,\tau}|] \\ &\overset{(1)}{\leq} \{\textstyle\sum_r [(|U^*_{r,\nu,\tau}|)^{1-1/\alpha}]^{\alpha/(\alpha-1)}\}^{1-1/\alpha} \{\textstyle\sum_r [|(U^*_{r,\nu,\tau})^{1/\alpha} X^*_{r,\nu,\tau}|]^\alpha\}^{1/\alpha} \\ &= \left(\textstyle\sum_r |U^*_{r,\nu,\tau}|\right)^{(\alpha-1)/\alpha} \left(\textstyle\sum_r |(U^*_{r,\nu,\tau}||X^*_{r,\nu,\tau}|^\alpha\right)^{1/\alpha},\end{aligned}$$

where (1) follows from Hölder's inequality. Therefore, from **(A1)**,

$$\mathrm{E}(|S_{\nu,N}|^\alpha | U_t, t \in \mathbb{Z}) \leq \left(\textstyle\sum_r |U^*_{r,\nu,\tau}|\right)^{(\alpha-1)} \textstyle\sum_r |U^*_{r,\nu,\tau}| \, \mathrm{E}\left(|X^*_{r,\nu,\tau}|^\alpha\right) \leq K \left(\textstyle\sum_r |U^*_{r,\nu,\tau}|\right)^\alpha,$$

since it can be shown that **(A3)** and **(A4)** and Minkowski and Cauchy-Schwarz inequalities ensure $\mathrm{E}(|X^*_{r,\nu,\tau}|^\alpha) \leq K < \infty$. Thus,

$$\mathrm{E}(|\tfrac{1}{N} S_{\nu,N}|^\alpha) = \mathrm{E}\left(\left|\tfrac{1}{N} \textstyle\sum_r U^*_{r,\nu,\tau} X^*_{r,\nu,\tau}\right|^\alpha\right) \leq N^{-\alpha} K \, \mathrm{E}[(\textstyle\sum_r |U^*_{r,\nu,\tau}|)^\alpha] \overset{(1)}{\leq} N^{-\alpha} K \textstyle\sum_r \mathrm{E}(|U^*_{r,\nu,\tau}|^\alpha).$$

From **(A2)** and Cauchy-Schwarz inquality, we have that $\mathrm{E}(|U^*_{r,\nu,\tau}|^\alpha) \leq K < \infty$. Hence,

$$\mathrm{E}(|\tfrac{1}{N} S_{\nu,N}|^\alpha) = \mathrm{E}(|\tfrac{1}{N} \textstyle\sum_r U_{r\mathcal{S}+\nu} U_{r\mathcal{S}+\nu-\tau} (X_{r\mathcal{S}+\nu} X_{r\mathcal{S}+\nu-\tau} - \varsigma_{\nu,X}(\tau))|^\alpha) \leq K N^{1-\alpha} = K N^{-\delta/2}.$$

We now use the method of subsequences to show that $\frac{1}{N} S_{\nu,N} \to 0$, $N \to \infty$, a.s.. More specifically, we have to find a subsequence $\frac{1}{N_k} S_{\nu,N_k}$, $k \geq 1$, such that $\frac{1}{N_k} S_{\nu,N_k} \to 0$ and $M_k \to 0$, $k \to \infty$, a.s., where

$$M_k = \max_{N_k \leq m < N_{k+1}} \left| \frac{1}{m} S_{\nu,m} - \frac{1}{N_k} S_{\nu,N_k} \right|.$$

