PERIODIC ARMA MODEL: FORECASTING, PARSIMONY, ASYMPTOTIC NORMALITY AND AIC

By

Kai Zhang

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Statistics - Doctor of Philosophy

2014

ABSTRACT

PERIODIC ARMA MODEL: FORECASTING, PARSIMONY, ASYMPTOTIC NORMALITY AND AIC

By

Kai Zhang

Periodic autoregressive moving average (PARMA) models are indicated for time series whose mean, variance, and covariance function vary with the season. In this thesis, I develop and implement forecasting procedures for PARMA models. The required computations are documented in detail. An application to monthly river flow forecasting is provided.

For my father.

ACKNOWLEDGMENTS

I was lucky to meet with my thesis advisor Dr. Mark Meerschaert, who led me into the fascinating world of time series and academic research. I want to say a million thanks, however it will not be enough for my gratitude. I could not make it without his encouragement and belief in me. I will also benefit from his life attitude for the rest of my life.

I am also grateful for the tremendous help from Dr. Paul Anderson, who introduced me into the periodic ARMA modeling. Thanks him for numerous of transits to East Lansing for discussing my thesis.

Special thanks go to my thesis committee member and department Chairperson Dr. Hira Koul. I am indebted for his help in every little detail of my doctoral years, including the learning in STT 954, failing and passing two prelim exams, student travel fund applications, summer research fellowships, job-hunting and thesis updates, etc.

Additionally, Dr. Chae Young Lim was always supportive in my research and job-hunting process. I am deeply grateful of her time spent reading my thesis and valuable feedbacks.

Lastly, I want to express my gratitude for Ms. Sue Watson.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Mathematical modeling and simulation of river flow time series are critical issues in hydrology and water resources. Most river flow time series are periodically stationary, that is, their mean and covariance functions are periodic with respect to time. To account for the periodic correlation structure, a periodic autoregressive moving average (PARMA) model can be useful. PARMA models are also appropriate for a wide variety of time series applications in geophysics and climatology. In a PARMA model, the parameters in a classical ARMA model are allowed to vary with the season. Since PARMA models explicitly describe seasonal fluctuations in mean, standard deviation and autocorrelation, they have been used to generate more faithful models and simulations of natural river flows.

Historically, Gladyshev [18] first defined the concept of periodically correlated stochastic process; In [20], some early work of Jones and Brelsford studied the problem of predicting time series with periodic structure, including the estimation of necessary parameters; In [32] and [47], Pagano and Troutman studied the elementary properties of univariate processes,

and connections with stationary multivariate processes; For stationary time series models, Box and Jenkins [10] presented a systematic approach for modeling time series based on three stages: (1) Model identification; (2) parameter estimation; and (3) diagnostic checks or tests of goodness of fit. Since then, applications and extensions of this modeling approach to hydrology have been widespread.

Adams and Goodwin [1] described an on-line parameter estimation technique, based on methods from automatic control, which provides consistent estimates of PARMA model parameters. Anderson and Vecchia [3] obtained the asymptotic distribution for the sample autocovariance and sample autocorrelation functions of PARMA process, and they also studied the asymptotic properties of the discrete Fourier transform of the estimated periodic autocovariance and autocorrelation functions. Anderson and Meerschaert [5] established the basic asymptotic theory for periodic moving averages of i.i.d. random variables with regularly varying tails. They showed that when the underlying random variables have finite variance but infinite fourth moment, the sample autocorrelations are asymptotically stable. Lund and Basawa [22, 23] explored recursive prediction and likelihood evaluation techniques for PARMA models.

Time series analysis involves four general steps: model identification, parameter estimation, diagnostic checking, and forecasting. Model identification is the most difficult step for PARMA modeling. Noakes et al. [31] suggested examining the plots of the periodic partial autocorrelation function as the best approach to identify PARMA models. This method is highly recommended when the parameter space is not constrained, however it requires a high level of user experience. Another method is to use an automatic selection criterion, such as the Akaike Information Criterion (AIC) [2], or the Bayesian information criterion

(BIC) [37] when all possible candidates of models are examined. However, this procedure requires investigating quite a large number of models, especially when the number of parameter estimates is fairly large, for example, monthly data with 12 seasons. Ursu and Turkman [52] applied the genetic algorithm as a method of identifying the PAR model, which greatly improved the model selection efficiency. Salehi H. has also made tremendous contributions on periodic stochastic process, see [26, 27, 28, 29].

Additionally, model identification for general PARMA times series was discussed in Tesfaye, Meerschaert and Anderson [42]. Anderson, Meerschaert and Vecchia [6] developed an innovations algorithm for PARMA parameter estimation. Tesfaye, Meerschaert and Anderson [42] demonstrated model diagnostics for a PARMA model of monthly river flows for the Fraser River in British Columbia, Canada. In this thesis, I develop a practical method for forecasting PARMA models, and I demonstrate the method by forecasting monthly river flows for the same time series of monthly flows.

A stochastic process $\{\tilde{X}_t\}_{t\in\mathbb{Z}}$ is periodically stationary if its mean $\mathrm{E}\tilde{X}_t$ and covariance $\mathrm{Cov}(\tilde{X}_t, \tilde{X}_{t+h})$ for $h \in \mathbb{Z}$ are periodic functions of time t with the same period S, i.e., for some integer $S \geq 1$, for $i = 0, 1, ..., S-1$, and for all integers k and h, I have

$$\mathrm{E}\tilde{X}_i = \mathrm{E}\tilde{X}_{i+kS} \quad \text{and} \quad \mathrm{Cov}\left(\tilde{X}_i, \tilde{X}_{i+h}\right) = \mathrm{Cov}\left(\tilde{X}_{i+kS}, \tilde{X}_{i+kS+h}\right).$$

A periodically stationary process $\{\tilde{X}_t\}$ ia called a $\mathrm{PARMA}_S(p,q)$ process if the mean-centered process $X_t = \tilde{X}_t - \mu_t$ is of the form

$$X_t - \sum_{k=1}^{p} \phi_t(k) X_{t-k} = \varepsilon_t + \sum_{j=1}^{q} \theta_t(j)\varepsilon_{t-j} \tag{1.0.1}$$

where $\{\varepsilon_t\}$ is a sequence of random variables with mean zero and standard deviation $\sigma_t > 0$ such that $\{\delta_t = \sigma_t^{-1}\varepsilon_t\}$ is independent and identically distributed. ($\{\varepsilon_t\}$ is called periodic i.i.d. Gaussian noise if X_t is a Gaussian process.) Here the autoregressive parameters $\phi_t(j)$, the moving average parameters $\theta_t(j)$, and the residual standard deviations σ_t are all assumed to be periodic functions of t with the same period $S \geq 1$. Throughout this paper I will also assume:

(i) The model (1.0.1) admits a causal representation

$$X_t = \sum_{j=0}^{\infty} \psi_t(j)\varepsilon_{t-j} \tag{1.0.2}$$

where $\psi_t(0) = 1$ and $\sum_{j=0}^{\infty} |\psi_t(j)| < \infty$ for all t. Note that $\psi_t(j) = \psi_{t+kS}(j)$ for all j.

(ii) The model (1.0.1) also satisfies an invertibility condition

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_t(j)X_{t-j} \tag{1.0.3}$$

where $\pi_t(0) = 1$ and $\sum_{j=0}^{\infty} |\pi_t(j)| < \infty$ for all t, and define $X_{t-j} = 0$ when $t - j < 0$. Again, $\pi_t(j) = \pi_{t+kS}(j)$ for all j.

The notation used in this paper is consistent with: Anderson and Vecchia [3]; Anderson and Meerschaert [4, 5]; Anderson, Meerschaert, and Vecchia [6]; Anderson and Meerschaert [7]; Tesfaye, Meerschaert, and Anderson [42]; and Tesfaye, Anderson, and Meerschaert [43]. This notation is also an extension of the notation in Brockwell and Davis [11].

Suppose I have N years of data, consisting of $n = N \times S$ data points, $\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_{n-1}$,

where S is the number of seasons. For example, for monthly data I have $S = 12$, and our convention is to let $i = 0$ represent the first month, $i = 1$ represent the second, ..., and $i = S - 1 = 11$ represent the last.

The sample mean for season i is

$$\hat{\mu}_i = N^{-1} \sum_{k=0}^{N-1} \tilde{X}_{kS+i}. \tag{1.0.4}$$

The sample autocovariance for season i at lag ℓ is

$$\hat{\gamma}_i(\ell) = N^{-1} \sum_{j=0}^{N-1-h_i} \left(\tilde{X}_{jS+i} - \hat{\mu}_i\right) \left(\tilde{X}_{jS+i+\ell} - \hat{\mu}_{i+\ell}\right), \tag{1.0.5}$$

where $\ell \geq 0$, $h_i = \lfloor (i+\ell)/S \rfloor$ and $= \lfloor \cdot \rfloor$ is the greatest integer function.

The sample autocorrelation for season i at lag ℓ is

$$\hat{\rho}_i(\ell) = \frac{\hat{\gamma}_i(\ell)}{\sqrt{\hat{\gamma}_i(0)\hat{\gamma}_{i+\ell}(0)}}, \tag{1.0.6}$$

which is also the sample cross-correlation between two different seasons.

In (1.0.5) the divisor N is used rather than $N - h_i$, since this ensures that the autocovariance matrix at season i, $\hat{\Gamma}_N^{(i)} = \left[\hat{\gamma}_i(j-\ell)\right]_{j,\ell=1}^n$ is non-negative definite, where

$$\hat{\Gamma}_N^{(i)} = \left[\hat{\gamma}_i(j-\ell)\right]_{j,\ell=1}^n = \begin{pmatrix} \hat{\gamma}_i(0) & \hat{\gamma}_i(1) & \dots & \hat{\gamma}_i(N-1) \\ \hat{\gamma}_i(1) & \hat{\gamma}_i(0) & \dots & \hat{\gamma}_i(N-2) \\ \vdots & & \ddots & \vdots \\ \hat{\gamma}_i(N-1) & \hat{\gamma}_i(N-2) & \dots & \hat{\gamma}_i(0) \end{pmatrix}.$$

To see this we may write $\hat{\Gamma}_N^{(i)} = \frac{1}{N}\Gamma\Gamma'$, where Γ is the $N \times 2N$ matrix

$$\Gamma = \begin{pmatrix} 0 & \dots & \dots & 0 & Y_1 & Y_2 & \dots & \dots & Y_N \\ 0 & \dots & 0 & Y_1 & Y_2 & \dots & \dots & Y_N & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & Y_1 & Y_2 & \dots & Y_N & 0 & \dots & \dots & 0 \end{pmatrix},$$

$j = 1, \dots, N$, and $Y_j = \tilde{X}_{jS+i} - \hat{\mu}_i$. Then $\forall N \times 1$ vector $\mathbf{a}$, we have $\mathbf{a}'\hat{\Gamma}_N^{(i)}\mathbf{a} = N^{-1}(\mathbf{a}'\Gamma)(\mathbf{a}'\Gamma)' \geq 0$. By definition of non-negative definiteness, since $\mathbf{a}$ is arbitrary, then the matrix $\hat{\Gamma}_N^{(i)} = \left[\hat{\gamma}_i(j-\ell)\right]_{j,\ell=1}^n$ is non-negative definite.

Then the sample covariance matrix

$$\hat{\Gamma}_k^{(i)} = \begin{pmatrix} \hat{\gamma}_i(0) & \hat{\gamma}_i(1) & \dots & \hat{\gamma}_i(k-1) \\ \hat{\gamma}_i(1) & \hat{\gamma}_i(0) & \dots & \hat{\gamma}_i(k-2) \\ \vdots & & \ddots & \vdots \\ \hat{\gamma}_i(k-1) & \hat{\gamma}_i(k-2) & \dots & \hat{\gamma}_i(0) \end{pmatrix}$$

converges in the operator norm $\|A\| = \sup\{\|Ax\| : \|x\| = 1\}$ to the covariance matrix

$$\Gamma_k^{(i)} = \begin{pmatrix} \gamma_i(0) & \gamma_i(1) & \dots & \gamma_i(k-1) \\ \gamma_i(1) & \gamma_i(0) & \dots & \gamma_i(k-2) \\ \vdots & & \ddots & \vdots \\ \gamma_i(k-1) & \gamma_i(k-2) & \dots & \gamma_i(0) \end{pmatrix}$$

in probability as $N \to \infty$, if $k \to \infty$ in such a way that $k^2/N \to 0$. This result also assumes

a spectral bound, see [6, Theorem 3.1], and that the underlying noise sequence has finite fourth moment. Note that, if every $\sigma_i > 0$, then Proposition 4.1 of Lund and Basawa [22] shows that the covariance matrix $\Gamma_n^{(i)} = \left[\gamma_i(j-\ell)\right]_{j,\ell=1}^{n}$ is invertible for every $n \geq 1$ and each $i = 0, 1, \ldots, S-1$. Since the set of invertible matrices is open, the convergence in probability from [6, Theorem 3.1] implies that the sample covariance matrix is invertible with probability approaching 1 as $k \to \infty$.

Given a $\text{PARMA}_S(p,q)$ model (1.0.1) for a periodic time series, a recursive forecasting algorithm is developed in this thesis based on minimizing mean squared error. I detail the computation of h-step ahead forecasts for a PARMA model, based on the innovations algorithm, and an idea of Ansley (see Ansley [9]; Lund and Basawa [23]). I also have developed R codes to implement these forecasts, and compute the asymptotic variance of the forecast errors. All R codes are listed in the appendix. This thesis is laid out as follows. Chapter 2 develops the algorithms for computing h-step ahead forecasts for any $h \geq 1$, and computes the associated forecast error variances. Chapter 3 specifies the details of computation in this paper, mainly by R software. Chapter 4 illustrates the methods of this thesis by forecasting average monthly flows for the Fraser River. Chapter 5 develops a reduced $\text{PARMA}_S(p,q)$ model to achieve parsimony. Chapter 6 derives the asymptotic theory of PARMA models, and Chapter 7 discusses the periodic AIC for automatic model selection.

1.1 Computation of PARMA Autocovariances

Given a PARMA$_S(p,q)$ time series (1.0.1), Define the covariance function

$$\gamma_j(\ell - j) = \mathrm{E}(X_j X_\ell) = \mathrm{Cov}(X_j, X_\ell), \tag{1.1.1}$$

and note that

$$\gamma_j(\ell - j) = \mathrm{E}(X_j X_\ell) = \mathrm{E}(X_\ell X_j) = \gamma_\ell(j - \ell)$$

for all $j, \ell \in \mathbb{Z}$. Then the covariance function $\gamma_j(\ell - j) = \mathrm{E}(X_j X_\ell)$ can be explicitly computed by two methods. The first approach is to use the causal representation in (1.0.2). Given the model parameters, we may write

$$\begin{aligned}
\mathrm{E}\left(X_j X_\ell\right) &= \mathrm{E}\left(\left(\sum_{k=0}^{\infty} \psi_j(k)\varepsilon_{j-k}\right)\left(\sum_{r=0}^{\infty} \psi_\ell(r)\varepsilon_{\ell-r}\right)\right) \\
&= \sum_{k=0}^{\infty}\sum_{r=0}^{\infty} \psi_j(k)\psi_\ell(r)\mathrm{E}\left(\varepsilon_{j-k}\varepsilon_{\ell-r}\right).
\end{aligned}$$

Notice that $\mathrm{E}\left(\varepsilon_{j-k}\varepsilon_{\ell-r}\right) = \sigma^2_{j-k}$, when $j-k = \ell-r$, and $\mathrm{E}\left(\varepsilon_{j-k}\varepsilon_{\ell-r}\right) = 0$, otherwise. Letting $r = \ell - j + k$, therefore

$$\gamma_j(\ell - j) = \mathrm{E}\left(X_j X_\ell\right) = \sum_{k=0}^{\infty} \psi_j(k)\psi_\ell(\ell - j + k)\sigma^2_{j-k}. \tag{1.1.2}$$

However, (1.1.2) is computationally impractical since it requires determination and infinite summation of $\psi_j(k)$ and $\psi_\ell(\ell - j + k)$.

Let $\gamma_t(h) = \mathrm{Cov}\left(X_t, X_{t+h}\right)$ be the autocovariance of X_t at season $i = t - S\lfloor t/S \rfloor$ and lag $h \geq 0$. Now I will consider the second method, by mimicking Yule-Walker methods for

stationary ARMA series. Multiplying both sides of (1.0.1) by X_{t-h}, where $h > \max(p, q)$, I have

$$X_{t-h}\left(X_t - \sum_{k=1}^{p} \phi_t(k) X_{t-k}\right) = X_{t-h}\left(\varepsilon_t + \sum_{j=1}^{q} \theta_t(j)\varepsilon_{t-j}\right).$$

Take expectations on both sides, and use causal representation to compute the right hand side:

$$\begin{aligned}
\mathrm{E}\left[X_{t-h}\left(X_t - \sum_{k=1}^{p} \phi_t(k) X_{t-k}\right)\right] &= \mathrm{E}\left[X_{t-h}\left(\varepsilon_t + \sum_{j=1}^{q} \theta_t(j)\varepsilon_{t-j}\right)\right] \\
\gamma_{t-h}(h) - \sum_{k=1}^{p} \phi_t(k)\gamma_{t-h}(h-k) &= \mathrm{E}\left(\sum_{k=0}^{\infty} \psi_{t-h}(k)\varepsilon_{t-h-k}\right)\left(\varepsilon_t + \sum_{j=1}^{q} \theta_t(j)\varepsilon_{t-j}\right) \\
&= 0,
\end{aligned}$$

since $\varepsilon_{t-h-k} \perp \varepsilon_{t-j}, \forall j = 0, 1, \ldots, q, k = 0, 1, \ldots$, where $x \perp y$ if and only if $\langle x, y\rangle = 0$, and $\langle x, y\rangle$ is the inner product of x and y in Hilbert space $\mathcal{H}$. Recall from Chapter 2 of Brockwell and Davis [11] that $\langle x, y\rangle$ is called the inner product of x and y, such that

(a) $\langle x, y\rangle = \overline{\langle y, x\rangle}$, the bar denoting complex conjugation,

(b) $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$ for all $x, y, z \in \mathcal{H}$,

(c) $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$ for all $x, y \in \mathcal{H}$ and $\alpha \in R$,

(d) $\langle x, x\rangle \geq 0$ for all $x \in \mathcal{H}$,

(e) $\langle x, x\rangle = 0$ if and only if $x = 0$.

In this way I obtain

$$\gamma_{t-h}(h) = \sum_{k=1}^{p} \phi_t(k)\gamma_{t-h}(h-k). \tag{1.1.3}$$

Equation (1.1.3) expresses $\gamma_{t-h}(h)$ in terms of autocovariances of the previous p lags when $h > \max(p,q)$. The advantage of this method is that the computational complexity of (1.1.3) does not increase with increasing h. Therefore, once $\gamma_{t-h}(h)$ is identified for all lags $0 \leq h \leq \max(p,q)$ and for all seasons $i = 0, 1, \ldots, S-1$, then the PARMA autocovariances at higher lags can be efficiently computed. Next I will focus on computation of $\gamma_{t-h}(h)$ for $0 \leq h \leq \max(p,q)$ and all seasons $i = 0, 1, \ldots, S-1$. With a similar technique, multiply both sides of (1.0.1) by X_{t-h} and take expectations:

$$\mathrm{E}\left[X_{t-h}\left(X_t - \sum_{k=1}^{p}\phi_t(k)X_{t-k}\right)\right] = \mathrm{E}\left[X_{t-h}\left(\varepsilon_t + \sum_{j=1}^{q}\theta_t(j)\varepsilon_{t-j}\right)\right].$$

Notice that for $0 \leq h \leq \max(p,q)$, then

$$\begin{aligned}
&\gamma_{t-h}(h) - \sum_{k=1}^{p}\phi_t(k)\gamma_{\min(t-k,t-h)}(|h-k|) \\
&= \mathrm{E}\left[\sum_{k=0}^{\infty}\psi_{t-h}(k)\varepsilon_{t-h-k}\left(\varepsilon_t + \sum_{j=1}^{q}\theta_t(j)\varepsilon_{t-j}\right)\right] \\
&= \sum_{j=h}^{q}\theta_t(j)\sigma^2_{t-j}\psi_{t-h}(j-h),
\end{aligned}$$

where $\theta_t(0) = 1$ and $\mathrm{E}(\varepsilon_{t-h-k}\varepsilon_{t-j}) = \sigma^2_{t-j}$, when $k = j-h$, and $\mathrm{E}(\varepsilon_{t-h-k}\varepsilon_{t-j}) = 0$, otherwise. In this way I get the general form of autocovariance function for $0 \leq h \leq \max(p,q)$,

$$\gamma_{t-h}(h) = \sum_{k=1}^{p}\phi_t(k)\gamma_{\min(t-k,t-h)}(|h-k|) + \sum_{j=h}^{q}\theta_t(j)\sigma^2_{t-j}\psi_{t-h}(j-h). \qquad (1.1.4)$$

In a straightforward way, (1.1.4) is actually an $S \times [\max(p,q)+1]$ dimensional linear system.

Table 1.1: Parameters in $\text{PARMA}_{12}(1,1)$ for Example 1.1.1.

season i	ϕ_i	θ_i	σ_i
0	0.198	0.687	11875.479
1	0.568	0.056	11598.254
2	0.560	-0.052	7311.452
3	0.565	-0.050	5940.845
4	0.321	0.470	4160.214
5	0.956	-0.389	4610.209
6	1.254	-0.178	15232.867
7	0.636	-0.114	31114.514
8	-1.942	2.393	32824.370
9	-0.092	0.710	29712.190
10	0.662	-0.213	15511.187
11	0.355	0.322	12077.991

The matrix associated with this linear system is invertible as long as the PARMA model is causal. See the appendix for a simple R code written to solve the linear system, for $\gamma_{t-h}(h)$, for all $0 \leq h \leq \max(p,q)$ and $i = 0, 1, \ldots, S-1$. Note that (1.1.4) only requires $\psi_{t-h}(j-h)$ for $j \leq q$, which great reduces the computation, compared with the first method in (1.1.2).

Example 1.1.1. Consider the $\text{PARMA}_{12}(1,1)$ model in Table 1.1.1, used in Tesfaye et al. [42] to fit the 72-year monthly observations of Fraser river.

By (1.1.4), when $0 \leq h \leq \max(1,1)$, I have $h = 0, 1$, and

$$\begin{cases} \gamma_t(0) - \phi_t(1)\gamma_{t-1}(1) = \theta_t(0)\psi_t(0)\sigma_t^2 + \theta_t(1)\psi_t(1)\sigma_{t-1}^2 \\ \gamma_{t-1}(1) - \phi_t(1)\gamma_{t-1}(0) = \theta_t(1)\sigma_{t-1}^2 \end{cases} \tag{1.1.5}$$

Note that $\theta_t(0) = 1$, $\psi_t(0) = 1$, $\psi_t(1) = \phi_t(1) + \theta_t(1)$, and (1.1.5) is a 12×2 dimensional linear system containing 24 unknown paramters $\gamma_t(0)$ and $\gamma_t(1)$, for $t = 0, 1, \ldots, 11$. Apply the computation in R, I can get $\gamma_t(0)$ and $\gamma_t(1)$ for all the seasons, which are shown in the

Table 1.2: Part of autocovariances in $\text{PARMA}_{12}(1,1)$ for Example 1.1.1.

season i	$\gamma_t(0)$	$\gamma_t(1)$	$\gamma_t(2)$	$\gamma_t(3)$
0	261385575	156364519	87564130	49473734
1	228262590	120832037	68270101	21914702
2	117569804	63754073	20465057	19564595
3	69938164	39038161	37320482	46799885
4	42959747	34336947	43058531	27385226
5	50262780	59246310	37680653	-73175828
6	302264368	165787551	-321959424	29620267
7	1059745614	258668383	-23797491	-15753939
8	1619934424	615947912	407757518	144753919
9	1298905828	671836226	238501860	47223368
10	600922799	290799803	57578361	32704509
11	301560482	159927070	90838576	50869602

first two columns of Table 1.1.1. For $h > 1$,

$$\gamma_{t-h}(h) = \phi_t \gamma_{t-h}(h-1), \tag{1.1.6}$$

therefore $\gamma_t(h)$ at higher lags $h > 1$ can be computed. Partial results for higher lags are shown in Table 1.1.1.

Chapter 2

Forecasting and forecast error

2.1 Forecasting

The PARMA prediction equations are based on orthogonal projection, to minimize the mean squared prediction error among all linear predictors. If the PARMA process has Gaussian innovations, then this will also minimize the mean squared prediction error among all predictors.

To simplify notation, I denote the first season to be forecast as season 0, and define the season of the oldest data point as season $S-1$. If the total number of available data is not a multiple of S, I discard a few($< S$) of the oldest observations, to obtain the data set $\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_{n-1}$, where $n = N \times S$.

Recall that $X_t = \tilde{X}_t - \mu_t$ is the mean-centered process in (1.0.1). Fix a probability space on which the PARMA$_S(p,q)$ model (1.0.1) is defined, and let

$$\tilde{\mathcal{H}}_n = \bar{\text{sp}}\{1, \tilde{X}_0, \ldots, \tilde{X}_{n-1}\} = \bar{\text{sp}}\{1, X_0, \ldots, X_{n-1}\}$$

denote the set of all linear combinations of these random variables in the Hilbert space of random variables on that probability space, with the inner product $\langle X, Y \rangle = \mathrm{E}(XY)$. Note that

$$\begin{aligned} P_{\tilde{\mathcal{H}}_n} \tilde{X}_n &= P_{\tilde{\mathcal{H}}_n} (X_n + \mu_n) \\ &= \mu_n + P_{\tilde{\mathcal{H}}_n} X_n \\ &= \mu_n + P_{\bar{\mathrm{sp}}\{1, X_0, \ldots, X_{n-1}\}} X_n \\ &= \mu_n + P_{\bar{\mathrm{sp}}\{X_0, \ldots, X_{n-1}\}} X_n, \end{aligned} \tag{2.1.1}$$

so that forecasting for the original data $\tilde{X}_t$ can be accomplished by forecasting the mean-centered process X_t, and then adding the seasonal mean. In order to develop a more efficient forecasting algorithm, it is useful to consider a transformed process (cf. Ansley [9]; Lund and Basawa [6]) defined by

$$W_t = \begin{cases} X_t, & t = 0, \ldots, m-1 \\ X_t - \sum_{k=1}^{p} \phi_t(k) X_{t-k}, & t \geq m \end{cases} \tag{2.1.2}$$

where $m = \max(p, q)$. Then it follows from (1.0.1) that, for $t \geq m$, the transformed process (2.1.2) has the moving average representation

$$W_t = \sum_{j=0}^{q} \theta_t(j) \varepsilon_{t-j}, \tag{2.1.3}$$

where $\theta_t(0) = 1$ for all t. Notice that $\phi_t(k)$ and $\theta_t(k)$ are periodic in S, such that $\phi_t(k) = \phi_{\langle t \rangle}(k)$ and $\theta_t(k) = \theta_{\langle t \rangle}(k)$, where $\langle t \rangle$ is the season corresponding to index t, so that $\langle t \rangle = t$

mod S.

Proposition 2.1.1.

$$\mathcal{H}_n = \bar{\text{sp}}\{X_0, \ldots, X_{n-1}\} = \bar{\text{sp}}\{W_0, \ldots, W_{n-1}\} \tag{2.1.4}$$

for all $n \geq 1$.

Proof. Define $\mathcal{H}_n^* = \bar{\text{sp}}\{W_0, \ldots, W_{n-1}\}$. For $j = 1$, obviously $\mathcal{H}_1 = \bar{\text{sp}}\{X_0\} = \bar{\text{sp}}\{W_0\} = \mathcal{H}_1^*$. Assume that when $j = n-1$, (2.1.4) holds, i.e. $\mathcal{H}_{n-1} = \bar{\text{sp}}\{X_0, \ldots, X_{n-2}\} = \bar{\text{sp}}\{W_0, \ldots, W_{n-2}\} = \mathcal{H}_{n-1}^*$. By (2.1.2),

$$W_{n-1} = \begin{cases} X_{n-1}, & n-1 < m \\ X_{n-1} - \sum_{k=1}^{p} \phi_{n-1}(k) X_{n-1-k}, & n-1 \geq m. \end{cases}$$

Then W_{n-1} can be expressed as a linear combination in the span $\mathcal{H}_n = \bar{\text{sp}}\{X_0, \ldots, X_{n-1}\}$. Together with the inductive assumption at $j = n-1$, $\mathcal{H}_{n-1} = \mathcal{H}_{n-1}^*$, I conclude that $\mathcal{H}_n^* \subset \mathcal{H}_n$. Next I will prove the other direction, $\mathcal{H}_n \subset \mathcal{H}_n^*$. By a transformation on (2.1.2), I have,

$$X_{n-1} = \begin{cases} W_{n-1}, & n-1 < m \\ W_{n-1} + \sum_{k=1}^{p} \phi_{n-1}(k) X_{n-1-k}, & n-1 \geq m \end{cases}$$

By the induction hypothesis, $\mathcal{H}_{n-1} = \bar{\text{sp}}\{X_0, \ldots, X_{n-2}\} = \bar{\text{sp}}\{W_0, \ldots, W_{n-2}\} = \mathcal{H}_{n-1}^*$, it follows that X_{n-1} can be expressed as a linear combination in the span

$$\mathcal{H}_n^* = \bar{\text{sp}}\{W_0, \ldots, W_{n-2}, W_{n-1}\}.$$

Then $\mathcal{H}_n \subset \mathcal{H}_n^*$. Together, this proves that

$$\mathcal{H}_n = \bar{\text{sp}}\{X_0, \ldots, X_{n-1}\} = \bar{\text{sp}}\{W_0, \ldots, W_{n-1}\} = \mathcal{H}_n^*$$

for all $n \geq 1$. □

Proposition 2.1.2. *The process W_t satisfies an invertibility condition*

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_{t,j} W_{t-j} \tag{2.1.5}$$

where $\pi_{t,0} = 1$ and $\sum_{j=0}^{\infty} \left|\pi_{t,j}\right| < \infty$ for all t. Also $\pi_{t,j} = \pi_{t+kS,j}$ for all j.

Proof. First define $W_{t-j} = 0$ if $t - j < 0$. By (2.1.2), when $t \geq m$,

$$X_t = W_t + \sum_{k=1}^{p} \phi_t(k) X_{t-k},$$

and then by Proposition 2.1.1,

$$X_t \in \bar{\text{sp}}\{W_t, X_{t-1}, \ldots, X_{t-p}\} = \bar{\text{sp}}\{W_t, W_{t-1}, \ldots, W_{t-p}\}.$$

Then X_t can be written as

$$X_t = a_t W_t + a_{t-1} W_{t-1} + \ldots + a_{t-p} W_{t-p} = \sum_{k=0}^{q} a_{t-k} W_{t-k},$$

where $a_t, \ldots, a_{t-p}$ are finite constants. Then by (1.0.3),

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_t(j) X_{t-j} = \sum_{j=0}^{\infty} \pi_t(j) \left(\sum_{k=0}^{q} a_{t-j-k} W_{t-j-k} \right).$$

Therefore I can write

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_{t,j} W_{t-j},$$

where $\pi_{t,j}$ is a finite linear combination of $\pi_t(j)$, such that

$$\pi_{t,j} = \begin{cases} a_{t-j} \sum_{k=0}^{j} \pi_t(j-k), & 0 \leq j < q \\ a_{t-j} \sum_{k=0}^{q} \pi_t(j-k), & j \geq q \end{cases}$$

Or

$$\pi_{t,j} = a_{t-j} \sum_{k=0}^{\min(q,j)} \pi_t(j-k).$$

If we let $a^* = \max\{a_t, \ldots, a_{t-p}\}$ which is bounded, by taking maximum of $p+1$ finite

constants. Then

$$\begin{aligned}
\sum_{j=0}^{\infty}\left|\pi_{t,j}\right| &= \sum_{j=0}^{\infty}\left|a_{t-j}\sum_{k=0}^{\min(q,j)}\pi_t(j-k)\right| \\
&\leq \left|a^*\right|\sum_{j=0}^{\infty}\sum_{k=0}^{\min(q,j)}|\pi_t(j-k)| \\
&= \left|a^*\right|\left(\sum_{j=q}^{\infty}\sum_{k=0}^{q}|\pi_t(j-k)| + \sum_{j=0}^{q-1}\sum_{k=0}^{j}|\pi_t(j-k)|\right) \\
&= \left|a^*\right|\left(\sum_{k=0}^{q}\sum_{j=q}^{\infty}|\pi_t(j-k)| + \sum_{k=0}^{j}\sum_{j=0}^{q-1}|\pi_t(j-k)|\right) \\
&= \left|a^*\right|\sum_{k=0}^{\min(q,j)}\sum_{j=k}^{\infty}|\pi_t(j-k)| < \infty,
\end{aligned}$$

since $\left|a^*\right| < \infty$ and $\sum_{j=0}^{\infty}|\pi_t(j)| < \infty$ for all t in (1.0.3). □

Define $\hat{X}_0 = \hat{W}_0 = 0$ and, for $n \geq 1$, let

$$\begin{aligned}
\hat{X}_n &= P_{\mathcal{H}_n}(X_n) \\
\hat{W}_n &= P_{\mathcal{H}_n}(W_n)
\end{aligned} \tag{2.1.6}$$

denote the one-step projections of X_n and W_n onto $\mathcal{H}_n$, respectively.

The next result computes the covariance function of the transformed process $\{W_t\}$. This result was stated without proof in Lund and Basawa [23, Equation (3.16)], in a different notation.

Proposition 2.1.3. *Given a PARMA$_S(p,q)$ process* (1.0.1)*, the covariance function* $C(j,\ell) =$

$\mathrm{E}(W_jW_\ell)$ *of the transformed process* (2.1.2) *is given by*

$$C(j,\ell) = \begin{cases} \gamma_j(\ell - j) & 0 \le j \le \ell \le m-1 \\ \gamma_j(\ell - j) - \sum_{k=1}^{p} \phi_\ell(k)\gamma_j(\ell - k - j) & 0 \le j \le m-1 < \ell \le 2m-1 \\ 0 & 0 \le j \le m-1, \ell \ge 2m \\ \sum_{k=0}^{q} \theta_j(k)\theta_\ell(k+\ell-j)\sigma^2_{j-k} & j, \ell \ge m \end{cases}$$

for all $j, \ell \in \mathbb{Z}$*, where I define* $\theta_j(0) = 1$ *for all* $j, \in \mathbb{Z}$*.*

Proof. For the first case, where $0 \le j \le \ell \le m-1$, I have from (2.1.2) that $W_j = X_j$ and $W_\ell = X_\ell$. Then $\mathrm{E}(W_jW_\ell) = \mathrm{E}(X_jX_\ell) = \gamma_j(\ell - j)$. For the second case, when $j \le m-1$, but $m \le \ell \le 2m-1$, I have

$$\begin{aligned} \mathrm{E}(W_jW_\ell) &= \mathrm{E}\big[X_j(X_\ell - \phi_\ell(1)X_{\ell-1} - \ldots - \phi_\ell(p)X_{\ell-p})\big] \\ &= \gamma_j(\ell - j) - \sum_{k=1}^{p} \phi_\ell(k)\gamma_\ell(\ell - k - j). \end{aligned}$$

In the third case, if $j \le m-1$ and $\ell \ge 2m$, then I have using (1.0.1) with $\theta_t(0) = 1$ that

$$\mathrm{E}(W_jW_\ell) = \mathrm{E}\left[X_j\Big(\sum_{k=0}^{q} \theta_\ell(k)\varepsilon_{\ell-k}\Big)\right],$$

where $\mathrm{E}(X_j\,\varepsilon_{\ell-k}) = 0$ for $k = 0, 1, \ldots, q$ since $\ell - k \ge \ell - q \ge 2m - q \ge 2m - m = m >$

$m-1 \geq j$. Finally, for the last case, when $m \leq j \leq \ell$, using (1.0.1) again I have

$$\begin{aligned} \mathrm{E}(W_j W_\ell) &= \mathrm{E}\left[\Big(\sum_{k=0}^{q} \theta_j(k) \varepsilon_{j-k} \Big) \Big(\sum_{r=0}^{q} \theta_\ell(r) \varepsilon_{\ell-r} \Big) \right] \\ &= \sum_{k=0}^{q} \sum_{r=0}^{q} \theta_j(k) \theta_\ell(r) \mathrm{E}(\varepsilon_{j-k} \varepsilon_{\ell-r}), \end{aligned}$$

and $\mathrm{E}(\varepsilon_{j-k}\varepsilon_{\ell-r}) = 0$ unless $j-k = \ell - r$. □

Remark 2.1.4. Since $\theta_0(k) = 0$ for $k > q$, it follows from the final case in Proposition 2.1.3 that $C(j, \ell) = 0$ whenever $\ell > m$ and $\ell > j + q$.

Using the covariance function $C(j, \ell)$ computed in Proposition 2.1.3, I can now apply the innovations algorithm from Brockwell and Davis [11, Proposition 5.2.2] to the transformed process (2.1.2) to compute the one-step ahead predictor

$$\hat{W}_n = \sum_{j=1}^{n} \theta_{n,j} \left(W_{n-j} - \hat{W}_{n-j} \right), \tag{2.1.7}$$

where $\theta_{n,1}, \ldots, \theta_{n,n}$ are the unique projection coefficients that minimize the mean squared error

$$v_n = \mathrm{E}\left(W_n - \hat{W}_n \right)^2. \tag{2.1.8}$$

The coefficients $\theta_{n,j}$ in (2.1.7) and v_n in (2.1.8) are computed from the system of equations

$$\begin{aligned} v_0 &= C(0,0) \\ \theta_{n,n-k} &= v_k^{-1}\left[C(n,k) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}v_j\right] \\ v_n &= C(n,n) - \sum_{j=0}^{n-1}\left(\theta_{n,n-j}\right)^2 v_j \end{aligned} \tag{2.1.9}$$

solved in the order v_0, $\theta_{1,1}$, v_1, $\theta_{2,2}$, $\theta_{2,1}$, v_2, $\theta_{3,3}$, $\theta_{3,2}$, $\theta_{3,1}$, $v_3, \ldots$ and so forth.

Proposition 2.1.5. *Given a PARMA$_S(p,q)$ process* (1.0.1)*, the innovations algorithm* (2.1.9) *applied to the transformed process* (2.1.2) *with covariance function $C(j,\ell)$ from Proposition 2.1.3 yields*

$$\hat{W}_n = \begin{cases} 0 & n = 0, \\ \displaystyle\sum_{j=1}^{n} \theta_{n,j}\left(W_{n-j} - \hat{W}_{n-j}\right) & 1 \le n < m, \\ \displaystyle\sum_{j=1}^{q} \theta_{n,j}\left(W_{n-j} - \hat{W}_{n-j}\right) & n \ge m. \end{cases} \tag{2.1.10}$$

In particular, I have $\theta_{n,j} = 0$ whenever $j > q$ and $n \ge m$.

Proof. If $n \ge m$ and $j > q$, then Remark 2.1.4 implies that $C(n, n-j) = \mathrm{E}(W_n W_{n-j}) = 0$. Since

$$\hat{W}_{n-j} \in \bar{\mathrm{sp}}\{W_0, \ldots, W_{n-j-1}\},$$

another application of Remark 2.1.4 shows that $\mathrm{E}(W_n \hat{W}_{n-j}) = 0$, and then it follows using the linearity of the expectation operator that

$$\mathrm{E}\left[W_n(W_{n-j} - \hat{W}_{n-j})\right] = 0. \tag{2.1.11}$$

The proof of the innovations algorithm in [11, Proposition 5.2.2] shows that the random variables

$$\{W_{n-j} - \hat{W}_{n-j} : j = 1, 2, \ldots, n\}$$

are uncorrelated, and hence the coefficients

$$\theta_{n,j} = v_{n-j}^{-1} \mathrm{E}\left[W_n (W_{n-j} - \hat{W}_{n-j})\right] \tag{2.1.12}$$

are uniquely defined. Then (2.1.10) follows from (2.1.11), (2.1.12), and (2.1.9). □

Remark 2.1.6. Proposition 2.1.5 shows the advantage of the transformed process (2.1.2) for computing the innovations, since the sum in (2.1.10) terminates after q lags when $n \geq m$. Since the forecast equations developed later in this paper all depend on the innovations, this fact will be used to speed up the forecast computations.