Take an $h \in \{2, 3, \ldots\}$ and put $\alpha = 1 + \frac{\delta}{2}$ and $N_k = h^k$, $k \in \mathbb{N}$. Therefore,

$$\mathsf{P}(|\tfrac{1}{N_k} S_{\nu,N_k}| > \epsilon) \leq \tfrac{1}{\epsilon^\alpha} \mathrm{E}(|\tfrac{1}{N_k} S_{\nu,N_k}|^\alpha) \leq K N_k^{1-\alpha} = K N_k^{-\delta/2} = K r^k,$$

where $0 < r = h^{-\delta/2} < 1$. Hence $\sum_{k=1}^\infty \mathsf{P}(|\frac{1}{N_k} S_{\nu,N_k}| > \epsilon) < \infty$ and, from Borel-Cantelli Lemma, we have that $\frac{1}{N_k} S_{\nu,N_k} \to 0$, $k \to \infty$, a.s.. It remains to show that $M_k \to 0$, $k \to \infty$, a.s.. Let $D_k = \max_{N_k \leq m < N_{k+1}} |S_{\nu,m} - S_{\nu,N_k}|$ and note that, for $N_k \leq m < N_{k+1}$, $|S_{\nu,m}| \leq |S_{\nu,N_k}| + |D_k|$ and $1/m \leq 1/N_k$. Hence, $|\frac{1}{m} S_{\nu,m}| \leq |\frac{1}{N_k} S_{\nu,N_k}| + |\frac{1}{N_k} D_k|$. From triangle inequality, we have

$$M_k \leq \left| \frac{1}{N_k} S_{\nu,N_k} \right| + \max_{N_k \leq m < N_{k+1}} \left| \frac{1}{m} S_{\nu,m} \right| \leq 2 \left| \frac{1}{N_k} S_{\nu,N_k} \right| + \left| \frac{D_k}{N_k} \right|.$$

The first term in the last inequality tends to 0 a.s., such that we just have to study the convergence of the second term. Note that, for $m > N_k$, $S_{\nu,m} - S_{\nu,N_k} = \sum_{r=N_k+1}^m U^*_{r,\nu,\tau} X^*_{r,\nu,\tau}$, thus Hölder's inequality and **(A1)**, **(A3)** and **(A4)** give

$$\mathrm{E}(D_k^\alpha | U_t, t \in \mathbb{Z}) \leq K \left(\textstyle\sum_{r=N_k+1}^{N_{k+1}-1} |U^*_{r,\nu,\tau}| \right)^\alpha$$

which implies

$$\mathrm{E}(D_k^\alpha) \leq \textstyle\sum_{r=N_k+1}^{N_{k+1}-1} K \leq K(N_{k+1} - N_k) \leq K h^k,$$

from **(A2)** and Cauchy-Schwarz inquality. Therefore, $\mathrm{E}\left(\frac{D_k^\alpha}{N_k^\alpha}\right) \leq K\frac{h^k}{h^{\alpha k}} = Kh^{k(1-\alpha)} = Kr^k$, where $0 < r = h^{-\delta/2} < 1$. From Markov's inequality and Borel-Cantelli Lemma, we obtain that $D_k/N_k \to 0$ (and $M_k \to 0$), $k \to \infty$, a.s., which ensures that

$$\lim_{N\to\infty} c_{\nu,Y,N}(\tau) = \varsigma_{\nu,U}(\tau)\varsigma_{\nu,X}(\tau),$$

almost surely.

4.7.2 Proof of Theorem 4.2

Firstly, observe that

$$\begin{aligned}
\mathrm{E}([I_{\mathbf{Y}}(\omega)]_{l,m}) &= \frac{1}{N}\sum_{r,s=0}^{N-1} \mathrm{E}(U_{r\mathcal{S}+l}X_{r\mathcal{S}+l}U_{s\mathcal{S}+m}X_{s\mathcal{S}+m})e^{-\mathbf{i}\omega(r-s)} \\
&= \frac{1}{N}\sum_{r,s=0}^{N-1} \mathrm{E}(U_{r\mathcal{S}+l}U_{s\mathcal{S}+m})[\Gamma_{\mathbf{X}}(r-s)]_{l,m}e^{-\mathbf{i}\omega(r-s)} \\
&= \sum_{|\tau|<N} [\Gamma_{\mathbf{X}}(\tau)]_{l,m}\,\mathrm{E}([\hat{\mathbf{C}}_{\mathbf{U}}(\tau)]_{l,m})e^{-\mathbf{i}\omega\tau},
\end{aligned}$$