The next result gives the one-step ahead predictors $\hat{X}_n$ for the best linear predictor, that minimize the mean squared prediction error. Equation (2.1.13) was proven by Lund and Basawa [23, Equation (3.4)] in a different notation.

Theorem 2.1.7. *The one-step predictors* (2.1.6) *for a* $PARMA_S(p, q)$ *process* (1.0.1) *can be computed recursively using*

$$\hat{X}_n = \begin{cases} \displaystyle\sum_{j=1}^{n} \theta_{n,j} \left(X_{n-j} - \hat{X}_{n-j}\right) & 1 \leq n < m \\ \displaystyle\sum_{j=1}^{p} \phi_n(j) X_{n-j} + \sum_{j=1}^{q} \theta_{n,j} \left(X_{n-j} - \hat{X}_{n-j}\right) & n \geq m \end{cases} \tag{2.1.13}$$

where the coefficients $\theta_{n,j}$ *are computed via the innovations algorithm* (2.1.9) *applied to the transformed process* (2.1.2).

Proof. Project each side of (2.1.2) onto the space $\mathcal{H}_n$ in (2.1.4) to get

$$\hat{W}_t = \begin{cases} \hat{X}_t & t = 0, \ldots, m-1, \\ \hat{X}_t - \phi_t(1)X_{t-1} - \ldots - \phi_t(p)X_{t-p} & t \geq m. \end{cases} \tag{2.1.14}$$

Subtract each side of (2.1.2) from the corresponding expression in (2.1.14) to see that

$$\hat{W}_t - W_t = \hat{X}_t - X_t \tag{2.1.15}$$

for all $t \geq 0$. Solve for $\hat{X}_t$ in (2.1.14), substitute (2.1.10), and then use (2.1.15) to arrive at (2.1.13). □

Remark 2.1.8. Theorem 2.1.7 is the basis for our forecasting computations. The fact that the sum in (2.1.13) terminates after q innovations when $n \geq m$ simplifies and speeds up the computations.

Recall from (2.1.4) that $\mathcal{H}_n = \bar{\text{sp}}\{X_0, \ldots, X_{n-1}\}$. The next result gives the h-step ahead predictors $P_{\mathcal{H}_n} X_{n+h}$ that minimize the mean squared prediction error. This result was proven by Lund and Basawa [23, Equation (3.36)] in a different notation.

Theorem 2.1.9. *The h-step predictors for a $PARMA_S(p,q)$ process* (1.0.1) *can be computed recursively using*

$$P_{\mathcal{H}_n} X_{n+h} = \sum_{j=h+1}^{n+h} \theta_{n+h,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right) \tag{2.1.16}$$

when $n < m$ *and* $0 \leq h \leq m - n - 1$, *and*

$$P_{\mathcal{H}_n} X_{n+h} = \sum_{k=1}^{p} \phi_{n+h}(k) P_{\mathcal{H}_n} X_{n+h-k} + \sum_{j=h+1}^{q} \theta_{n+h,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right). \tag{2.1.17}$$

otherwise, where the coefficients $\theta_{n,j}$ *are computed via the innovations algorithm* (2.1.9) *applied to the transformed process* (2.1.2).

Proof. Since $\mathcal{H}_n$ is a subspace of $\mathcal{H}_{n+h}$, I can write

$$\begin{aligned} P_{\mathcal{H}_n} W_{n+h} &= P_{\mathcal{H}_n} P_{\mathcal{H}_{n+h}} W_{n+h} \\ &= P_{\mathcal{H}_n} \hat{W}_{n+h} \\ &= P_{\mathcal{H}_n} \left(\sum_{j=1}^{n+h} \theta_{n+h,j} \left(W_{n+h-j} - \hat{W}_{n+h-j} \right) \right). \end{aligned}$$

Since $W_{n+h-j} - \hat{W}_{n+h-j}$ is orthogonal to $\mathcal{H}_n$ for $j \leq h$, and contained in $\mathcal{H}_n$ for $j > h$, it follows using Brockwell and Davis [11, Proposition 2.3.2] that

$$P_{\mathcal{H}_n} W_{n+h} = \sum_{j=h+1}^{n+h} \theta_{n+h,j} \left(W_{n+h-j} - \hat{W}_{n+h-j} \right). \tag{2.1.18}$$

Since each $W_{n+h-j} - \hat{W}_{n+h-j}$ lies in $\mathcal{H}_n$, for $j > h$, and $W_{n+h} - P_{\mathcal{H}_n} W_{n+h}$ is orthogonal to any element of $\mathcal{H}_n$, it follows that the mean square error of the h-step prediction is

$$\mathrm{E} \left(W_{n+h} - P_{\mathcal{H}_n} W_{n+h} \right)^2 = C(n+h, n+h) - \sum_{j=h+1}^{n+h} \left(\theta_{n+h,j} \right)^2 v_{n+h-j}.$$

From (2.1.2), I can write $P_{\mathcal{H}_n} W_{n+h} = P_{\mathcal{H}_n} X_{n+h}$ when $0 \leq h \leq m-n-1$, and

$$P_{\mathcal{H}_n} W_{n+h} = P_{\mathcal{H}_n} X_{n+h} - \sum_{k=1}^{p} \phi_{n+h}(k) P_{\mathcal{H}_n} X_{n+h-k} \tag{2.1.19}$$

when $h \geq m-n$. Substitute (2.1.15) into (2.1.18) to get

$$P_{\mathcal{H}_n} W_{n+h} = \sum_{j=h+1}^{n+h} \theta_{n+h,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right). \tag{2.1.20}$$

If $0 \leq h \leq m-n-1$, then (3.0.6) and (2.1.2) imply that (2.1.16) holds. If $h \geq m-n$, substitute (3.0.6) into (2.1.19) to get (2.1.17), using the fact that $\theta_{n+h,j} = 0$ if $j > q$ and $h \geq m-n$. □

Given a $\text{PARMA}_S(p,q)$ time series data set $\tilde{X}_0, \ldots, \tilde{X}_{n-1}$, I can forecast future values using the h-step ahead predictors from Theorem 2.1.9. Note that this requires computing all of the innovations

$$X_t - \hat{X}_t \quad \text{for } t = 0, 1, \ldots, n-1$$

using the recursive equation (2.1.13). Complete details will be provided in Chapter 3. In the next chapter, I explicitly compute the forecast errors, which will be needed to compute prediction bands for this forecast.

2.2 Forecast errors

The next theorem is the main theoretical result of this chapter. It explicitly computes the variance of the forecast errors, and a simpler asymptotic variance that is useful for computations. Formula (2.2.12) for the covariance matrix of the forecast errors was established

by Lund and Basawa [23, Equation (3.41)] in a different notation. However, that paper did not develop an explicit formula for the forecast error variance. I begin with the data set $\{X_0, X_1, \ldots, X_{n-1}\}$ as in Chapter 2, where $n = N \times S$.

Theorem 2.2.1. *Define* $\chi_j(0) = 1$ *for all* $j \geq 0$, $\chi_j(j-\ell) = 0$ *for all* $j \geq 0$ *and* $\ell > j$, *and recursively*

$$\chi_j(\ell) = \sum_{k=1}^{\min(p,\ell)} \phi_j(k)\chi_{j-k}(\ell - k) \tag{2.2.1}$$

for all $j \geq 0$ *and* $0 \leq \ell < j$.

Then the mean-squared error $\sigma_n^2(h) = \mathrm{E}\left[\left(X_{n+h} - P_{\mathcal{H}_n} X_{n+h}\right)^2\right]$ *of the* h*-step predictors* $P_{\mathcal{H}_n} X_{n+h}$ *for the* $PARMA_S(p,q)$ *process* (1.0.1) *can be computed recursively using*

$$\sigma_n^2(h) = \sum_{j=0}^{h} \left(\sum_{k=0}^{j} \chi_h(k)\theta_{n+h-k,j-k} \right)^2 v_{n+h-j} \tag{2.2.2}$$

when $n \geq m := \max(p,q)$, *where the coefficients* $\theta_{n+h-k,j-k}$ *and* v_{n+h-j} *are computed via the innovations algorithm* (2.1.9) *applied to the transformed process* (2.1.2). *Furthermore, the asymptotic mean squared error is given by*

$$\sigma_n^2(h) \to \sum_{j=0}^{h} \psi_h^2(j)\sigma_{h-j}^2 \quad \text{as } n = NS \to \infty, \tag{2.2.3}$$

where

$$\psi_h(j) = \sum_{k=0}^{j} \chi_h(k)\theta_{h-k}(j-k).$$

Proof. Recall that $X_t = \tilde{X}_t - \mu_t$ is the mean-centered process in (1.0.1), and W_t is the transformed process in (2.1.2). Note that the mean squared prediction error $\sigma_n^2(h) =$

$\mathrm{E}(X_{n+h} - P_{\mathcal{H}_n} X_{n+h})^2$ for the mean-centered PARMA process (1.0.1) is not the same as the mean squared prediction error $\mathrm{E}(W_{n+h} - P_{\mathcal{H}_n} W_{n+h})^2$ for the transformed process (2.1.2). When $n \geq m := \max(p, q)$, Theorem 2.1.9 implies that (2.1.17) holds for any $h \geq 0$, and note that the second term in (2.1.17) vanishes when $h + 1 > q$. Write

$$X_{n+h} = \hat{X}_{n+h} + (X_{n+h} - \hat{X}_{n+h}),$$

and since $n = NS$, then $\phi_{n+h}(j) = \phi_h(j)$. Substitute (2.1.13) with $n \geq m$ to get

$$\begin{aligned} X_{n+h} = \phi_h(1) X_{n+h-1} + \ldots + \phi_h(p) X_{n+h-p} \\ + \sum_{j=0}^{q} \theta_{n+h,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right), \end{aligned} \tag{2.2.4}$$

with $\theta_{n,0} = 1$ for all n. Subtract (2.1.17) from (2.2.4) and rearrange terms to get

$$\begin{aligned} X_{n+h} - P_{\mathcal{H}_n} X_{n+h} - \sum_{k=1}^{p} \phi_h(k) \left(X_{n+h-k} - P_{\mathcal{H}_n} X_{n+h-k} \right) \\ = \sum_{j=0}^{h} \theta_{n+h,j} \left(X_{n+h-j} - \hat{X}_{n+h-j} \right). \end{aligned} \tag{2.2.5}$$

Define the random vectors

$$M_n = \begin{pmatrix} X_n - \hat{X}_n \\ \vdots \\ X_{n+h} - \hat{X}_{n+h} \end{pmatrix} \quad \text{and} \quad F_n = \begin{pmatrix} X_n - P_{\mathcal{H}_n} X_n \\ \vdots \\ X_{n+h} - P_{\mathcal{H}_n} X_{n+h}. \end{pmatrix} \tag{2.2.6}$$

Write

$$\Phi_h = \Big[-\phi_j(j-\ell)\Big]_{j,\ell=0}^{h} \tag{2.2.7}$$

where I define $\phi_j(0) = -1$ for all j, and $\phi_j(k) = 0$ for $k > p$ or $k < 0$. Note that $\phi_j(k)$ is periodic in S, such that $\phi_j(k) = \phi_{\langle j \rangle}(k)$, where $\langle j \rangle$ is the season corresponding to index j, so that $\langle j \rangle = j \bmod S$. Then from the innovations algorithm, write

$$\Theta_n = \Big[\theta_{n+j,j-\ell}\Big]_{j,\ell=0}^{h} \tag{2.2.8}$$

where I define $\theta_{n,0} = 1$, and $\theta_{n,k} := 0$ for $k > q$ or $k < 0$. Then I can use (5.1.3) to write

$$\Phi_h F n = \Theta_n M_n, \tag{2.2.9}$$

where I note that

$$\Phi_h = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ -\phi_1(1) & 1 & 0 & 0 & \dots & 0 \\ -\phi_2(2) & -\phi_2(1) & 1 & 0 & \dots & 0 \\ -\phi_3(3) & -\phi_3(2) & -\phi_3(1) & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\phi_h(h) & -\phi_h(h-1) & -\phi_h(h-2) & \dots & \dots & 1 \end{pmatrix}$$

and

$$\Theta_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \theta_{n+1,1} & 1 & 0 & \dots & 0 \\ \theta_{n+2,2} & \theta_{n+2,1} & 1 & \dots & 0 \\ \theta_{n+3,3} & \theta_{n+3,2} & \theta_{n+3,1} & 1 & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta_{n+h,h} & \theta_{n+h,h-1} & \theta_{n+h,h-2} & \cdots & 1 \end{pmatrix}$$

are lower triangular matrices.

The entries of the innovations vector M_n in (2.2.6) are uncorrelated, with covariance matrix

$$V_n := \mathrm{E}\left[M_n M_n'\right] = \mathrm{diag}\left(v_n, v_{n+1}, \dots, v_{n+h}\right), \tag{2.2.10}$$

where I recall that

$$v_n = \mathrm{E}\left[\left(X_n - \hat{X}_n\right)^2\right] \tag{2.2.11}$$

are the one step ahead prediction errors. Then the covariance matrix of the vector $F_n = \Phi_h^{-1}\Theta_n M_n$ of prediction errors is

$$C_n := \mathrm{E}\left[F_n F_n'\right] = \Psi_n V_n \Psi_n' \quad \text{where} \quad \Psi_n = \Phi_h^{-1}\Theta_n \tag{2.2.12}$$

and $'$ denotes the transpose.

In order to compute the matrix C_n in (2.2.12), I first need to compute the inverse matrix

$$\Phi_h^{-1} = \left[\chi_j(j-\ell)\right]_{j,\ell=0}^{h} \tag{2.2.13}$$

and show that (2.2.1) holds. An elegant way to compute the inverse matrix (5.1.11) is to use operator notation along with the z-transform. Use (1.0.1) to write

$$\Phi(B)X_{n+j} = \Theta(B)\varepsilon_{n+j} \quad \text{for } j \geq 0, \tag{2.2.14}$$

where

$$\begin{aligned}
\Phi(z) &= 1 - \sum_{k=1}^{p} \phi_{n+j}(k)z^k = -\sum_{k=0}^{p} \phi_{n+j}(k)z^k = -\sum_{k=0}^{p} \phi_j(k)z^k \\
\Theta(z) &= 1 + \sum_{k=1}^{q} \theta_{n+j}(k)z^k = \sum_{k=0}^{q} \theta_{n+j}(k)z^k = \sum_{k=0}^{q} \theta_j(k)z^k,
\end{aligned}$$

since $n = NS$, $BX_t = X_{t-1}$ is the backward shift operator. Then the infinite order moving average representation (1.0.2) for the mean-centered process X_t can be written in the form

$$X_{n+j} = \Psi(B)\varepsilon_{n+j} \quad \text{for } j \geq 0, \tag{2.2.15}$$

where

$$\Psi(z) = \Phi^{-1}(z)\Theta(z) \tag{2.2.16}$$

Notice that $\Phi_h = \Phi$, however $\Theta_n \neq \Theta$ and $\Psi_n \neq \Psi$. Write

$$\begin{aligned}
\Psi(z) &= \sum_{k=0}^{\infty} \psi_j(k)z^k \\
\Phi^{-1}(z) &= \sum_{k=0}^{\infty} \chi_j(k)z^k
\end{aligned}$$

extending the notation (5.1.11). If $n \geq m = \max(p, q)$, then the infinite order moving

average representation (2.1.3) of the transformed process W_t can be written in the form

$$W_{n+j} = \Theta(B)\varepsilon_{n+j} \quad \text{for } j \geq 0 \text{ and } n \geq m. \tag{2.2.17}$$

Equating (2.2.15) and (2.2.17) and using (2.2.16) shows that

$$X_{n+j} = \Phi^{-1}(B)W_{n+j} \quad \text{for } j \geq 0. \tag{2.2.18}$$

Since

$$\Phi^{-1}(z)\Phi(z) = I,$$

I have

$$\left(\sum_{k=0}^{\infty} \chi_{j+k}(k)z^k\right)\left(1 - \sum_{k=1}^{p} \phi_j(k)z^k\right) = 1$$

for all $j \geq 0$.

By expanding the above equation, I have

$$\Big(\chi_j(0) + \chi_{j+1}(1)z + \cdots + \chi_{j+k}(k)z^k + \cdots\Big)\Big(1 - \phi_j(1)z - \cdots - \phi_j(p)z^p\Big) = 1,$$

which could be separated as

$$\begin{aligned}&\chi_j(0) + \Big(\chi_{j+1}(1) - \chi_j(0)\phi_j(1)\Big)z + \cdots \\ &+ \Big(\chi_{j+p}(p) - \chi_{j+p-1}(p-1)\phi_j(1) - \cdots - \chi_j(0)\phi_j(p)\Big)z^p + \cdots = 1\end{aligned}$$

Setting every coefficient of z^k to 0, for $k = 1, 2, \ldots$, I have for all $j \geq 0$,

$$\chi_j(0) = 1$$

$$\chi_{j+1}(1) = \chi_j(0)\phi_j(1)$$

$$\vdots$$

$$\chi_{j+p}(p) = \chi_{j+p-1}(p-1)\phi_j(1) - \cdots - \chi_j(0)\phi_j(p).$$

Hence,

$$\chi_j(\ell) = \sum_{k=1}^{\ell} \phi_j(k)\chi_{j-k}(\ell - k) = \sum_{k=1}^{min(p,\ell)} \phi_j(k)\chi_{j-k}(\ell - k),$$

using the fact that $\phi_j(k) = 0, k > p$ and $\chi_j(\ell - k) = 0, k > \ell$, which proves (2.2.1).

Since $\Phi_h = \Phi$, this shows that the inverse matrix (5.1.11) can be computed for any $h \geq 0$ by taking $\chi_j(0) = 1$ for all $j \geq 0$, $\chi_j(j - \ell) = 0$ for all $j \geq 0$ and $\ell > j$, and recursively applying the formula (2.2.1) for all $j > 0$ and $0 \leq \ell < j$.

Now I can proceed to compute the matrix C_n in (2.2.12). First write

$$\Psi_n = \Phi_h^{-1}\Theta_n = \Big[\psi_{n+j,j-\ell}\Big]_{j,\ell=0}^{h}$$

where

$$\begin{aligned}\psi_{n+j,j-\ell} &= \sum_{s=0}^{h} \left(\Phi_h^{-1}\right)_{j,s} (\Theta n)_{s,\ell} \\ &= \sum_{s=0}^{h} \chi_j(j-s)\theta_{n+s,s-\ell} \\ &= \sum_{s=\ell}^{j} \chi_j(j-s)\theta_{n+s,s-\ell}\end{aligned} \tag{2.2.19}$$

since $\chi_j(j-s) = 0$ for $s > j$, and $\theta_{n+s,s-\ell} = 0$ for $s < \ell$. Substitute $s = j - k$ in the sum (5.1.5) to obtain

$$\psi_{n+j,j-\ell} = \sum_{k=0}^{j-\ell} \chi_j(k)\theta_{n+j-k,j-\ell-k} \tag{2.2.20}$$

Now it follows from (5.1.5), (5.1.9), and (5.1.14) that

$$\begin{aligned}
\sigma_n^2(h) &= \mathrm{E}\left[\left(X_{n+h} - P_{\mathcal{H}_n} X_{n+h}\right)^2\right] \\
&= C_{h,h} \\
&= \sum_{u=0}^{h} \sum_{w=0}^{h} \Psi_{h,u} V_{u,w} \Psi'_{w,h} \\
&= \sum_{u=0}^{h} \sum_{w=0}^{h} \Psi_{h,u} V_{u,w} \Psi_{h,w} \\
&= \sum_{u=0}^{h} \Psi_{h,u} V_{u,u} \Psi_{h,u} \\
&= \sum_{u=0}^{h} \psi_{n+h,h-u}^2 v_{n+u} \\
&= \sum_{u=0}^{h} \left(\sum_{k=0}^{h-u} \chi_h(k) \theta_{n+h-k,h-u-k} \right)^2 v_{n+u}
\end{aligned} \tag{2.2.21}$$

Substitute $u = h - j$ to obtain (2.2.2). □

Next I require a few preparatory results.

Lemma 2.2.2. *With the innovations algorithm (2.1.9) applied to the transformed process (2.1.2), I have* $\left| v_r - \sigma_r^2 \right| \to 0$, *as* $r \to \infty$.

Proof. Recall that $\mathcal{H}_r = \bar{\mathrm{sp}}\{X_0, \ldots, X_{r-1}\} = \bar{\mathrm{sp}}\{W_0, \ldots, W_{r-1}\}$, and define

$$H_r = \bar{\mathrm{sp}}\{W_j, -\infty < j \leq r-1\}.$$

By the invertibility condition of W_r process in Proposition 2.1.2,

$$\sigma_r^2 \ = \mathrm{E}(\varepsilon_r^2) = \mathrm{E}(W_r + \sum_{j=1}^{\infty} \pi_r(j) W_{r-j})^2 = \mathrm{E}(W_r - P_{H_r} W_r)^2$$

where

$$\sum_{j=1}^{\infty} \pi_r(j) W_{r-j} = P_{H_r}(\varepsilon_r - W_r) = -P_{H_r} W_r,$$

since $\varepsilon_r \perp H_r$ and $W_{r-j} \in H_r, j = 1, 2, \ldots, \infty$. Since $\mathcal{H}_r \subseteq H_r$, I have $P_{\mathcal{H}_r} W_r \in H_r$, thus by the Projection Theorem in Brockwell and Davis [11, Theorem 2.3.1], it follows that

$$\sigma_r^2 \ = \mathrm{E}(W_r - P_{H_r} W_r)^2 \leq \mathrm{E}(W_r - P_{\mathcal{H}_r} W_r)^2 = \mathrm{E}(W_r - \hat{W}_r)^2 = v_r.$$

On the other hand,

$$-\sum_{j=1}^{r} \pi_r(j) W_{r-j} \in \mathcal{H}_r,$$

so that by another application of the Projection Theorem in Brockwell and Davis [11, Theorem 2.3.1], I obtain

$$\begin{aligned} v_r = \mathrm{E}(W_r - \hat{W}_r)^2 &= \mathrm{E}(W_r - P_{\mathcal{H}_r} W_r)^2 \\ &\leq \mathrm{E}\big(W_r - (-\sum_{j=1}^{r} \pi_r(j) W_{r-j})\big)^2 \\ &= \mathrm{E}\big(\pi_r(0) W_r + \sum_{j=1}^{r} \pi_r(j) W_{r-j}\big)^2 \\ &= \mathrm{E}(\sum_{j=0}^{r} \pi_r(j) W_{r-j})^2 \end{aligned}$$

Therefore, since $\varepsilon_r \perp H_r$,

$$\begin{aligned}
v_r &\leq \mathrm{E}\left(\sum_{j=0}^{r} \pi_r(j) W_{r-j}\right)^2 \\
&= \mathrm{E}\left(\varepsilon_r - \sum_{j>r} \pi_r(j) W_{r-j}\right)^2 \\
&= \mathrm{E}(\varepsilon_r)^2 + \mathrm{E}\left(\sum_{j>r} \pi_r(j) W_{r-j}\right)^2 \\
&= \sigma_r^2 + \mathrm{E}\left(\sum_{j>r} \pi_r(j) W_{r-j} \sum_{k>r} \pi_r(k) W_{r-k}\right) \\
&\leq \sigma_r^2 + \sum_{j,k>r} \Big(\mid \pi_r(j) \mid\mid \pi_r(k) \mid \mathrm{E} \mid W_{r-j} W_{r-k} \mid \Big) \\
&\leq \sigma_r^2 + \sum_{j,k>r} \Big(\mid \pi_r(j) \mid\mid \pi_r(k) \mid \sqrt{C(r-j, r-j) C(r-k, r-k)} \Big) \\
&\leq \sigma_r^2 + \left(\sum_{j>r} \mid \pi_r(j) \mid\right)^2 M,
\end{aligned}$$

where $M = \max\{C(i,i) : i = 0, 1, \ldots, S-1\} < \infty$. In summary, I have proved

$$\sigma_r^2 \leq v_r \leq \sigma_r^2 + \big(\sum_{j>r} \mid \pi_r(j) \mid \big)^2 M$$

where $\big(\sum_{j>r} \mid \pi_r(j) \mid \big)^2 \to 0$ uniformly over all seasons, as $r \to \infty$. Hence , as $r \to \infty$, I have

$$|v_r - \sigma_r^2| \leq \big(\sum_{j>r} \mid \pi_r(j) \mid \big)^2 M \to 0,$$

which completes the proof. □

Lemma 2.2.3. *Let W_r be given by (2.1.2) and $\hat{W}_r$ be given by (2.1.10). Then*

$$\lim_{r\to\infty} \mathrm{E}\big[(W_r - \hat{W}_r - \varepsilon_r)^2\big] = 0.$$

Proof. By the causal representation $W_r = \sum_{j=0}^{q} \theta_r(j)\varepsilon_{r-j}$ in (2.1.3), I may write

$$\mathrm{E}[\varepsilon_r(W_r - \hat{W}_r)] = \mathrm{E}[\varepsilon_r(\varepsilon_r + \sum_{j=1}^{q} \theta_r(j)\varepsilon_{r-j} - \hat{W}_r)] = \mathrm{E}(\varepsilon_r^2) = \sigma_r^2,$$

since $\varepsilon_r \perp \varepsilon_{r-j}$, for $j = 1, \ldots, q$, and $\varepsilon_r \perp \mathcal{H}_r$, therefore $\varepsilon_r \perp \hat{W}_r$. Hence $\mathrm{E}[\varepsilon_r(W_r - \hat{W}_r)] = \sigma_r^2$. Then

$$\begin{aligned}
&\mathrm{E}\big[(W_r - \hat{W}_r - \varepsilon_r)^2\big] \\
&= \mathrm{E}\big[(W_r - \hat{W}_r)^2\big] - 2\mathrm{E}[\varepsilon_r(W_r - \hat{W}_r)] + \mathrm{E}(\varepsilon_r^2) \\
&= v_r - 2\sigma_r^2 + \sigma_r^2 \\
&= v_r - \sigma_r^2 \to 0,
\end{aligned}$$

using Lemma 2.2.2. □

Proposition 2.2.4. *Let $\theta_{r,k}$ be the projection coefficients from (2.1.7) and let $\theta_r(k)$ be the moving average coefficients from (1.0.1). Then*

$$|\theta_{r,k} - \theta_r(k)| \to 0 \quad \textit{as } r \to \infty, \textit{ for all } k = 1, 2, \ldots$$

Proof. I know that by (2.1.12)

$$\theta_{r,k} = v_{r-k}^{-1}\mathrm{E}\big[W_r(W_{r-k} - \hat{W}_{r-k})\big].$$

I also have

$$\theta_r(k) = \sigma_{r-k}^{-2}\mathrm{E}(W_r\varepsilon_{r-k}),$$

since

$$\mathrm{E}(W_r\varepsilon_{r-k}) = \mathrm{E}\big(\theta_r(k)\varepsilon_{r-k}^2\big) = \theta_r(k)\sigma_{r-k}^2,$$

using the causal representation (2.1.3). By the triangle inequality,

$$\begin{aligned}
\Big|\theta_{r,k} - \theta_r(k)\Big| &\leq \Big|\theta_{r,k} - \sigma_{r-k}^{-2}\mathrm{E}\big(W_r(W_{r-k} - \hat{W}_{r-k})\big)\Big| \\
&\quad + \Big|\sigma_{r-k}^{-2}\mathrm{E}\big(W_r(W_{r-k} - \hat{W}_{r-k} - \varepsilon_{r-k})\big)\Big| \\
&= \Big|\theta_{r,k} - \sigma_{r-k}^{-2}\theta_{r,k}v_{r-k}\Big| \\
&\quad + \Big|\sigma_{r-k}^{-2}\mathrm{E}\big(W_r(W_{r-k} - \hat{W}_{r-k} - \varepsilon_{r-k})\big)\Big| \\
&\leq \Big|\theta_{r,k}\Big|\Big|1 - \sigma_{r-k}^{-2}v_{r-k}\Big| \\
&\quad + \Big|\sigma_{r-k}^{-2}\Big| C(r,r)^{\frac{1}{2}}\big[\mathrm{E}(W_r - \hat{W}_r - \varepsilon_r)^2\big]^{\frac{1}{2}},
\end{aligned} \tag{2.2.22}$$

using the Cauchy-Schwarz inequality. Note that $\theta_{r,k}$ is uniformly bounded in r over all seasons, since (2.1.12) and the Cauchy-Schwarz Inequality imply,

$$\begin{aligned}
\theta_{r,k} &= v_{r-k}^{-1}\mathrm{E}\big(W_r(W_{r-k} - \hat{W}_{r-k})\big) \\
&\leq v_{r-k}^{-1}\sqrt{C(r,r)}\sqrt{\mathrm{E}(W_{r-k} - \hat{W}_{r-k})^2} \\
&= v_{r-k}^{-1/2}\sqrt{C(r,r)}.
\end{aligned}$$

Then, as $r \to \infty$, the first term on the right-hand side of (2.2.22) approaches 0 by Lemma 2.2.2. The second term on the right-hand side of (2.2.22) approaches 0 by Lemma 2.2.3 and the fact that $\sigma_{r-k}^{-2}C(r,r)^{\frac{1}{2}}$ is uniformly bounded in r. Thus, $\Big|\theta_{r,k} - \theta_r(k)\Big| \to 0$ as

$r \to \infty$. $\square$

Proof of Theorem 2.2.1 continued. Using the covariance function $C(j, \ell)$ computed in Proposition 2.1.3, I can now apply the innovations algorithm from Brockwell and Davis [11, Proposition 5.2.2] to the transformed process (2.1.2) to compute the one-step ahead predictor

$$\hat{W}_n = \sum_{j=1}^{n} \theta_{n,j} \left(W_{n-j} - \hat{W}_{n-j} \right)$$

for $n > 0$, where $\hat{W}_0 = 0$. Proposition 2.2.4 shows that

$$\left| \theta_{s,\ell} - \theta_s(\ell) \right| \to 0 \quad \text{as } s \to \infty, \quad \text{for all } \ell > 0. \tag{2.2.23}$$

Note that (2.2.23) also holds for $\ell = 0$, since $\theta_{s,0} = \theta_s(0) = 1$ by definition. Substitute $s = n + h - k$ and $\ell = j - k$ to see that

$$\left| \theta_{n+h-k,j-k} - \theta_{n+h-k}(j-k) \right| \to 0 \quad \text{as } n \to \infty \tag{2.2.24}$$

for all $j \geq k \geq 0$. Since $n = NS$, then $\theta_{n+h-k}(j-k) = \theta_{NS+h-k}(j-k) = \theta_{h-k}(j-k)$. Therefore (5.1.19) is better written as

$$\left| \theta_{NS+h-k,j-k} - \theta_{h-k}(j-k) \right| \to 0 \quad \text{as } N \to \infty. \tag{2.2.25}$$

The mean-centered process X_t has moving average representation (2.2.15), which can be related to the moving average representation (2.2.17) of the transformed process W_t by the

relation (2.2.16). Expanding both sides I obtain

$$\sum_{j=0}^{\infty} \psi_h(j) z^j = \Big(\sum_{j=0}^{\infty} \chi_h(j) z^j \Big) \Big(\sum_{j=0}^{q} \theta_h(j) z^j \Big).$$

Expanding the product on the right-hand side and equating coefficients, we have

$$\begin{aligned}
\sum_{j=0}^{\infty} \psi_h(j) z^j &= \Big(\sum_{j=0}^{\infty} \chi_h(j) z^j \Big) \Big(\theta_h(0)1 + \theta_h(1)z + \cdots + \theta_h(q) z^q \Big) \\
&= \chi_h(0)\theta_h(0)1 + \Big(\chi_h(0)\theta_h(1) + \chi_h(1)\theta_{h-1}(0) \Big) z \\
&\quad + \cdots + \Big(\sum_{k=0}^{j} \chi_h(k)\theta_{h-k}(j-k) \Big) z^j + \cdots \\
&= \sum_{j=0}^{\infty} \Big(\sum_{k=0}^{j} \chi_h(k)\theta_{h-k}(j-k) \Big) z^k.
\end{aligned}$$

Then

$$\psi_h(j) = \sum_{k=0}^{j} \chi_h(k)\theta_{h-k}(j-k) \tag{2.2.26}$$

for all $h \geq 0$ and $j \geq 0$. Now it follows from (5.1.5), (2.2.25) and (2.2.26) that

$$\begin{aligned}
&\lim_{N\to\infty} \Big| \psi_h(j) - \psi_{NS+h,j} \Big| \\
&= \lim_{N\to\infty} \left| \sum_{k=0}^{j} \chi_h(k)\theta_{h-k}(j-k) - \sum_{k=0}^{j} \chi_h(k)\theta_{NS+h-k,j-k} \right| \\
&\leq \lim_{N\to\infty} \sum_{k=0}^{j} \big| \chi_h(k) \big| \Big| \theta_{h-k}(j-k) - \theta_{NS+h-k,j-k} \Big| \\
&= 0,
\end{aligned} \tag{2.2.27}$$

where for each $h \geq 0$, $k = 0, \ldots, j$ and $0 \leq j \leq h$, $\chi_h(k)$ is a constant independent of

$n = NS$. Notice that from (2.2.25), as $N \to \infty$ I have

$$\sum_{k=0}^{j} \chi_h(k)\theta_{NS+h-k,j-k} \to \sum_{k=0}^{j} \chi_h(k)\theta_{h-k}(j-k)$$

and therefore,

$$\left(\sum_{k=0}^{j} \chi_h(k)\theta_{NS+h-k,j-k}\right)^2 \to \left(\sum_{k=0}^{j} \chi_h(k)\theta_{h-k}(j-k)\right)^2 \tag{2.2.28}$$

Furthermore, by (2.2.28), since h is finite and σ^2_{h-j} is periodic with period S and thus bounded, then as $N \to \infty$,

$$\lim_{N\to\infty} \sum_{j=0}^{h} \left| \left(\sum_{k=0}^{j} \chi_h(k)\theta_{NS+h-k}(j-k)\right)^2 - \left(\sum_{k=0}^{j} \chi_h(k)\theta_{NS+h-k,j-k}\right)^2 \right| \sigma^2_{h-j} \to 0 \tag{2.2.29}$$

Then it follows from (2.2.2), (2.2.27), (2.2.29) and Lemma 2.2.2 that

$$
\begin{aligned}
&\lim_{n\to\infty}\left|\sigma_n^2(h)-\sum_{j=0}^{h}\psi_{n+h}^2(j)\sigma_{n+h-j}^2\right| \\
&=\lim_{N\to\infty}\left|\sigma_{NS}^2(h)-\sum_{j=0}^{h}\psi_h^2(j)\sigma_{h-j}^2\right| \\
&=\lim_{N\to\infty}\left|\sigma_{NS}^2(h)-\sum_{j=0}^{h}\left(\sum_{k=0}^{j}\chi_h(k)\theta_{h-k}(j-k)\right)^2\sigma_{h-j}^2\right| \\
&\le\lim_{N\to\infty}\left|\sigma_{NS}^2(h)-\sum_{j=0}^{h}\left(\sum_{k=0}^{j}\chi_h(k)\theta_{NS+h-k,j-k}\right)^2\sigma_{h-j}^2\right| \\
&+\lim_{N\to\infty}\sum_{j=0}^{h}\left|\left(\sum_{k=0}^{j}\chi_h(k)\theta_{h-k}(j-k)\right)^2-\left(\sum_{k=0}^{j}\chi_h(k)\theta_{NS+h-k,j-k}\right)^2\right|\sigma_{h-j}^2 \\
&=\lim_{N\to\infty}\left|\sigma_{NS}^2(h)-\sum_{j=0}^{h}\left(\sum_{k=0}^{j}\chi_h(k)\theta_{NS+h-k,j-k}\right)^2\sigma_{h-j}^2\right| \\
&=\lim_{N\to\infty}\left|\sigma_{NS}^2(h)-\sum_{j=0}^{h}\left(\sum_{k=0}^{j}\chi_h(k)\theta_{NS+h-k,j-k}\right)^2\left(v_{NS+h-j}+\sigma_{h-j}^2-v_{NS+h-j}\right)\right| \\
&\le\lim_{N\to\infty}\left|\sigma_{NS}^2(h)-\sum_{j=0}^{h}\left(\sum_{k=0}^{j}\chi_h(k)\theta_{NS+h-k,j-k}\right)^2 v_{NS+h-j}\right| \\
&+\lim_{N\to\infty}\sum_{j=0}^{h}\left|\left(\sum_{k=0}^{j}\chi_h(k)\theta_{NS+h-k,j-k}\right)^2\right|\left|\left(\sigma_{h-j}^2-v_{NS+h-j}\right)\right| \\
&=\lim_{n\to\infty}\left|\sigma_n^2(h)-\sum_{j=0}^{h}\left(\sum_{k=0}^{j}\chi_h(k)\theta_{n+h-k,j-k}\right)^2 v_{n+h-j}\right|+0 \\
&=0.
\end{aligned}
$$

This proves (2.2.3). □

The following corollary gives asymptotic prediction intervals for the forecast, in the Gaussian case.

Corollary 2.2.5. *If $\{X_t\}$ is a 0-mean Gaussian process, then the probability that X_{n+h} lies between the bounds $P_{\mathcal{H}_n}X_{n+h} \pm z_{\alpha/2}\big(\sum_{j=0}^{h}\psi_h^2(j)\sigma_{h-j}^2\big)^{\frac{1}{2}}$ approaches $(1-\alpha)$ as $n \to \infty$, where z_α is the $(1-\alpha)$-quantile of the standard normal distribution.*

Proof. Since $(X_0, X_1, \ldots, X_{n+h})'$ has a multivariate normal distribution, then by Problem 2.20 in Brockwell and Davis [11],

$$P_{\mathcal{H}_n}X_{n+h} = \mathrm{E}_{\overline{\mathrm{sp}}\{X_0,\ldots,X_{n-1}\}}X_{n+h} = \mathrm{E}(X_{n+h}|X_0,\ldots,X_{n-1}).$$

Then since (2.2.3) holds, letting $\Phi(t)$ denote the cumulative distribution function of the standard normal distribution, it follows that, as $n \to \infty$,

$$\begin{aligned}
&\mathbb{P}\Big(P_{\mathcal{H}_n}X_{n+h} - z_{\alpha/2}\big(\sum_{j=0}^{h}\psi_h^2(j)\sigma_{h-j}^2\big)^{\frac{1}{2}}\\
&\leq X_{n+h} \leq P_{\mathcal{H}_n}X_{n+h} + z_{\alpha/2}\big(\sum_{j=0}^{h}\psi_h^2(j)\sigma_{h-j}^2\big)^{\frac{1}{2}}\Big)\\
&= \mathbb{P}\big(|X_{n+h} - P_{\mathcal{H}_n}X_{n+h}| \leq z_{\alpha/2}\sigma_n(h)\big)\\
&\to \Phi(z_{\alpha/2}) - \Phi(-z_{\alpha/2})\\
&= 1-\alpha.
\end{aligned}$$

□

Chapter 3

Computation

In this chapter, I outline the computations needed to produce forecasts and the associated prediction intervals. All the calculations are carried out by R software, and the codes are in the appendix. This chapter details the codes and their usage. The first step is model selection, i.e., the number of seasons S, the order p of the autoregressive part, the order q of the moving average part must be chosen. Usually S is known from the application. For example, I use $S = 4$ for quarterly data and $S = 12$ for monthly data. The next step is to estimate the autoregressive parameters $\phi_t(k)$ for $k = 1, \ldots, p$, the moving average parameters $\theta_t(j)$ for $j = 1, \ldots, q$, and the residual standard deviations σ_t of a $\text{PARMA}_S(p, q)$ model for the sample-mean corrected data, $X_t = \tilde{X}_t - \hat{\mu}_t$. These two steps are closely connected, since the process of model selection requires fitting a proposed model to judge its adequacy. The entire procedure of model selection and parameter estimation is outlined in Tesfaye, *et al.*[42]. A brief synopsis is given in the following paragraph.