where $\hat{\mathbf{C}}_{\mathbf{U}}(\tau) = \frac{1}{N}\sum_r \mathbf{U}_r\mathbf{U}'_{r-\tau}$. From **(A2)**, we can show that $\mathrm{E}(|[\hat{\mathbf{C}}_{\mathbf{U}}(\tau)]_{l,m}|) \leq K$ independent of N. Therefore, $\mathrm{E}([I_{\mathbf{Y}}(\omega)]_{l,m})$ converges, because Assumption **(A3)** implies that $\sum_{\tau=-\infty}^{\infty}|[\Gamma_{\mathbf{X}}(\tau)]_{l,m}| < \infty$. Let $b_\tau = [\Gamma_{\mathbf{X}}(\tau)]_{l,m}e^{-\mathbf{i}\omega\tau}$ and $a_{\tau,N} = \mathrm{E}([\hat{\mathbf{C}}_{\mathbf{U}}(\tau)]_{l,m})$, $|\tau| \leq N-1$, $a_{\tau,N} = 0$, $|\tau| > N-1$. Then

$$\lim_{N\to\infty}\mathrm{E}([I_{\mathbf{Y}}(\omega)]_{l,m}) = \lim_{N\to\infty}\sum_{\tau=-\infty}^{\infty} b_\tau a_{\tau,N} \overset{(1)}{=} \sum_{\tau=-\infty}^{\infty} b_\tau \lim_{N\to\infty} a_{\tau,N},$$

where (1) follows from the dominated convergence theorem. In addition, under **(A2')**, $|U_t| \leq U$, $t \in \mathbb{Z}$, and $\mathrm{E}(U) < \infty$, so that the dominated convergence theorem can be used again to show that

$$\lim_{N\to\infty} a_{\tau,N} = \lim_{N\to\infty}\mathrm{E}([\hat{\mathbf{C}}_{\mathbf{U}}(\tau)]_{l,m}) = \mathrm{E}\left(\lim_{N\to\infty}[\hat{\mathbf{C}}_{\mathbf{U}}(\tau)]_{l,m}\right) = [\mathbf{C}_{\mathbf{U}}(\tau)]_{l,m}.$$

Hence

$$\lim_{N\to\infty}\mathrm{E}([I_{\mathbf{Y}}(\omega)]_{l,m}) = \sum_{\tau=-\infty}^{\infty}[\Gamma_{\mathbf{X}}(\tau)]_{l,m}[\mathbf{C}_{\mathbf{U}}(\tau)]_{l,m}e^{-\mathbf{i}\omega\tau} = \mathbf{f}_{\mathbf{Y}}(\omega).$$

5 Conclusão

Esta tese investiga a estimação de modelos PARMA para ajustar séries temporais em diversas situações encontradas na prática. As conclusões aqui expostas se baseiam em resultados teóricos e empíricos obtidos nos capítulos anteriores.

O problema de identificabilidade desses modelos foi tratado diretamente e restrições simples que asseguram essa propriedade foram encontradas. Em adição, a consistência e a normalidade assintóticas do estimador de Whittle foram obtidas sob suposições usuais na literatura. O estudo de simulação mostrou as vantagens da metodologia proposta, principalmente no que tange ao tempo de convergência, comparado com o estimador de máxima verossimilhança gaussiano.

O ajuste de modelos PARMA a séries temporais, completas ou com dados faltantes, acometidas por outliers aditivos também foi abordado. Um estimador robusto da matriz de densidade espectral foi proposto. Essa alternativa ao periodograma foi utilizada na verossimilhança de Whittle com o objetivo de fornecer estimativas robustas para os parâmetros do modelo PARMA. O estudo de simulação demonstrou a robustez do estimador proposto quando a série temporal é acometida por outliers aditivos.