For any data set $\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_{\tilde{n}}$, we can extract a subset $\tilde{X}_i, \tilde{X}_{i+1}, \ldots, \tilde{X}_{i+n-1}$, where i represents the i-th season, $n = NS$, N equals the number of years of data and S equals the

number of seasons in a year. Use (1.0.4) and (1.0.5), respectively, to compute the seasonal sample means and autocovariances. The means $\hat{\mu}_i$ for $i = 0, 1, \ldots, S-1$ are stored in an $S \times 1$ array MU(I) for I= $1, \ldots, S$ and the autocovariances $\hat{\gamma}_i(\ell)$ for $i = 0, 1, \ldots, S-1$ and $\ell = 0, \ldots, N-1$ are stored in an $S \times N$ array GAMMA(I,L) for I= $1, 2, \ldots, S$ and L= $1, \ldots, N$. Since our notation begins with season $i = 0$ and lag $\ell = 0$, and since many coding platforms (including R) do not allow zero or negative subscripts, there is a change of notation I= $i+1$ and L= $\ell+1$. In this way, MU(I) = $\hat{\mu}_i$ and GAMMA(I,L) = $\hat{\gamma}_i(\ell)$.

Now consider a general innovations algorithm for all seasons i, where $i = 0, 1, \ldots, S-1$:

$$\begin{aligned} v_{0,i} &= C(i,i) \\ \theta^{(i)}_{n,n-k} &= v^{-1}_{k,i} \left[C(i+n, i+k) - \sum_{j=0}^{k-1} \theta^{(i)}_{k,k-j} \theta^{(i)}_{n,n-j} v_{j,i} \right] \\ v_{n,i} &= C(i+n, i+n) - \sum_{j=0}^{n-1} \left(\theta^{(i)}_{n,n-j} \right)^2 v_{j,i}, \end{aligned} \tag{3.0.1}$$

solved in the order $v_{0,i}$, $\theta^{(i)}_{1,1}$, $v_{1,i}$, $\theta^{(i)}_{2,2}$, $\theta^{(i)}_{2,1}$, $v_{2,i}$, $\theta^{(i)}_{3,3}$, $\theta^{(i)}_{3,2}$, $\theta^{(i)}_{3,1}$, $v_{3,i}, \ldots$ and so forth. Now $k+1$ iterations of the innovations algorithm (3.0.1) with $C(j,\ell) = \hat{\gamma}_j(\ell - j)$ must be computed for *every* initial season $i = 0, 1, \ldots, S-1$ to obtain the estimates of the seasonal variances

$$\hat{\sigma}^2_i = v_{k,\langle i-k \rangle}, \tag{3.0.2}$$

and estimates of the coefficients in the infinite order moving average representation (1.0.2),

$$\hat{\psi}_i(\ell) = \theta^{(\langle i-k \rangle)}_{k,\ell} \tag{3.0.3}$$

for $\ell = 0, \ldots, k$, where $\langle t \rangle$ is the season corresponding to the index t, so that $\langle jS + i \rangle = i$. See Anderson, Meerschaert, and Vecchia [6] for more details.

It is necessary to repeat the innovations algorithm for every initial season $i = 0, 1, \ldots, S-1$ because the final estimates of the seasonal variances $\hat{\sigma}_i^2$ and the infinite order moving average coefficients $\hat{\psi}_i(j)$ depend on both the starting season i and the number of iterations. The number of iterations $k+1$ should be chosen so that all the parameter estimates $\hat{\sigma}_i^2$ for $i = 0, 1, \ldots, S-1$ and $\hat{\psi}_i(j)$ for $i = 0, 1, \ldots, S-1$ and $\ell = 0, \ldots, m$ show evidence of convergence, where $m = \max\{p, q\}$. I use the idea of relative error to show the convergence. For example, as the number of iterations increases, for a fixed season i, define $\hat{\sigma}_i^2(k)$ as the seasonal sample variance after k iterations, and $\hat{\sigma}_i^2(k+1)$ the seasonal variance after $k+1$ iterations. Since the value of variance can be large, I consider the relative change $\big[\hat{\sigma}_i^2(k+1) - \hat{\sigma}_i^2(k)\big]/\hat{\sigma}_i^2(k)$. As a general rule of thumb, I interpret a relative error less than 0.05 after $k+1$ iterations as evidence of convergence. The estimated seasonal variances are stored in an $S \times 1$ array VAR(I) for I= $1, \ldots, S$ and the estimated coefficients in the infinite order moving average representation are stored in an $S \times N$ array PSI(I,L) for I= $1, \ldots, S$ and L= $1, \ldots, N$. In this way, VAR(I) $= \hat{\sigma}_i^2$ and PSI(I,L) $= \hat{\psi}_i(\ell)$, where I= $i+1$ and L= $\ell + 1$.

Anderson and Meerschaert (2005) show that

$$v_{k,\langle i-k \rangle} \xrightarrow{P} \sigma_i^2, \tag{3.0.4}$$

$$\hat{\theta}_{k,j}^{(\langle i-k \rangle)} \xrightarrow{P} \psi_i(j), \tag{3.0.5}$$

and

$$N^{1/2}\big(\hat{\theta}_{k,j}^{(\langle i-k\rangle)} - \psi_i(j)\big) \Rightarrow \mathcal{N}\Big(0, \sum_{n=0}^{j-1} \frac{\sigma_{i-n}^2}{\sigma_{i-j}^2}\,\psi_i^2(n)\Big) \tag{3.0.6}$$

as $N \to \infty$ and $k \to \infty$ for any fixed $i = 0, 1, \ldots, S-1$, where " $\Rightarrow$ " indicates convergence in distribution, and $\mathcal{N}(m, v)$ is a normal random variable with mean m and variance v. The main technical condition for the convergence (3.0.6) to hold is that the noise sequence ε_t has a finite fourth moment.

In practical applications, N is the number of years of data, k is the number of iterations of the innovations algorithm (typically on the order of $k = 10$, 15 or 20, see the discussion later), and the convergence in distribution is used to approximate the quantity on the left-hand side of (3.0.6) by a normal random variable. Equation (3.0.6) can be used to produce confidence intervals and hypothesis tests for the moving average parameters in (1.0.1). For example, an α-level test statistic rejects the null hypothesis $\big(H_0 : \psi_i(\ell) = 0\big)$ in favor of the alternative $\big(H_a : \psi_i(\ell) \neq 0$, indicating that the model parameter is statistically significantly different from zero) if $|Z| > z_{\alpha/2}$. The p-value for this test is

$$\begin{gathered} p = P\left(|Z| > |z|\right) \qquad \text{where} \quad Z \sim \mathcal{N}(0,1), \\ z = \frac{N^{1/2}\hat{\theta}_{k,\ell}^{(\langle i-k\rangle)}}{W}, \qquad \text{and} \qquad W^2 = \frac{\sum_{n=0}^{\ell-1} \hat{v}_{k,\langle i-k-n\rangle}\left(\hat{\theta}_{k,n}^{(\langle i-k\rangle)}\right)^2}{\hat{v}_{k,\langle i-k-\ell\rangle}}. \end{gathered} \tag{3.0.7}$$

The innovations algorithm allows us to identify an appropriate model for the periodic time series at hand, and the p-value formula gives us a way to determine which coefficients in the identified PARMA model are statistically significantly different from zero (those with a

small p-value, say, $p < 0.05$).

Once estimates of the infinite order moving average coefficients $\hat{\psi}_i(j)$ have been computed, a system of vector difference equations must be solved to determine estimates of the autoregressive parameters $\hat{\phi}_i(j)$ for $i = 0, 1, \ldots, S-1$ and $j = 0, \ldots, p$, and estimates of the moving average parameters $\hat{\theta}_i(j)$ for $i = 0, 1, \ldots, S-1$ and $j = 0, \ldots, q$. See Tesfaye, Anderson and Meerschaert [43] for complete details. In the special case of a PARMA$_S(1,1)$ model, it is possible to solve those difference equations by hand to obtain

$$\hat{\phi}_t(1) = \hat{\psi}_t(2)/\hat{\psi}_{t-1}(1) \quad \text{and} \quad \hat{\theta}_t(1) = \hat{\psi}_t(1) - \hat{\phi}_t(1) \tag{3.0.8}$$

where $\hat{\psi}_t(0) = 1$. Hence $k+1$ iterations of the innovations algorithm for *every* initial season $i = 0, 1, \ldots, S-1$ are sufficient to estimate these parameters, assuming that k is large enough to ensure convergence for the variance estimates $\hat{\sigma}_i^2$ and the infinite order moving average coefficients $\hat{\psi}_i(j)$ for all seasons $i = 0, 1, \ldots, S-1$ and for all lags $j = 0, \ldots, m$, where $m = \max\{p, q\}$. In general, the number of iterations needed for convergence will depend on the order of the model being fit. I use an $S \times (m+1)$ array to store $\hat{\theta}_t(j)$, for a PARMA$_S(p,q)$ model, and use the same size array to store $\hat{\phi}_t(j)$. In my R code, the corresponding names of these arrays are named as THETA and PHI respectively.

Table 3 lists moving average parameter estimates $\hat{\psi}_i(\ell)$ at season i and lag $\ell = 1, 2, \ldots, 6$, and p-values, after $k = 20$ iterations of the innovations algorithm applied to average monthly flow series for the Fraser River near Hope BC. In the discussion of a PARMA$_S(1,1)$ model, by (3.0.8) I only consider $\hat{\psi}_i(\ell)$ at lag 1 and lag 2, and the ones with a higher p-value are shown in bold font in Table 3. In order to study the convergence of $\hat{\psi}_i(\ell)$ as iterations k increase, I exclude $\hat{\psi}_i(\ell)$ if its corresponding p-value is more than p_0 of 0.05, since that value

Table 3.1: Moving average parameter estimates $\hat{\psi}_i(\ell)$ at season i and lag $\ell = 1, 2, \ldots, 6$, and p-values, after $k = 20$ iterations of the innovations algorithm applied to average monthly flow series for the Fraser River near Hope BC.

i	$\hat{\psi}_i(1)$	p	$\hat{\psi}_i(2)$	p	$\hat{\psi}_i(3)$	p	$\hat{\psi}_i(4)$	p	$\hat{\psi}_i(5)$	p	$\hat{\psi}_i(6)$	p
0	0.885	.00	**0.134**	**.28**	0.105	.10	0.163	.01	0.006	.93	0.038	
1	0.625	.00	0.503	.00	0.085	.46	0.140	.02	0.077	.17	-0.004	.94
2	0.508	.00	0.350	.00	0.419	.00	0.032	.72	0.097	.03	0.019	.65
3	0.515	.00	0.287	.00	0.140	.07	0.239	.00	0.034	.60	0.030	.37
4	0.791	.00	**0.165**	**.10**	0.295	.00	0.112	.12	0.160	.03	0.045	.43
5	0.567	.00	0.757	.00	0.057	.61	0.250	.00	0.062	.40	0.139	.06
6	1.076	.01	**0.711**	**.11**	0.856	.01	0.415	.13	0.241	.17	0.112	.52
7	0.522	.03	**0.684**	**.41**	0.988	.28	1.095	.09	0.350	.51	0.198	.56
8	0.451	.00	-1.014	.00	-0.062	.66	-0.745	.50	0.128	.87	-0.635	.31
9	0.618	.00	**-0.041**	**.77**	-0.746	.01	-1.083	.26	-0.047	.97	0.514	.50
10	0.448	.00	0.409	.00	0.026	.78	-0.241	.20	-1.125	.08	0.799	.26
11	0.677	.00	0.159	.01	0.194	.00	0.050	.46	-0.190	.17	-0.402	.38

would not be significantly different from zero by (3.0.7). And consider the relative error

$$\mathrm{ERR}_i^k(\ell) = \left| \frac{\hat{\psi}_i(\ell)^{(k+1)} - \hat{\psi}_i(\ell)^{(k)}}{\hat{\psi}_i(\ell)^{(k)}} \right|. \tag{3.0.9}$$

And define the maximum of the relative error

$$R_k = \max\{\mathrm{ERR}_i^k(\ell) : (i, \ell) \notin I_k\},$$

where I_k is defined as

$$I_k = \{i, \ell : p < p_0 \quad \text{in} \quad (3.0.7)\}$$

Figure 3.1 shows plots of R_k against number of iterations k, and the convergence of $\hat{\psi}_i(\ell)$ as k increase. Figure 3.1 also shows plots excluding $\hat{\psi}_i(\ell)$ with p values above p_0 of 0.01 and 0.10. Similar plots for $v_{k,\langle i-k \rangle}$ are given in Figure 3.2.

Once the model is fit, the adequacy of the model can be judged. One way to do this is

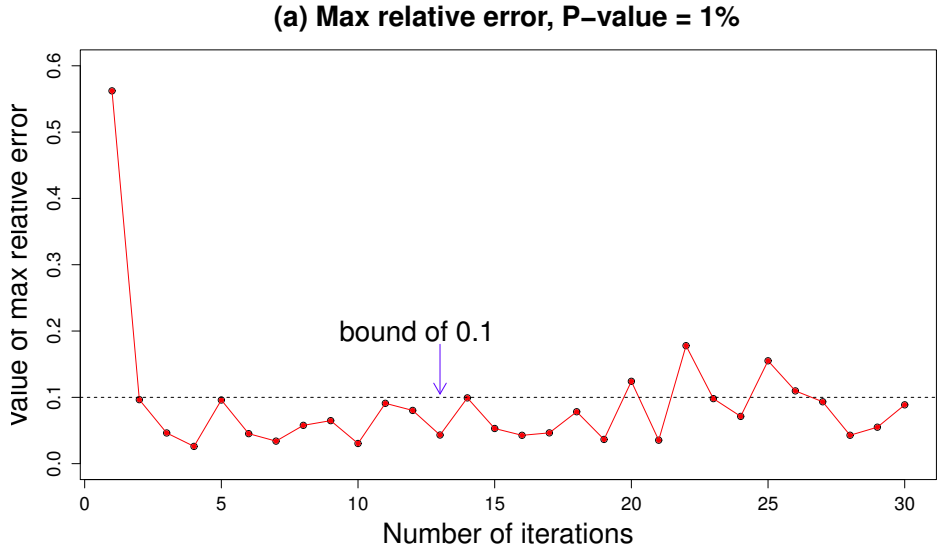


Figure 3.1: Convergence of $\hat{\psi}_i(\ell)$ as iterations k increase,where $0 \leq i < S-1, 1 \leq \ell \leq 2, k = 1, 2, \ldots, 30$. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

to compute the residuals and check for any remaining serial correlation. To compute the residuals from the data, the invertible representation (1.0.3) is used and the weights $\pi_t(j)$ must be computed by solving another system of vector difference equations. In the special case of a PARMA$_S(1,1)$ model, it is possible to solve those difference equations by hand (see Anderson, Meerschaert and Tesfaye [8]) to obtain

$$\begin{aligned}\hat{\delta}_t = \hat{\sigma}_t^{-1}\Big[X_t - \Big(\hat{\phi}_t(1) + \hat{\theta}_t(1)\Big) X_{t-1} \\ + \sum_{j=2}^{t-i} (-1)^j \Big(\hat{\phi}_{t-j+1}(1) + \hat{\theta}_{t-j+1}(1)\Big) \hat{\theta}_t(1)\hat{\theta}_{t-1}(1)...\hat{\theta}_{t-j+2}(1) X_{t-j}\Big]\end{aligned} \tag{3.0.10}$$

where i is the season of the first data point. This produces $n-1$ residuals $\hat{\delta}_t$ for $t = i+1, \ldots, i+n-1$. In general, one obtains $n-m$ residuals where $m = \max\{p, q\}$ depends on the order of the model being fit. Now plot the autocorrelation function (ACF) and the partial

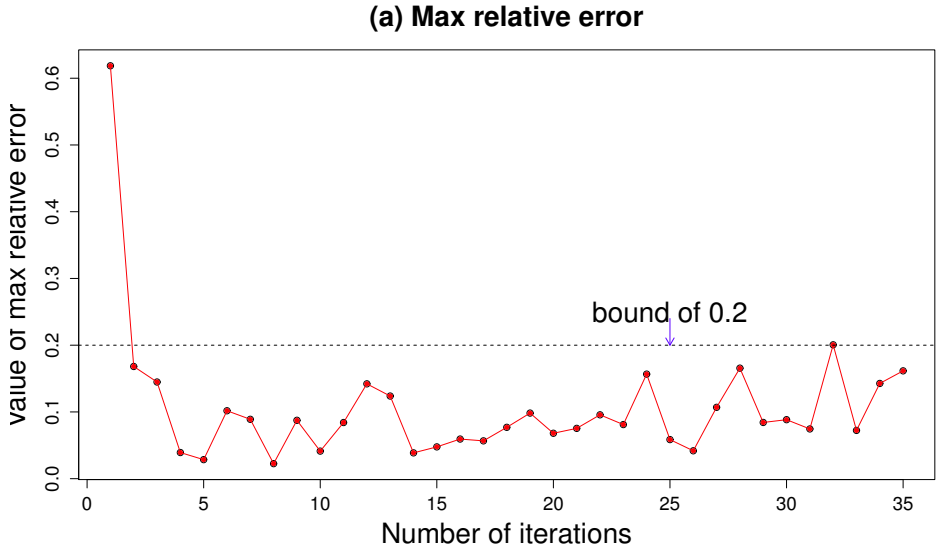


Figure 3.2: Convergence of $v_{k,\langle i-k\rangle}$ as iterations k increase,where $0 \leq i < S-1, 1 \leq \ell \leq 2, k = 1, 2, \ldots, 30$.

autocorrelation function (PACF) of the residuals and check for any remaining serial correlation, in exactly the same way as for ARMA modeling. An example could be seen in Figure 4.3 of next chapter. Since the standardized errors $\delta_t = \sigma_t^{-1}\varepsilon_t$ in a PARMA$_S(p,q)$ model are also iid observations under this model, 95% of the ACF and PACF values should fall within the confidence bands $\pm 1.96/\sqrt{n-m}$ if the model is adequate, see Tesfaye, *et al* [42]. The principle of parsimony suggests that I choose an adequate model with $m = \max(p,q)$ as small as possible. Once the order p of the autoregressive part and the order q of the moving average part are chosen and found adequate, it is then advisable to fit a reduced model with fewer parameters. One method for finding a reduced model using discrete Fourier transforms is discussed in Anderson, Meerschaert and Tesfaye [8].

Validation of a time series model is tantamount to the application of diagnostic checks to

the model residuals, to see if they resemble white noise. The Ljung-Box test can be used to test the white-noise null hypothesis (see [11]). If the null hypothesis of white-noise residuals is not rejected, and if the autocorrelation and partial autocorrelation functions of the residuals show no evidence of serial correlation, then I judge the model to be adequate. Fitting a suitable distribution to the residuals allows for a faithful simulation based on this model. To obtain additional parsimony, it is also advisable to consider simpler models where some statistically insignificant model parameters are set to zero. If the resulting model residuals pass the same diagnostic tests, then the simplified model is also deemed adequate.

Next I detail the computations required to produce forecasts, and their prediction intervals. As before, I assume that the first season to be forecast is season $i = 0$, and write the data in the form $\tilde{X}_0, \tilde{X}_1, \ldots, \tilde{X}_{n-1}$, discarding a few of the oldest observations if necessary. Then I compute the sample means $\hat{\mu}_i$ using (1.0.4) and set $X(I) = \tilde{X}_i - \hat{\mu}_i$ for $I = 1, \ldots, S$.

Step 0: Compute the sample autocovariance by (1.0.5). Apply innovations algorithm with $\hat{\gamma}_i(\ell)$, and get estimates for $\hat{\sigma}_i^2$ and $\hat{\psi}_i(\ell)$ by (3.0.2) and (3.0.3). Then model parameter estimates $\hat{\theta}_t(j)$ and $\hat{\phi}_t(k)$ could be computed by (3.0.8). From the constructed model, create an $S \times Q$ array to store $\hat{\theta}_t(j)$, $j = 0, 1, \ldots, q$ with $\hat{\theta}_t(0) = 1$ and $t = 0, 1, \ldots, S-1$, where $Q = q + 1$. Also, create an $S \times p$ array to store $\hat{\phi}_t(k)$, $k = 1, \ldots, p$ and an $S \times 1$ array to store $\hat{\sigma}_t^2$. The corresponding array names in my R codes are THETA, PHI and SIGMA respectively.