Um estimador da função de autocovariância periódica apropriado a séries temporais incompletas foi proposto e sua consistência forte foi determinada sob suposições similares ao caso estacionário investigado por Dunsmuir & Robinson (1981a). Dois estimadores para séries temporais PARMA com dados faltantes foram introduzidos. Ambos se destacam por não sofrer perda significante de eficiência comparados aos resultados para amostras completas. Entretanto, sob a influência simultânea de observações atípicas e faltantes, o estimador baseado no periodograma clássico é completamente afetado, enquanto que a alternativa robusta tem desempenho quase inalterado.

Este estudo cria diversas linhas de pesquisa promissoras e que podem ser perseguidas futuramente, tais como: investigação de propriedades assintóticas dos estimadores propostos; ajuste do modelo PARMA à séries temporais incompletas, com volatilidade e sob efeito de observações aberrantes; desenvolvimento de estimadores robustos da função de autocovariância e da matriz de densidade espectral de séries temporais com dados faltantes, entre outras.

Referências Bibliográficas

ANDERSON, P. L.; MEERSCHAERT, M. M. Parameter estimation for periodically stationary time series. *Journal of Time Series Analysis*, v. 26, n. 14, p. 489–518, 2005. Citado na página 12.

ANDERSON, P. L.; MEERSCHAERT, M. M.; VECCHIA, A. V. Innovations algorithm for periodically stationary time series. *Stochastic Processes and their Applications*, Elsevier, v. 83, n. 1, p. 149–169, 1999. Citado na página 12.

BASAWA, I. V.; LUND, R. B. Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis*, v. 22, p. 651–663, 2001. Citado 6 vezes nas páginas 12, 16, 17, 24, 39 e 52.

BLOOMFIELD, P.; HURD, H. L.; LUND, R. B. Periodic correlation in stratospheric ozone data. *Journal of Time Series Analysis*, v. 15, p. 127–150, 1994. Citado 4 vezes nas páginas 11, 15, 38 e 51.

BONDON, P.; BAHAMONDE, N. Least squares estimation of ARCH models with missing observations. *Journal of Time Series Analysis*, v. 33, p. 880–891, may 2012. Citado 2 vezes nas páginas 14 e 52.

BOX, G.; JENKINS, G.; REINSEL, G. *Time Series Analysis: Forecasting and Control*. 4th. ed. [S.l.]: Prentice Hall, 2008. Citado na página 53.

BROCKWELL, P. J.; DAVIS, R. A. *Time Series: Theory and Methods*. 2nd. ed. New York, NY: Springer Science, 2006. (Springer Series in Statistics). Citado 4 vezes nas páginas 11, 12, 20 e 41.

DEISTLER, M.; DUNSMUIR, W.; HANNAN, E. J. Vector linear time series models: Corrections and extensions. *Advances in Applied Probability*, v. 10, p. 360–372, 1978. Citado 8 vezes nas páginas 12, 16, 20, 21, 36, 37, 41 e 44.

DENBY, L.; MARTIN, R. D. Robust estimation of the first order autoregressive parameter. *Journal of the American Statistical Association*, v. 74, p. 140–46, 1979. Citado 2 vezes nas páginas 12 e 39.

DRAKE, C.; KNAPIK, O.; LEŚKOW, J. EM-based inference for cyclostationary time series with missing observations. In: *Cyclostationarity: Theory and Methods*. [S.l.]: Springer, 2014. p. 23–35. Citado 2 vezes nas páginas 13 e 14.

DRAKE, C.; KNAPIK, O.; LEŚKOW, J. Missing data analysis in cyclostationary models. *Czasopismo Techniczne*, p. 25–36, 2015. Citado 2 vezes nas páginas 13 e 14.

DRAKE, C.; LEŚKOW, J.; GARAY, A. M. Imputation of missing observations for heavy tailed cyclostationary time series. In: *Cyclostationarity: Theory and Methods-II*. [S.l.]: Springer, 2015. p. 179–187. Citado na página 14.