Step 1: Use Proposition 2.1.3 to compute the covariances, $C(j, \ell)$, of the transformed process W_t given in (2.1.2). Notice that in the computation process, $\gamma_j(\ell - j)$, $\hat{\theta}_t(j)$, $\hat{\phi}_t(k)$, and $\hat{\sigma}_t^2$ are obtained in Step 0. Create an $n \times n$ array C(J,L) to store $C(j, \ell)$, where C(J,L) $= C(j, \ell)$, for J $= j+1$, L $= \ell+1$. Speed and efficiency of the algorithm given in Proposition

2.1.3 is provided for by noting that $C(j,\ell) = 0$ if $\ell - j > q$ and $\ell \geq i + m$ where $m = \max(p,q)$. I may let the $n \times n$ array defined just now as a zero-array, in this way, with the previous condition, I could avoid a lot of computations.

Step 2: From the innovations algorithm given by (2.1.9), once $C(j,\ell)$ is given, calculate the coefficients $\theta_{n,j}$. Since $\theta_{n,j} = 0$ if $j > q$ and $n \geq m$, I will need an array of size $n \times Q$ (again, $\theta_{n,0} = 1$) to store the innovation coefficients, where the corresponding array name is THETA in my R code. The innovations algorithm is solved in the order v_0, $\theta_{1,1}$, v_1, $\theta_{2,2}$, $\theta_{2,1}$, v_2, $\theta_{3,3}$, $\theta_{3,2}$, $\theta_{3,1}$, $v_3, \ldots$ and so forth. Use an $n \times 1$ array to store v_k.

Step 3: Compute the one-step predictors, $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_{n-1}$, by (2.1.13), and they are stored in the array Xhat in my R code. At most only an m-step computation is needed, where $m = \max(p,q)$, since (2.1.13) only requires the storage of at most p past observations $X_{n-1}, \ldots, X_{n-p}$ and at most q past innovations $(X_{n-j} - \hat{X}_{n-j}), j = 1, \ldots, q$. Therefore computation is much faster when p and q are small, for example, PARMA(1,1) model. The one-step prediction is based on the mean-subtracted data, so I add the seasonal mean back to $\hat{X}_n$ once the prediction is obtained. See step 4 for specific storage of $\hat{X}_n$.

Step 4: The one- and h-step predictors are stored in the same vector Y of length $n + h$, where h is the desired number of forecast steps. For example, $Y[n+2]$ is the prediction at 2-step. Compute the h-step predictors $P_{\mathcal{H}_n} X_n, P_{\mathcal{H}_n} X_{n+1}, \ldots$, in order, using the recursion given by (2.1.17). The calculation of $Y[h]$ is based on the information on $Y[h-1]$, so the computation is recursive. At most only an m-step computation is needed.

Step 5: By Corollary 2.2.5, compute approximate 95% Gaussian prediction bounds for X_{n+h} given by $P_{\mathcal{H}_n} X_{n+h} \pm 1.96 \times \sigma_n(h)$. The large sample approximation of $\sigma_n(h)$ is obtained from (2.2.3). The weights, $\psi_t(j)$ in (2.2.3) are computed from the model parameters

ϕ_t and θ_t via the following recursions:

$$\psi_i(j) - \sum_{k=1}^{p} \phi_i(k)\psi_{i-k}(j-k) = 0, j \geq \max(p, q+1)$$

and

$$\psi_i(j) - \sum_{k=1}^{j} \phi_i(k)\psi_{i-k}(j-k) = 0, 0 \leq j < \max(p, q+1)$$

To avoid this computation, I may adopt the parameter estimate $\hat{\psi}_i(j)$ in (3.0.3). Finally, noting that $\tilde{X}_{n+h} - \hat{\mu}_{n+h} = X_{n+h}$, the approximate 95% prediction bounds for $\tilde{X}_{n+h}$ are $(\hat{\mu}_{n+h} + P_{\mathcal{H}_n} X_{n+h}) \pm 1.96 \times \sigma_n(h)$. Also, $\hat{\mu}_{n+h} = \hat{\mu}_{\langle n+h \rangle}$. Two h-length vectors U_BOUND and L_BOUND are defined to store the prediction bounds.

3.1 A simulation study for the convergence of the coefficients in innovations algorithm.

To better testify the convergence of $\hat{\psi}_i(\ell)$ and $\hat{v}_{k,\langle i-k \rangle}$ as iterations k increase in Figure 3.1 and Figure 3.2, we will conduct a detailed simulation to show the actual error in convergence of $\hat{\psi}_i(\ell)$ and $\hat{v}_{k,\langle i-k \rangle}$ in innovations algorithm. 72-year of monthly data for $\text{PARMA}_{12}(1,1)$ were simulated, as shown in Table 3.1, and the innovations algorithm was used to obtain parameter estimates. Some general conclusions can be drawn from this study, which in practice proves the results in (3.0.4) and (3.0.5). Define the relative error for $\hat{\psi}_i(\ell)$ as follows:

$$\text{ERR1}_i^k(\ell) = \left| \frac{\hat{\psi}_i(\ell)^{(k+1)} - \hat{\psi}_i(\ell)^{(k)}}{\hat{\psi}_i(\ell)^{(k)}} \right|. \tag{3.1.1}$$

And define the maximum of the relative error

$$R1_k = \max\{\text{ERR1}_i^k(\ell) : (i,\ell) \notin I_k\},$$

where I_k is defined as

$$I_k = \{i, \ell : p < p_0 \quad \text{in} \quad (3.0.7)\}$$

And define the relative error for $\hat{v}_{k,\langle i-k\rangle}$:

$$\text{ERR2}_i^k = \left| \frac{\hat{v}_{k+1,\langle i-(k+1)\rangle} - \hat{v}_{k,\langle i-k\rangle}}{\hat{v}_{k,\langle i-k\rangle}} \right|. \tag{3.1.2}$$

And define the maximum of the relative error

$$R2_k = \max\{\text{ERR2}_i^k : \forall i = 0, \ldots, 11\},$$

Figure 3.3 illustrates plots of $R1_k$ and $R2_k$ against iterations k, which shows the convergence of $\hat{\psi}_i(\ell)$ and $\hat{v}_{k,\langle i-k\rangle}$ as iterations k increase,where $0 \leq i \leq 11, 1 \leq \ell \leq 2, k = 1, 2, \ldots, 50$. For $\hat{\psi}_i(\ell)$, the maximum relative error over all seasons i drop below 5% when k is between 10 and 20, and it increases when k is larger than 20, since the sample autocovariance becomes relatively small when the lag is large. Similar observation could be seen for $\hat{v}_{k,\langle i-k\rangle}$, where the maximum relative error over all seasons i drop below 7% when k is between 10 and 20, and it increases when k is larger than 20. Therefore, we conclude that 10 to 20 iterations of the innovations algorithm are usually sufficient to obtain reasonable estimates of the model parameters. Furthermore, a convergence test for the model parameters $\hat{\phi}$ and $\hat{\theta}$ is shown in FIgure 3.4, and it shows similar results, which proves our previous

Table 3.2: Model parameters and estimates for simulated $\text{PARMA}_{12}(1,1)$ data.

season i	ϕ_i	$\hat{\phi}_i$	θ_i	$\hat{\theta}_i$	σ_i	$\hat{\sigma}_i$
0	0.20	0.34	0.70	0.58	11900	8766
1	0.60	0.72	0.10	0.10	11600	7626
2	0.60	0.56	-0.10	-0.16	7300	6693
3	0.60	0.45	-0.10	0.05	6000	4402
4	0.30	0.30	0.50	0.48	4200	3521
5	0.90	0.85	-0.40	-0.29	4600	4287
6	1.30	2.03	-0.20	-0.23	15200	15880
7	0.60	0.06	-0.10	0.20	31100	22339
8	-1.90	-4.85	2.40	5.43	32800	31761
9	-0.09	0.38	0.70	0.17	29700	29657
10	0.70	0.65	-0.20	-0.25	15500	13426
11	0.40	0.10	0.30	0.51	12100	11126

conclusion.

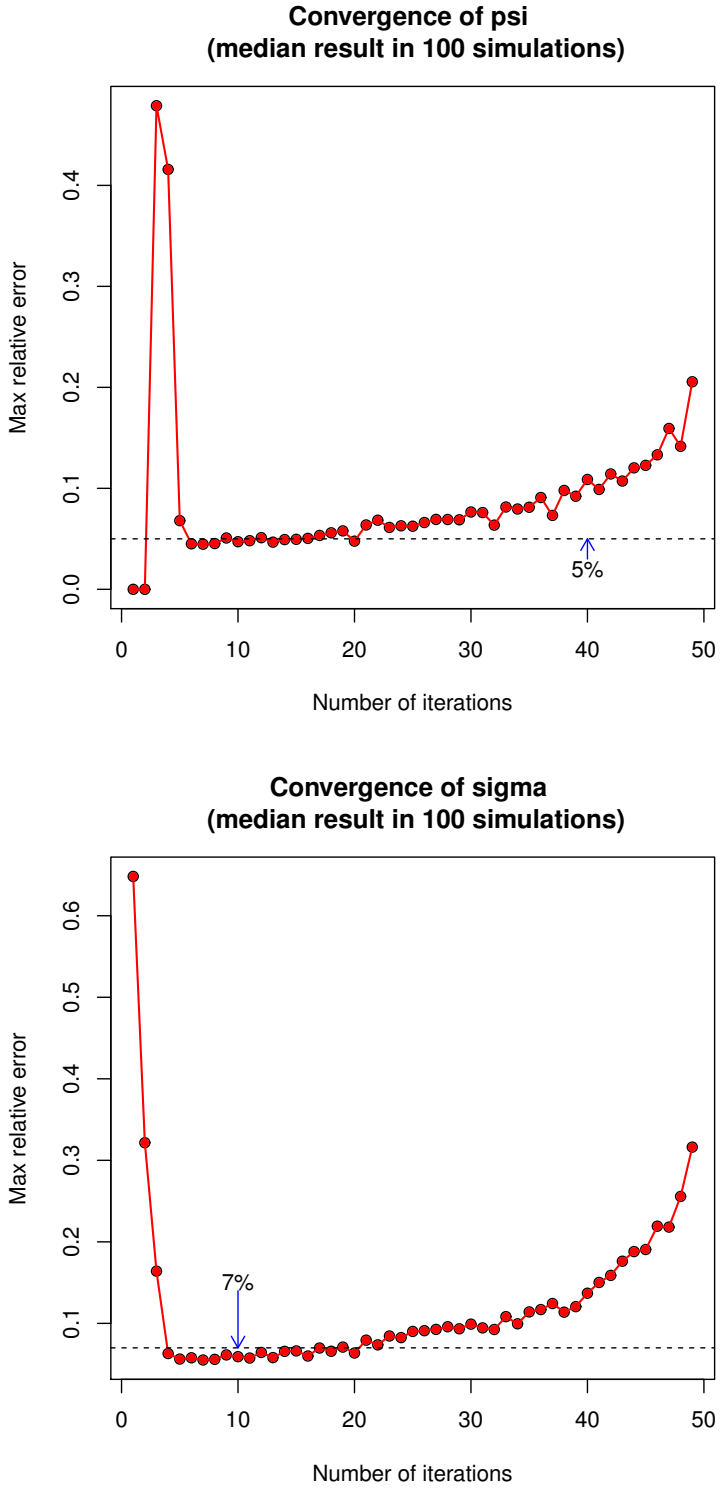
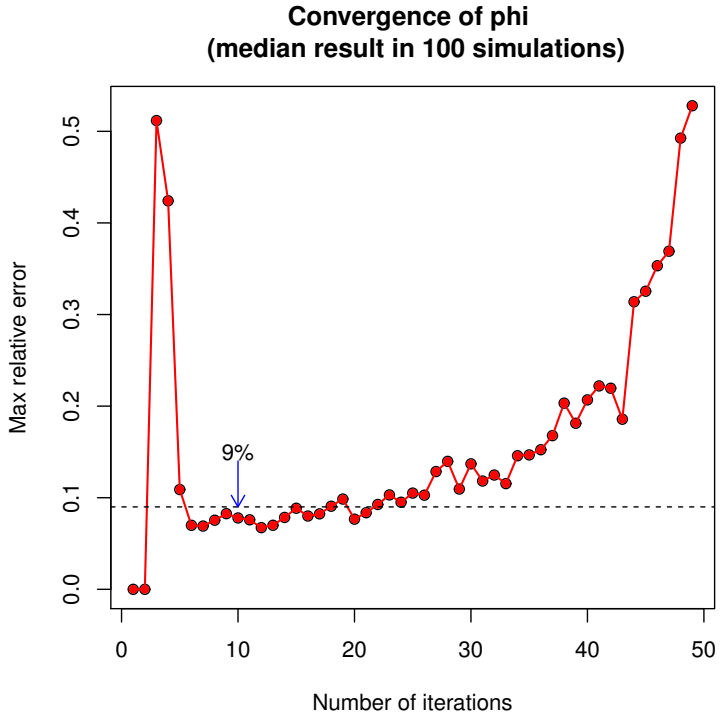


Figure 3.3: Plots of $R1_k$ and $R2_k$ against iterations k, which show the convergence of $\hat{\psi}_i(\ell)$ and $\hat{v}_{k,\langle i-k\rangle}$ as iterations k increase,where $0 \leq i \leq 11, 1 \leq \ell \leq 2, k = 1, 2, \ldots, 50$.



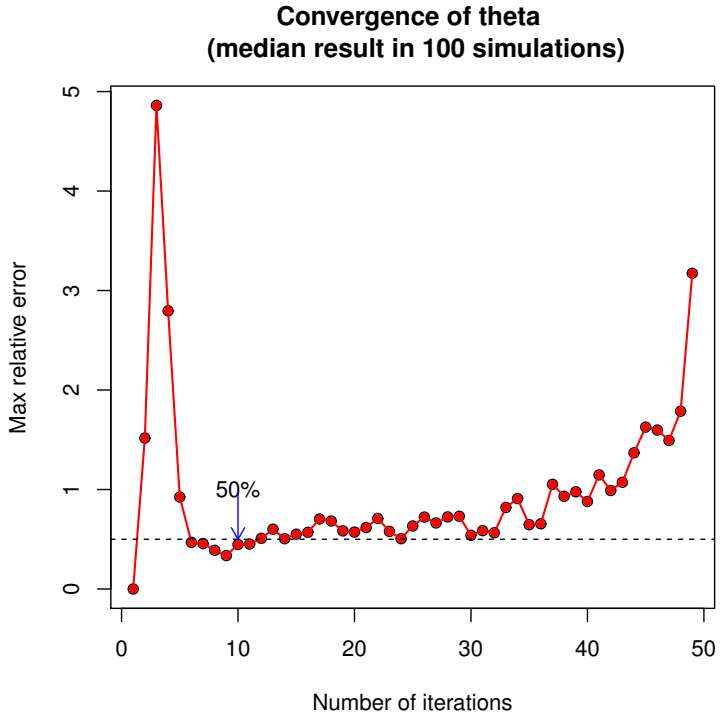


Figure 3.4: Convergence of $\hat{\phi}_i$ and $\hat{\theta}_i$ as iterations k increase,where $0 \leq i \leq 11, k = 1, 2, \ldots, 50$.

Chapter 4

Application to a Natural River Flow

In this chapter, I apply the R codes in Chapter 3 to forecast future flows for a time series of monthly river flows, get 95% Gaussian prediction bounds for these forecasts.

I model a monthly river flow time series from the Fraser River at Hope, British Columbia, which is the longest river in British Columbia, traveling almost 1400 km and sustained by a drainage area covering 220,000 square kilometers. See for maps and river flow data downloads at http://www.wateroffice.ec.gc.ca/ .

Daily discharge measurements, in cubic meters per second (cms), were averaged over each of the respective months to produce monthly Fraser River flow time series. The series contains 72 years of data from October 1912 to September 1984, and part of the data is shown in Figure 4.1. To better test the prediction results, I based our forecast on the first 70 years of data, from 1912 to 1982. Then, I computed a 24-month prediction from October 1982 to September 1984. In our analysis, $i = 0$ corresponds to October and $i = 11$ corresponds to September. Using the "water year" starting on 1 October is customary for modeling of river flows, because of low correlation between Fall monthly flows. The sample seasonal mean,

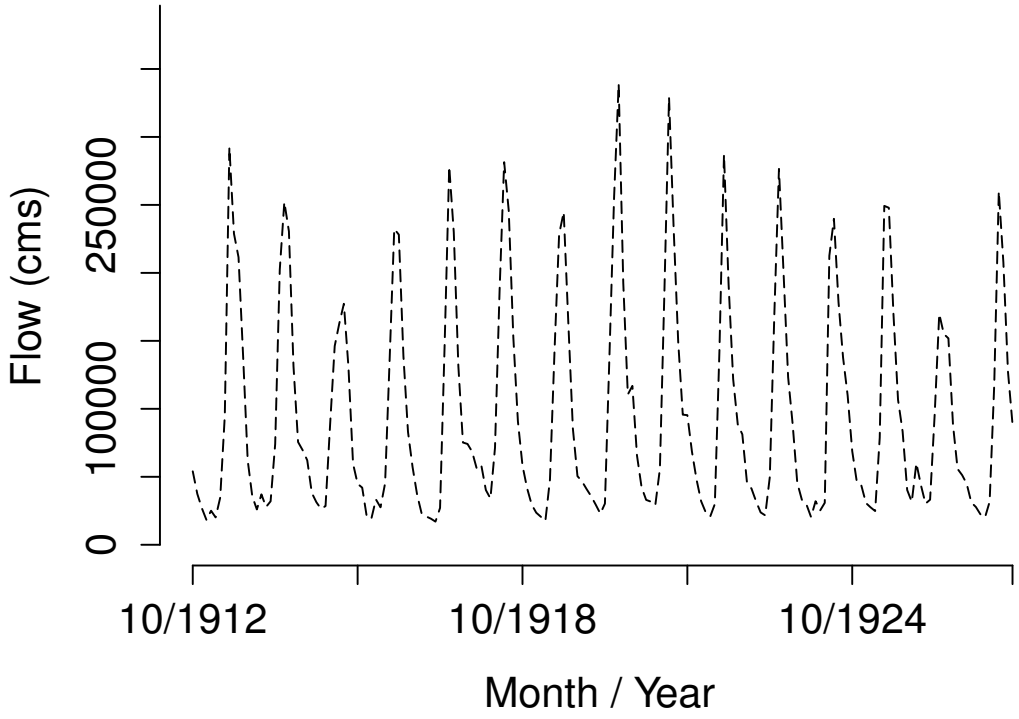


Figure 4.1: Average monthly flows in cubic meters per second (cms) for the Fraser River at Hope, BC indicate a seasonal pattern.

standard deviation and autocorrelations at lag 1 and lag 2 are given in Table 4 and Figure 4.2, with 95% confidence intervals. The non-stationary of the series is apparent, since the mean, standard deviation and correlation functions vary significantly from month to month (the confidence bands for some wet and dry seasons don't overlap). Removing the periodicity in mean and variance will not yield a stationary series. Therefore a periodically stationary series model is appropriate. Tesfaye, *et al.* [42], identified a $\text{PARMA}_{12}(1,1)$ model

$$X_t - \phi_i X_{t-1} = \varepsilon_t + \theta_i \varepsilon_{t-1}, \tag{4.0.1}$$

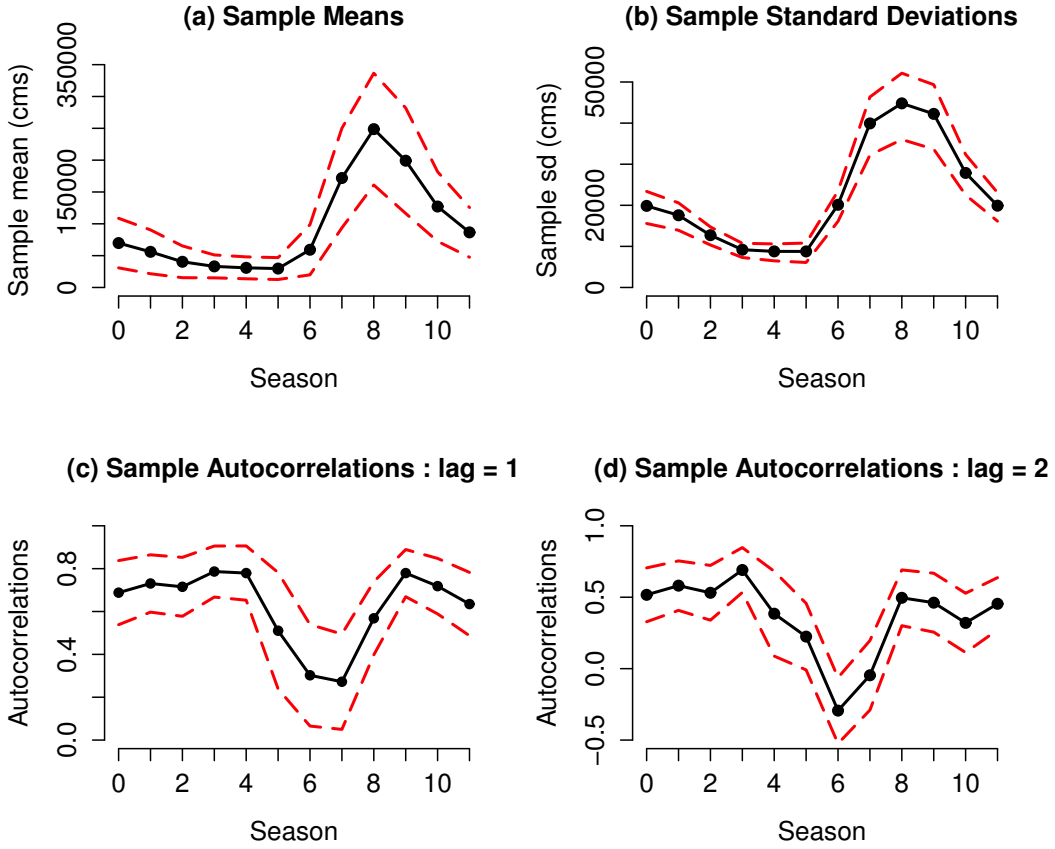


Figure 4.2: Statistics for the Fraser river time series: (a) seasonal mean; (b) standard deviation; (c,d) autocorrelations at lags 1 and 2. Dotted lines are 95% confidence intervals. Season = 0 corresponds to October and Season = 11 corresponds to September.

where $X_t = \tilde{X}_t - \mu_t$, $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{V}(\varepsilon_t) = \sigma_t^2$, $\sigma_t^{-1}\varepsilon_t$ iid, for the series with $S = 12$, and used the innovations algorithm at $k = 20$ iterations for periodically stationary processes by (3.0.1) and (3.0.8) to estimate $\phi_i(1)$, $\theta_i(1)$, and σ_i, $i = 0, 1, \ldots, 11$.

Table 4 gives the parameter estimates of the model where $\hat{\phi} = (\phi_0(1), \ldots, \phi_{11}(1))'$, $\hat{\theta} = (\theta_0(1), \ldots, \theta_{11}(1))'$, and $\hat{\sigma} = (\sigma_0, \ldots, \sigma_{11})'$. I employ these estimates as the parameters of the model. Although the model is periodically stationary, the standardized residuals (3.0.10) should be stationary, so the standard 95% limits (that is, $1.96/\sqrt{n}$) still apply. Figure 4.3 shows the ACF and PACF of the model residuals. Although a few values lie

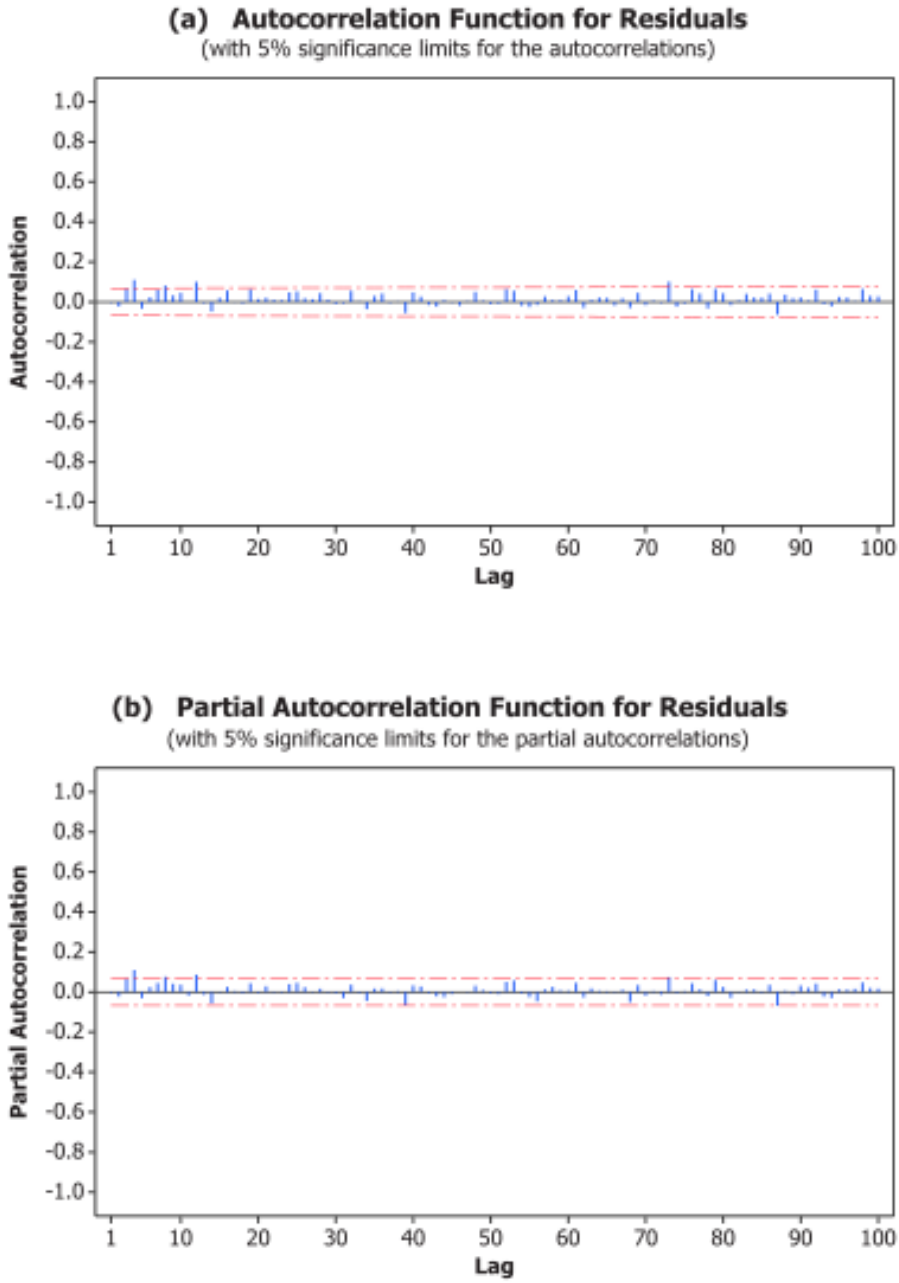


Figure 4.3: ACF and PACF for model residuals, showing the bounds $\pm 1.96/\sqrt{N}$, indicate no serial dependence. With no apparent pattern, these plots indicate that the $\text{PARMA}_{12}(1,1)$ model is adequate.

Table 4.1: Parameter estimates for the PARMA model (1.0.1) of average monthly flow series for the Fraser River near Hope BC from October 1912 to September 1982.

month	$\hat{\phi}$	$\hat{\theta}$	$\hat{\sigma}$
OCT	0.187	0.704	11761.042
NOV	0.592	0.050	11468.539
DEC	0.575	-0.038	7104.342
JAN	0.519	-0.041	5879.327
FEB	0.337	0.469	4170.111
MAR	0.931	-0.388	4469.202
APR	1.286	-0.088	15414.905
MAY	1.059	-0.592	30017.508
JUN	-2.245	2.661	32955.491
JUL	-1.105	0.730	30069.997
AUG	0.679	-0.236	15511.989
SEP	0.353	0.326	12111.919

Table 4.2: Sample mean, standard deviation and autocorrelation at lag 1 and 2 for an average monthly flow series for the Fraser River near Hope BC, from October 1912 to September 1982.

	Parameter			
month	$\hat{\mu}$	$\hat{\gamma}(0)^{\frac{1}{2}}$	$\hat{\rho}(1)$	$\hat{\rho}(2)$
OCT	69850	19976	0.712	0.515
NOV	55824	17709	0.748	0.577
DEC	40502	12858	0.731	0.541
JAN	33006	9269	0.786	0.697
FEB	30740	8878	0.787	0.380
MAR	29348	8864	0.504	0.286
APR	58959	20314	0.333	-0.286
MAY	173308	39437	0.260	-0.031
JUN	249564	45154	0.577	0.499
JUL	198844	42627	0.780	0.456
AUG	127138	28253	0.720	0.308
SEP	86437	20071	0.621	0.472

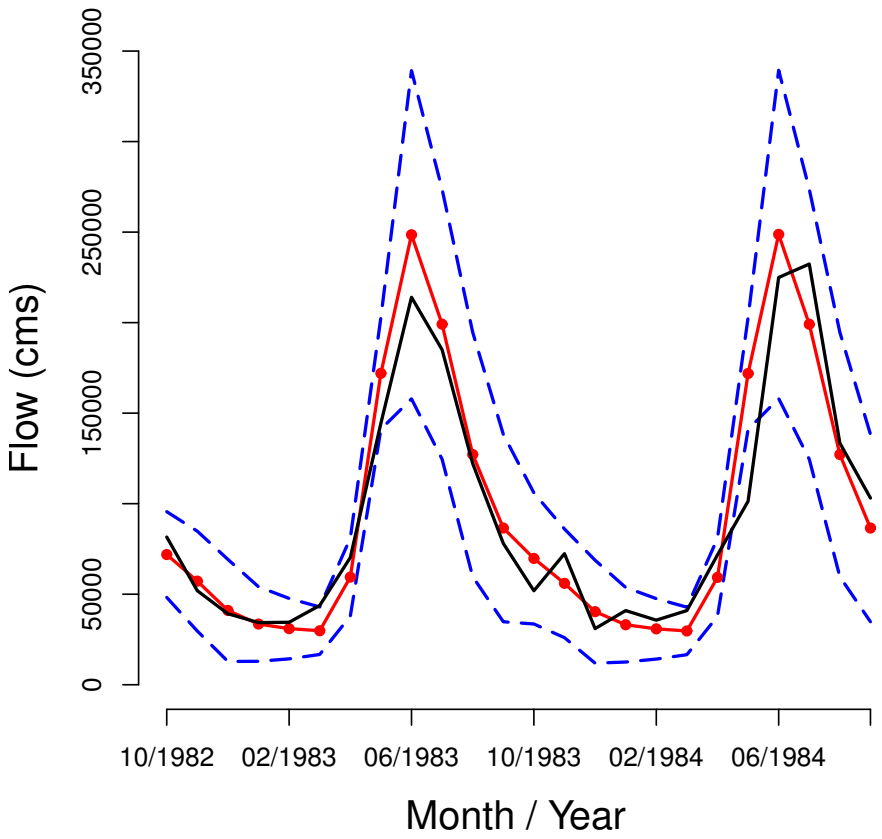


Figure 4.4: 24-month forecast (solid line with dots) based on 70 years of Fraser river data, with 95% prediction bounds (dotted lines). For comparison, the actual data (solid line) is also shown. This data was not used in the forecast.

slightly outside of the 95% confidence bands, there is no apparent pattern. The p value from the Ljung-Box test was 0.08 indicating that I do not reject the null hypothesis that the residuals resemble iid white noise. Hence we conclude that the $\text{PARMA}_{12}(1,1)$ model is adequate.

If I reject the null hypothesis that PARMA$(1,1)$ model residuals resemble iid noise, I would abandon that model and fit a PMA(q) model to the data, $q \geq 2$. Using Theorem 1 in [8], I would identify the order, q, of the pure moving average and then parsimoniously estimate the moving average parameters that I deem to be nonzero. However, from Tesfaye,

et al. [42], the PARMA$(1,1)$ model is generally adequate to model most river flow time series.

I then compute a 24-step future prediction for the Fraser river data, that is, a forecast for the next 24 months from October 1982 to September 1984.. The prediction is compared to the original data in Figure 4.4. Note that the forecast curve is close to the original data curve, and that the historical Fraser River data stay well within the 95% prediction bands.

Figure 4.5 illustrates how the width of the prediction intervals vary with the season. This is a consequence of the non-stationarity of the river flow series, and specifically the fact that the standard deviation and the correlation functions vary significantly from month to month.

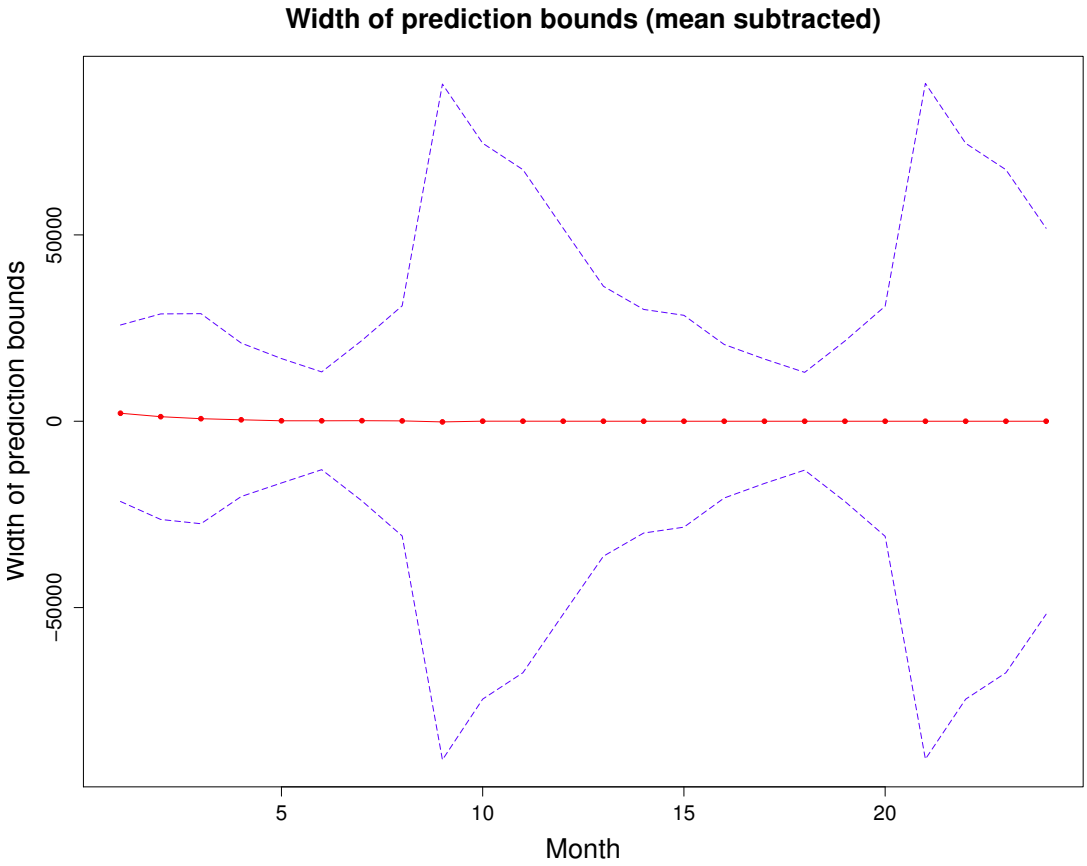


Figure 4.5: Width of 95% prediction bounds for the Fraser river.

Chapter 5

Maximum likelihood estimation and reduced model

5.1 Maximum likelihood Function for PARMA$_S(p,q)$ Model

By Yule-Walker equations, the vector of autoregressive coefficients $\phi_n^{(i)} = (\phi_{n,1}^{(i)}, \ldots, \phi_{n,n}^{(i)})'$ solves the prediction equations

$$\Gamma_{n,i}\phi_n^{(i)} = \gamma_n^{(i)} \tag{5.1.1}$$

with $\gamma_n^{(i)} = (\gamma_{i+n-1}(1), \gamma_{i+n-2}(2), \ldots, \gamma_i(n))'$ and

$$\Gamma_{n,i} = \left[\gamma_{i+n-\ell}(\ell - m)\right]_{\ell,m=1,\ldots,n} \tag{5.1.2}$$

is the covariance matrix of $\mathbf{X}_{n,i} = (X_i, ..., X_{i+n-1})'$ for each $i = 0, ..., S-1$. If $\sigma_i^2 > 0$ for $i = 0, \ldots, S-1$, then for a causal PARMA$_S(p,q)$ process the covariance matrix $\Gamma_{n,i}$ is

nonsingular for every $n \geq 1$ and each i in Lund and Basawa [22, Theorem 3.1]. Assuming that $\mathbf{X}_{n,i}$ is a Gaussian process, its likelihood function can be written using the general formula for a multivariate normal random vector with a given mean and covariance matrix

$$L(\Gamma_{n,i}) = (2\pi)^{-n/2}(\det \Gamma_{n,i})^{-1/2} \exp\left(-\frac{1}{2}\mathbf{X}'_{n,i}\Gamma^{-1}_{n,i}\mathbf{X}_{n,i}\right), \tag{5.1.3}$$

where $\Gamma_{n,i}$ is the covariance matrix of $\mathbf{X}_{n,i}$ defined in (5.1.2).

Whenever $\Gamma_{n,i}$ is invertible, the direct calculation of $\det \Gamma_{n,i}$ and $\Gamma^{-1}_{n,i}$ can be avoided by expressing it in terms of the one-step predictors $\hat{X}_{i+j}$ and the mean square errors $v_{j,i}$, both of which are easily calculated recursively from the innovations algorithm, where $v_{j,i}$ is from (3.0.1), and

$$X_{i+j} - \hat{X}^{(i)}_{i+j} \sim \mathcal{N}(0, v_{j,i}), \quad j = 0, 1, \ldots, n-1,$$

Recall the innovations representation for PARMA models in Anderson et al [6], where

$$\hat{X}^{(i)}_{i+t} = \begin{cases} 0, & t = 0 \\ \displaystyle\sum_{j=1}^{t} \theta^{(i)}_{t,j}(X_{i+t-j} - \hat{X}^{(i)}_{i+t-j}), & t = 1, \ldots, n-1. \end{cases} \tag{5.1.4}$$

Then the computation for (5.1.3) can be simplified in the following theorem:

Theorem 5.1.1. *When $\Gamma_{n,i}$ is invertible, the likelihood* (5.1.3) *can be equivalently written*

in the form of

$$L(\Gamma_{n,i}) = (2\pi)^{-n/2}(v_{0,i}v_{1,i}\cdots v_{n-1,i})^{-1/2} \exp\Big(-\frac{1}{2}\sum_{j=0}^{n-1}(X_{i+j}-\hat{X}_{i+j}^{(i)})^2/v_{j,i}\Big), \tag{5.1.5}$$

where $\hat{X}_{i+j}$ *are the one-step predictors in* (2.1.13) *and* $v_{j,i}$ *are the mean square errors in* (3.0.1).