DUNSMUIR, W. A central limit theorem for parameter estimation in stationary vector ttime series and its application to models for a signal observed with noise. *The Annals of Statistics*, p. 490–506, 1979. Citado na página 37.

DUNSMUIR, W.; HANNAN, E. J. Vector Linear Time Series Models. *Advances in Applied Probability*, v. 8, n. 2, p. 339, jun 1976. Disponível em: ⟨http://dx.doi.org/10.2307/1425908⟩. Citado 11 vezes nas páginas 12, 16, 20, 22, 26, 36, 37, 39, 41, 44 e 54.

DUNSMUIR, W.; ROBINSON, P. M. Asymptotic Theory for Time Series Containing Missing and Amplitude Modulated Observations. *Sankhyã: The Indian Journal of Statistics, Series A*, v. 43, p. 260–281, 1981. Citado 7 vezes nas páginas 14, 55, 56, 57, 59, 69 e 73.

DUNSMUIR, W.; ROBINSON, P. M. Estimation of Time Series Models in the Presence of Missing Data. *Journal of the American Statistical Association*, v. 76, n. 375, p. 560–568, sep 1981. Disponível em: ⟨http://dx.doi.org/10.1080/01621459.1981.10477687⟩. Citado 4 vezes nas páginas 14, 52, 68 e 69.

DUNSMUIR, W.; ROBINSON, P. M. Parametric Estimators for Stationary Time Series with Missing Observations. *Advances of Applied Probability*, v. 13, p. 129–146, 1981. Disponível em: ⟨http://dx.doi.org/10.1080/01621459.1981.10477687⟩. Citado 5 vezes nas páginas 14, 52, 56, 58 e 60.

EFROMOVICH, S. Efficient Non-Parametric Estimation of the Spectral Density in the Presence of Missing Observations. *Journal of Time Series Analysis*, v. 35, p. 407–427, Apr 2014. Citado 2 vezes nas páginas 14 e 52.

FAJARDO, F.; REISEN, V. A.; CRIBARI-NETO, F. Robust estimation in long-memory processes under additive outliers. *Journal of Statistical Planning and Inference*, v. 139, p. 2511–2525, 2009. Citado 4 vezes nas páginas 12, 13, 42 e 55.

FAJARDO, F.; REISEN, V. A.; LéVY-LEDUC, C.; TAQQU, M. S. Robust Periodogram Methods for Time Series with Long-Range Dependece: An Application to Pollution Levels. To appear. 2015. Citado 4 vezes nas páginas 13, 39, 43 e 52.

FOX, R.; TAQQU, M. S. Large-sample properties of parameters estimates for strongly dependent stationary gaussian time series. *The Annals of Statistics*, v. 14, n. 2, p. 517–532, 1986. Citado na página 16.

FRANSES, P. H.; OOMS, M. A periodic long-memory model for quarterly UK inflation. *International Journal of Forecasting*, v. 13, p. 117–126, 1997. Citado na página 12.

GARDNER, W. A.; FRANKS, L. E. Characterization of cyclostationary random signal processes. *IEEE Transactions on Signal Processing*, v. 21, p. 4–14, 1975. Citado 4 vezes nas páginas 11, 15, 38 e 51.

GARDNER, W. A.; NAPOLITANO, A.; PAURA, L. Cyclostationarity: Half a century of research. *Signal Processing*, v. 86, p. 639–697, 2006. Citado 3 vezes nas páginas 15, 38 e 51.

GLADYSHEV, E. G. Periodically correlated random sequences. *Sov. Math.*, v. 2, p. 385–388, 1961. Citado 7 vezes nas páginas 11, 15, 20, 38, 51, 53 e 54.

GOLUB, G. H.; LOAN, C. F. V. *Matrix computations*. 4th. ed. [S.l.]: Johns Hopkins University Press, Baltimore, MD, 2012. v. 3. Citado na página 33.

HAUSER, M. A. Maximum likelihood estimators for ARMA and ARFIMA models: a Monte Carlo study. *Journal of Statistical Planning and Inference*, v. 80, p. 229–255, 1999. Citado na página 29.

HEYMANN, M. *Structure and realization problems in the theory of dynamical systems*. [S.l.]: Springer-Verlag, Vienna-New York, 1975. Citado na página 20.

HUBER, P. J. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, v. 35, n. 1, p. 73–101, 1964. Citado 3 vezes nas páginas 43, 45 e 63.

HURD, H. L.; GERR, N. L. Graphical methods for determining the presence of periodic correlation. *Journal of the American Statistical Association*, v. 12, p. 337–50, 1991. Citado na página 11.

HURD, H. L.; MIAMEE, A. *Periodically correlated random sequences: Spectral theory and practice*. Hoboken, NJ: John Wiley & Sons, 2007. xviii+353 p. (Wiley Series in Probability and Statistics). ISBN 978-0-471-34771-2. Disponível em: ⟨http://dx.doi.org/10.1002/9780470182833⟩. Citado 6 vezes nas páginas 15, 20, 38, 51, 53 e 57.

KATKOVNIK, V. Robust M-Periodogram. *IEEE Transactions on Signal Processing*, v. 46, n. 11, p. 3104–3109, 1998. Citado 2 vezes nas páginas 42 e 55.

KOOPMAN, S. J.; OOMS, M.; CARNERO, M. A. Periodic seasonal Reg-ARFIMA-GARCH models for daily electricity spot prices. *Journal of the American Statistical Association*, v. 102, p. 16–27, 2007. Citado na página 12.

KOUL, H. L. M-estimators in linear models with long range dependent errors. *Statistics and Probability Letters*, v. 14, n. 2, p. 153–164, 1992. Citado 2 vezes nas páginas 13 e 39.

LI, T.-H. Laplace periodogram for time series analysis. *Journal of the American Statistical Association*, v. 103, n. 482, p. 757–768, 2008. Citado 2 vezes nas páginas 13 e 39.

LI, T.-H. A nonlinear method for robust spectral analysis. *IEEE Transactions on Signal Processing*, v. 58, n. 5, p. 2466–2474, 2010. Citado 2 vezes nas páginas 13 e 39.

LI, W. K.; HUI, Y. V. An algorithm for the exact likelihood of periodic autoregressive moving average models. In: *Communications in Statistics*. [S.l.]: Taylor and Francis, 1988, (Simulation and Computation, v. 17). p. 1483–94. Citado 2 vezes nas páginas 12 e 16.

LUND, R. B.; BASAWA, I. V. Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis*, v. 21, p. 75–93, 2000. Citado 6 vezes nas páginas 12, 16, 21, 26, 39 e 51.

LUND, R. B.; HURD, H. L.; BLOOMFIELD, P.; SMITH, R. L. Climatological time series with periodic correlation. *Journal of Climate*, v. 8, p. 2787–2809, 1995. Citado na página 11.

MA, Y.; GENTON, M. Highly robust estimation of the autocovariance function. *Journal of Time Series Analysis*, v. 21, p. 663–684, 2000. Citado 3 vezes nas páginas 12, 13 e 39.

MCLEOD, A. I. Diagnostic checking periodic autoregression models with application. *Journal of Time Series Analysis*, v. 15, p. 221–233, 1994. Citado 2 vezes nas páginas 11 e 31.

METAXOGLOU, K.; SMITH, A. Maximum Likelihood Estimation of VARMA Models Using a State-Space EM Algorithm. *Journal of Time Series Analysis*, v. 28, n. 5, p. 666–685, 2007. Citado na página 13.

PARZEN, E. *Spectral analysis of asymptotically stationary time series*. [S.l.], 1961. Citado na página 57.

PARZEN, E. On spectral analysis with missing observations and amplitude modulation. *Sankhyā: The Indian Journal of Statistics, Series A*, JSTOR, p. 383–392, 1963. Citado na página 56.

PRIESTLEY, M. B. *Spectral Analysis and Time Series*. [S.l.]: Academic press, 1981. Citado 3 vezes nas páginas 11, 13 e 39.

REINSEL, G. C. *Elements of Multivariate Time Series Analysis*. 2nd. ed. [S.l.]: Springer Series in Statistics, 1997. Citado 3 vezes nas páginas 12, 20 e 36.

SARNAGLIA, A. J. Q.; REISEN, V. A.; BONDON, P. Periodic ARMA models: Application to particulate matter concentrations. In: IEEE. *Signal Processing Conference (EUSIPCO), 2015 23rd European*. [S.l.], 2015. p. 2181–2185. Citado 6 vezes nas páginas 12, 39, 44, 45, 52 e 58.

SARNAGLIA, A. J. Q.; REISEN, V. A.; BONDON, P. Identifiability and Whittle Estimator of periodic ARMA models. To appear. 2016. Citado 10 vezes nas páginas 39, 48, 49, 52, 54, 58, 59, 67, 68 e 69.

SARNAGLIA, A. J. Q.; REISEN, V. A.; BONDON, P. M-regression spectral estimator for periodic ARMA models. A robust method against additive outliers and heavy tail distributions. Submitted to Computational Statistics & Data Analysis. 2016. Citado 7 vezes nas páginas 51, 52, 55, 58, 61, 62 e 63.

SARNAGLIA, A. J. Q.; REISEN, V. A.; LéVY-LEDUC, C. Robust estimation of periodic autoregressive processes in the presence of additive outliers. *Journal of Multivariate Analysis*, v. 101, p. 2168–2183, 2010. Disponível em: ⟨http://dx.doi.org/10.1016/j.jmva.2010.05.006⟩. Citado 6 vezes nas páginas 13, 16, 39, 41, 55 e 60.

SHAO, Q. Robust estimation for periodic autoregressive time series. *Journal of Time Series Analysis*, v. 29, n. 2, p. 251–263, 2008. Citado 2 vezes nas páginas 13 e 39.

THOMAS, H. A.; FIERING, M. B. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. *Design of water resource systems*, Harvard University Press Cambridge, Mass, p. 459–493, 1962. Citado na página 11.

TIAO, G. C.; GRUPE, M. R. Hidden periodic autoregressive-moving average models in time series data. *Biometrika*, v. 67, p. 365–73, 1980. Citado 4 vezes nas páginas 11, 15, 38 e 51.

TROUTMAN, B. M. Some results in periodic autoregression. *Biometrika*, v. 66, n. 2, p. 219–228, 1979. Citado na página 11.

VECCHIA, A. V. Maximum likelihood estimation for periodic autoregressive moving average models. *Technometrics*, v. 27, p. 375–84, 1985. Citado na página 12.

VECCHIA, A. V. Periodic autoregressive-moving average (PARMA) modeling with applications to water resources. *Journal of the American Water Resources Association*, v. 21, n. 5, p. 721–730, 1985. Citado na página 40.

VECCHIA, A. V.; BALLERINI, R. Testing for periodic autocorrelations in seasonal time series data. *Biometrika*, v. 78, p. 53–63, 1991. Citado na página 11.

WHITTLE, P. The analysis of multiple stationary time series. *Journal of the Royal Statistical Society. Series B*, v. 15, p. 125–139, 1953. Citado na página 16.

YAJIMA, Y.; NISHINO, H. Estimation of the autocorrelation function of a stationary time series with missing observations. *The Indian Journal of Statistics*, v. 61, p. 189–207, 1999. Citado 2 vezes nas páginas 14 e 52.

ZYGMUND, A. *Trigonometric Series*. 3rd. ed. [S.l.]: Cambridge University Press, 2002. Citado na página 37.