Proof. For each i, define the $n \times n$ lower triangular matrix

$$\mathbf{C}_i = [\theta_{k,k-j}(i)]_{k,j=0}^{n-1}, \tag{5.1.6}$$

where $\det \mathbf{C}_i = 1$, since all the diagonal elements $\theta_{k,0}(i) = 1$ for $k = 0, \ldots, n-1$, and define the $n \times n$ diagonal matrix,

$$\mathbf{D}_i = \mathrm{diag}(v_{0,i}, v_{1,i}, \ldots, v_{n-1,i}). \tag{5.1.7}$$

Therefore $\hat{\mathbf{X}}_{n,i}^{(i)} = (\hat{X}_i^{(i)}, \ldots, \hat{X}_{i+n-1}^{(i)})'$ can be written in the from,

$$\hat{\mathbf{X}}_{n,i} = (\mathbf{C}_i - \mathbf{I})(\mathbf{X}_{n,i} - \hat{\mathbf{X}}_{n,i}^{(i)}), \tag{5.1.8}$$

where $\mathbf{I}$ is the $n \times n$ identity matrix. Hence,

$$\mathbf{X}_{n,i} = \mathbf{X}_{n,i} - \hat{\mathbf{X}}_{n,i} + \hat{\mathbf{X}}_{n,i} = \mathbf{C}_i(\mathbf{X}_{n,i} - \hat{\mathbf{X}}_{n,i}). \tag{5.1.9}$$

Since $\mathbf{D}_i = \mathrm{E}\left[\left(\mathbf{X}_{n,i} - \hat{\mathbf{X}}_{n,i}\right)\left(\mathbf{X}_{n,i} - \hat{\mathbf{X}}_{n,i}\right)'\right]$, and $\Gamma_{n,i} = \mathrm{E}\left(\mathbf{X}_{n,i}\mathbf{X}'_{n,i}\right)$, it follows that

$$\Gamma_{n,i} = \mathbf{C}_i \mathbf{D}_i \mathbf{C}_i{}'. \tag{5.1.10}$$

By (5.1.9) and (5.1.10), we obtain

$$\mathbf{X}'_{n,i}\Gamma_{n,i}^{-1}\mathbf{X}_{n,i} = (\mathbf{X}_{n,i} - \hat{\mathbf{X}}_{n,i})'\mathbf{D}_i{}^{-1}(\mathbf{X}_{n,i} - \hat{\mathbf{X}}_{n,i}) = \sum_{j=0}^{n-1}(X_{i+j} - \hat{X}_{i+j}^{(i)})^2/v_{j,i}, \tag{5.1.11}$$

and

$$\det\Gamma_{n,i} = (\det\mathbf{C}_i)^2(\det\mathbf{D}_i) = v_{0,i}v_{1,i}\cdots v_{n-1,i}. \tag{5.1.12}$$

Therefore the likelihood (5.1.3) of the vector $\mathbf{X}_{n,i}$ reduces to (5.1.5). Whenever $\Gamma_{n,i}$ is invertible, the likelihood in (5.1.3) has the equivalent innovations representation as (5.1.5). Actually $\Gamma_{n,i}$ is invertible for causal PARMA models, see the proof in Lund and Basawa [23, Proposition 4.1]. Given the seasonal one-step error variances $v_{j,i}$ from the innovations and the one-step ahead predictor $\hat{X}_{i+j}^{(i)}$, then $L(\Gamma_{n,i})$ in (5.1.3) can easily be calculated. □

For notation, let $\boldsymbol{\phi}_t = (\phi_t(1), \ldots, \phi_t(p))'$ and $\boldsymbol{\theta}_t = (\theta_t(1), \ldots, \theta_t(q))'$ denote the autoregressive and moving-average parameters during season t, respectively. Then we can write $L(\Gamma_{n,i}) = L(\boldsymbol{\beta})$, where we use $\boldsymbol{\beta} = (\boldsymbol{\phi}'_0, \boldsymbol{\theta}'_0, \boldsymbol{\phi}'_1, \boldsymbol{\theta}'_1, \ldots, \boldsymbol{\phi}'_{S-1}, \boldsymbol{\theta}'_{S-1})'$ to denote the collection of all PARMA$_S(p,q)$ parameters. The dimension of $\boldsymbol{\beta}$ is $(p+q)S \times 1$. Following the ideas from Basawa and Lund [24], we treat the white noise variances $\boldsymbol{\sigma}^{\mathbf{2}} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{S-1}^2)'$

as nuisance parameters. Then the likelihood function is given by

$$L_{\boldsymbol{X}}(\boldsymbol{\beta}) = (2\pi)^{-n/2}(v_{0,i}v_{1,i}\cdots v_{n-1,i})^{-1/2}\exp\Big(-\frac{1}{2}\sum_{j=0}^{n-1}(X_{i+j}-\hat{X}_{i+j}^{(i)})^2/v_{j,i}\Big). \tag{5.1.13}$$

Another way to compute the PARMA likelihood parameter estimates is to equivalently minimize the negative log likelihood

$$-2\log\{L(\boldsymbol{\beta})\} = n\log(2\pi) + \sum_{j=0}^{n-1}\log(v_{j,i}) + \sum_{j=0}^{n-1}(X_{i+j}-\hat{X}_{i+j}^{(i)})^2/v_{j,i}. \tag{5.1.14}$$

According to Basawa and Lund [24], once we obtain the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$, the MLE of σ_i^2 for $0 \le i \le S-1$, have the large sample form

$$\hat{\sigma}_i^2 = N^{-1}\sum_{j=0}^{N-1}\varepsilon_{jS+i}(\hat{\beta})^2,$$

where $\hat{\boldsymbol{\sigma}}^{\mathbf{2}} = (\hat{\sigma}_0^2,\ldots,\hat{\sigma}_{S-1}^2)'$. However, they didn't give a proof for the large sample form. In the following we propose our method of MLE computation, where there is an exact form of MLE for $\boldsymbol{\sigma}^{\mathbf{2}}$, and the computation is much faster. Given the data set $X_i,\ldots,X_{i+n-1}$, apply innovations algorithm to compute the 1-step ahead predictors $\hat{X}_{i+j}^{(i)}$, $j=0,\ldots,n-1$ and the forecast errors $v_{j,i} = \mathrm{E}\left[\left(X_{i+j}-\hat{X}_{i+j}^{(i)}\right)^2\right]$.

Corollary 5.1.2. *$|r_{j,i}-1| \to 0$ as $j \to \infty$.*

Proof. By Anderson et al. [6, Corollary 2.2.1]

$$v_{k,\langle i-k\rangle} \to \sigma_i^2,$$

we have

$$|v_{j,i} - \sigma^2_{i+j}| \to 0, \text{as } j \to \infty.$$

Recall that $v_{j,i} = \sigma^2_{i+j} r_{j,i}$, then we have

$$|\sigma^2_{i+j} r_{j,i} - \sigma^2_{i+j}| \to 0, \text{as } j \to \infty,$$

so that

$$\sigma^2_{i+j} |r_{j,i} - 1| \to 0, \text{as } j \to \infty.$$

Notice that for PARMA$_S(p,q)$ model σ^2_{i+j} is periodic of S seasons, so

$$0 < \sigma^2_{i+j} \leq \max\{\sigma^2_0, \ldots, \sigma^2_{\mathrm{S}-1}\} < \infty,$$

then we have proved

$$|r_{j,i} - 1| \to 0, \text{as } j \to \infty.$$

□

Then the negative log likelihood (5.1.14) becomes

$$\begin{aligned}\ell(\boldsymbol{\beta}) = -2\log\{L(\boldsymbol{\beta})\} &= n\log(2\pi) + \sum_{j=0}^{n-1}\log(v_{j,i}) + \sum_{j=0}^{n-1}\left(X_{i+j} - \hat{X}_{i+j}^{(i)}\right)^2 / v_{j,i} \\ &= n\log(2\pi) + \sum_{i=0}^{S-1}\sum_{k=0}^{N-1}\log\left(r_{kS+i}\sigma_i^2\right) \\ &+ \sum_{i=0}^{S-1}\sum_{k=0}^{N-1}\left(X_{kS+i} - \hat{X}_{kS+i}^{(i)}\right)^2 / \left(r_{kS+i}\sigma_i^2\right) \\ &= n\log(2\pi) + \sum_{j=0}^{N-1}\log(r_j) + N\log(\sigma_i^2) + \sum_{i=0}^{S-1} S_i/\sigma_i^2,\end{aligned} \tag{5.1.15}$$

where

$$S_i = \sum_{k=0}^{N-1}\frac{(X_{kS+i} - \hat{X}_{kS+i}^{(i)})^2}{r_{kS+i}}.$$

Then for $i = 0, \ldots, S-1$, we have

$$\frac{\partial \ell}{\partial \sigma_i^2} = \frac{N}{\sigma_i^2} + \frac{-1}{\left(\sigma_i^2\right)^2} S_i = 0,$$

therefore

$$N\sigma_i^2 - Si = 0,$$

and

$$\sigma_i^2 = \frac{S_i}{N}.$$

So the MLE of σ_i^2 is

$$\hat{\sigma}_i^2 = \frac{1}{N}\sum_{k=0}^{N-1}\left(X_{kS+i} - \hat{X}_{kS+i}^{(i)}\right)^2 / r_{kS+i} = S_i/N, \tag{5.1.16}$$

where $\hat{X}_{kS+i}^{(i)}$ and r_{kS+i} come from the innovations algorithm applied to the model. Now we can substitute $\hat{\sigma}_i^2$ for σ_i^2 in the negative log likelihood to get

$$-2\log\{L(\boldsymbol{\beta})\} = n\log(2\pi) + n + \sum_{j=0}^{n-1}\log(r_{j,i}) + N\sum_{i=0}^{S-1}\log\left(S_i/N\right).$$

Then it suffices to minimize

$$\ell^*(\boldsymbol{\beta}) = \frac{1}{N}\sum_{j=0}^{n-1}\log(r_{j,i}) + \sum_{i=0}^{S-1}\log\left(S_i/N\right).$$

Also, since $r_{j,i} \to 1$ as $j \to \infty$, then $\log(r_j) \to 0$, so

$$\frac{1}{N}\sum_{j=0}^{n-1}\log(r_j) \approx 0,$$

and approximate MLE minimizes $\sum_{i=0}^{S-1}\log\left(S_i/N\right)$.

Example 5.1.3. A PARMA$(1,1)$ example from Chapter 4 will be shown here, to demonstrate how to explicitly solve the mean square errors, and obtain the likelihood value and maximum likelihood estimate. We adopt the 70 years of monthly Fraser river data from October 1912 to September 1982, and since there are 840 observations, we denote them as $\{X_t\} = \{X_0, X_1, \ldots, X_{839}\}$. Therefore the negative log likelihood is simplified as

$$\begin{aligned} -2\log\{L(\boldsymbol{\beta},\boldsymbol{\sigma}^2)\} &= n\log(2\pi) + \sum_{j=0}^{n-1}\log(v_j) + \sum_{j=0}^{n-1}(X_j - \hat{X}_j)^2/v_j \\ &= n\log(2\pi) + \sum_{j=0}^{N-1}\log(r_j) + N\log(\sigma_i^2) + \sum_{i=0}^{S-1}S_i/\sigma_i^2, \end{aligned} \tag{5.1.17}$$

where $\boldsymbol{\beta} = (\phi_0, \theta_0, \phi_1, \theta_1, \ldots, \phi_{11}, \theta_{11})'$ and $\boldsymbol{\sigma}^2 = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{11}^2)'$. Computations in Lund and Basawa [23, Equations (2.7), (2.11)] show that the PARMA$(1,1)$ model is causal when $|\phi_0\phi_1 \cdots \phi_{11}| < 1$ and invertible when $|\theta_0\theta_1 \cdots \theta_{11}| < 1$. By Theorem 2.1.7, we may write

$$\begin{aligned} \hat{X}_0 &= 0 \\ \hat{X}_1 &= \theta_{1,1}\left(X_0 - \hat{X}_0\right) \\ \hat{X}_t &= \phi_t X_{t-1} + \theta_{t,1}\left(X_{t-1} - \hat{X}_{t-1}\right) \quad t = 2, \ldots, 839. \end{aligned} \tag{5.1.18}$$

For the covariance structure of $\{X_t\}$, we will use the $\gamma_t(h)$ results from Section 1.1.

The innovations algorithm explicitly identifies the prediction coefficients as

$$\theta_{t,1} = v_{t-1}^{-1} C\left(t, t-1\right) = v_{t-1}^{-1} \theta_t \sigma_{t-1}^2. \tag{5.1.19}$$

Use (5.1.19) in innovation algorithm gives

$$v_t = C(t,t) - \theta_{t,1}^2 v_{t-1} = (\sigma_t^2 + \theta_t^2 \sigma_{t-1}^2) - v_{t-1}^{-1} \theta_t^2 \sigma_{t-1}^4, \tag{5.1.20}$$

where $t \geq 1$, and the initial condition is $v_0 = \gamma_0(0)$. Equation (5.1.20) is a difference equation for v_t, which can be explicitly solved as follows. First, write

$$\begin{aligned} v_t &= \sigma_t^2 \frac{y_t}{y_t - 1}, \\ v_{t-1} &= \sigma_{t-1}^2 \frac{y_{t-1}}{y_{t-1} - 1}, \end{aligned} \tag{5.1.21}$$

and substitute (5.1.21) in (5.1.20) to get

$$\sigma_t^2 \frac{y_t}{y_t - 1} = (\sigma_t^2 + \theta_t^2 \sigma_{t-1}^2) - \frac{\theta_t^2 \sigma_{t-1}^4}{\sigma_{t-1}^2} \frac{(y_{t-1} - 1)}{y_{t-1}}$$

$$\sigma_t^2 y_t y_{t-1} = (\sigma_t^2 + \theta_t^2 \sigma_{t-1}^2) y_{t-1}(y_t - 1) - \theta_t^2 \sigma_{t-1}^2 (y_{t-1} - 1)(y_t - 1),$$

therefore

$$y_t = 1 + \frac{\sigma_t^2}{\sigma_{t-1} \theta_t^2} y_{t-1} \quad t \geq 1, \tag{5.1.22}$$

where the initial condition of (5.1.22) is $y_0 = \frac{\sigma_0^{-2} \gamma_0(0)}{\sigma_0^{-2} \gamma_0(0) - 1}$. Letting

$$P(t) = \frac{\sigma_t^2}{\sigma_{t-1}^2 \theta_t^2}, \tag{5.1.23}$$

then the solution to (5.1.22) is

$$\begin{aligned} y_t &= 1 + P(t) y_{t-1} \\ &= 1 + P(t)\Big(1 + P(t-1)\Big) y_{t-2} \\ &= 1 + P(t)\Big(1 + P(t-1)\Big)\Big(1 + P(t-2)\Big) y_{t-3} \\ &\vdots \\ &= 1 + \left(\prod_{r=1}^{t} P(r) \right) y_0 + \sum_{j=2}^{t} \left(\prod_{r=j}^{t} P(r) \right), \quad t \geq 1. \end{aligned} \tag{5.1.24}$$

In the computation, notice that $\sum_{j=2}^{t} \left(\prod_{r=j}^{t} P(r) \right)$ is the row-summation of the product

of each row in a matrix, where the matrix is

$$\begin{pmatrix} 1 & P(2) & P(3) & \dots & \dots & P(t) \\ 1 & 1 & P(3) & \dots & \dots & P(t) \\ 1 & 1 & 1 & P(4) & \dots & P(t) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 1 & 1 & \dots & 1 & P(t) \end{pmatrix}.$$

Use (5.1.24) in (5.1.21) to compute v_t. Then use v_t in (5.1.19) to compute $\theta_{t,1}$. Next obtain $\hat{X}_t$ in (5.1.18). Finally substitute into (5.1.17) to get the negative log likelihood function.

In Chapter 4, we fit a PARMA(1,1) model to the 70-year Fraser river flows data. The model parameters $\hat{\theta}_t$, $\hat{\phi}_t$ and $\hat{\sigma}_t$ are shown in Table 4 of Chapter 4. Using these values to compute v_t and $\hat{X}_t$ in this example, we obtain the negative log likelihood value of 18543 for this model. However, in the following we consider to take the logarithm of the data, for all computations in MLE. The model parameters for the logarithm of the data are shown in Table 5.1.3, where the negative log likelihood value is -362.655. To get the maximum

Season	0	1	2	3	4	5
$\hat{\theta}_i$	0.650	0.206	-0.026	-0.033	0.426	-0.400
$\hat{\phi}_i$	0.393	0.740	0.712	0.715	0.431	1.020
σ_i	0.165	0.187	0.164	0.159	0.135	0.134
Season	6	7	8	9	10	11
$\hat{\theta}_i$	0.446	-0.171	1.918	0.836	-0.299	0.540
$\hat{\phi}_i$	0.425	0.313	-1.628	-0.045	0.979	0.414
σ_i	0.260	0.184	0.127	0.143	0.108	0.138

Table 5.1: Estimated parameters for $\text{PARMA}_{12}(1,1)$ model, from innovations algorithm, using the logarithm of the original data. The resulting negative likelihood value is -362.655.

likelihood estimators, we will differentiate $-2\log L(\boldsymbol{\beta})$ partially with respect to the parame-

ters. A non-linear optimization algorithm will be applied, using R software to find the MLE. By our earlier assumption, we treat the white noise variances $\boldsymbol{\sigma^2} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{S-1}^2)'$ as nuisance parameters, so that the independent variables in the optimization are the model parameters $\boldsymbol{\beta} = (\phi_0, \theta_0, \phi_1, \theta_1, \ldots, \phi_{S-1}, \theta_{S-1})'$. After we compute the MLE $\hat{\boldsymbol{\beta}}$, we solve the MLE for $\hat{\boldsymbol{\sigma}}^{\boldsymbol{2}} = (\hat{\sigma}_0^2, \ldots, \hat{\sigma}_{S-1}^2)'$ using

$$\hat{\sigma}_i^2 = \frac{1}{N} \sum_{k=0}^{N-1} \left(X_{kS+i} - \hat{X}_{kS+i}^{(i)} \right)^2 / r_{kS+i} = S_i/N,$$

The function *optim* in R is used, a general-purpose optimization based on Nelder–Mead, quasi-Newton and conjugate-gradient algorithms. It includes an option for box-constrained optimization and simulated annealing. Several different options were explored, using our likelihood function, and the initial values from Table 4. A summary of the performance of the different options is provided in Table 5.2. The BFGS option was found to be the fastest convergent method. This option uses the results published simultaneously by Broyden, Fletcher, Goldfarb and Shanno in 1970 [12], [17], [19] and [38], which uses function values and gradients to approximate the optimization surface. The parameter estimates that resulted using the BFGS algorithm are listed in Table 5.3.

Method	$-2\log\{L(\hat{\boldsymbol{\beta}})\}$	Running time	Convergence	Depending on initial value
BFGS	463.8585	137s	Yes	Yes
CG	-505.37	282s	No	Yes
SANN	-362.65	378s	Yes	Yes
Nelder-Mead	1182.61	17.29s	No	Yes

Table 5.2: A compare of different algorithm in *optim* function.

In order to test the convergence, we use the result in Table 5.3 as an initial value, and run the optimization procedure again. The algorithm converges quickly to the same value

Season	0	1	2	3	4	5
$\hat{\theta}_i$	0.586	0.154	0.119	-0.008	0.085	-0.206
$\hat{\phi}_i$	0.546	0.800	0.709	0.679	0.738	0.884
σ_i	0.165	0.187	0.164	0.159	0.135	0.134
Season	6	7	8	9	10	11
$\hat{\theta}_i$	0.204	0.071	1.035	0.398	-0.110	0.415
$\hat{\phi}_i$	0.720	0.162	-0.690	0.417	0.855	0.558
σ_i	0.260	0.184	0.127	0.143	0.108	0.138

Table 5.3: MLE result by BFGS method, with σ_i as nuisance parameters. The resulting MLE value was $-2\log L(\hat{\boldsymbol{\beta}}) = -463.8585$.

of $-2\log\{L(\hat{\boldsymbol{\beta}})\}$ as before.

Next we plot $-2\log\{L(\beta, \sigma^2)\}$ as a function of each σ_i, for $i = 0, 1, \ldots, 11$. In this way, we obtain 12 plots, shown in Figure 5.1. There is a clear global local minimum point on each plot, which we compute using the *optimize* function in R to find the corresponding σ_i value, and denote it as σ_i^*, with $\boldsymbol{\sigma}^* = (\sigma_0^*, \sigma_1^*, \ldots, \sigma_{11}^*)'$. With this σ_i^* value, we run *optim* again for $-2\log\{L(\beta, \sigma^2)\}$, using the same BFGS algorithm, resulting in a lower negative likelihood value of -506.6053. For different starting values and different iterations, the routine converges to the same optimum. See results in Table 5.4.

Finally, to check our results, we use a one variable optimization to minimize the negative log likelihood function as a function of each individual variable, by similar techniques to find the global minimum points $\boldsymbol{\theta}^* = (\theta_0^*, \theta_1^*, \ldots, \theta_{11}^*)'$ on Figure 5.2 and $\boldsymbol{\phi}^* = (\phi_0^*, \phi_1^*, \ldots, \phi_{11}^*)'$ on Figure 5.3. Since these results have been consistent with the MLE results from BFGS optimization algorithm shown in Table 5.4, then we are confident that the values reported in Table 5.4 represent the true MLE. Therefore we take the results in Table 5.4 as our final optimization results for model parameters.

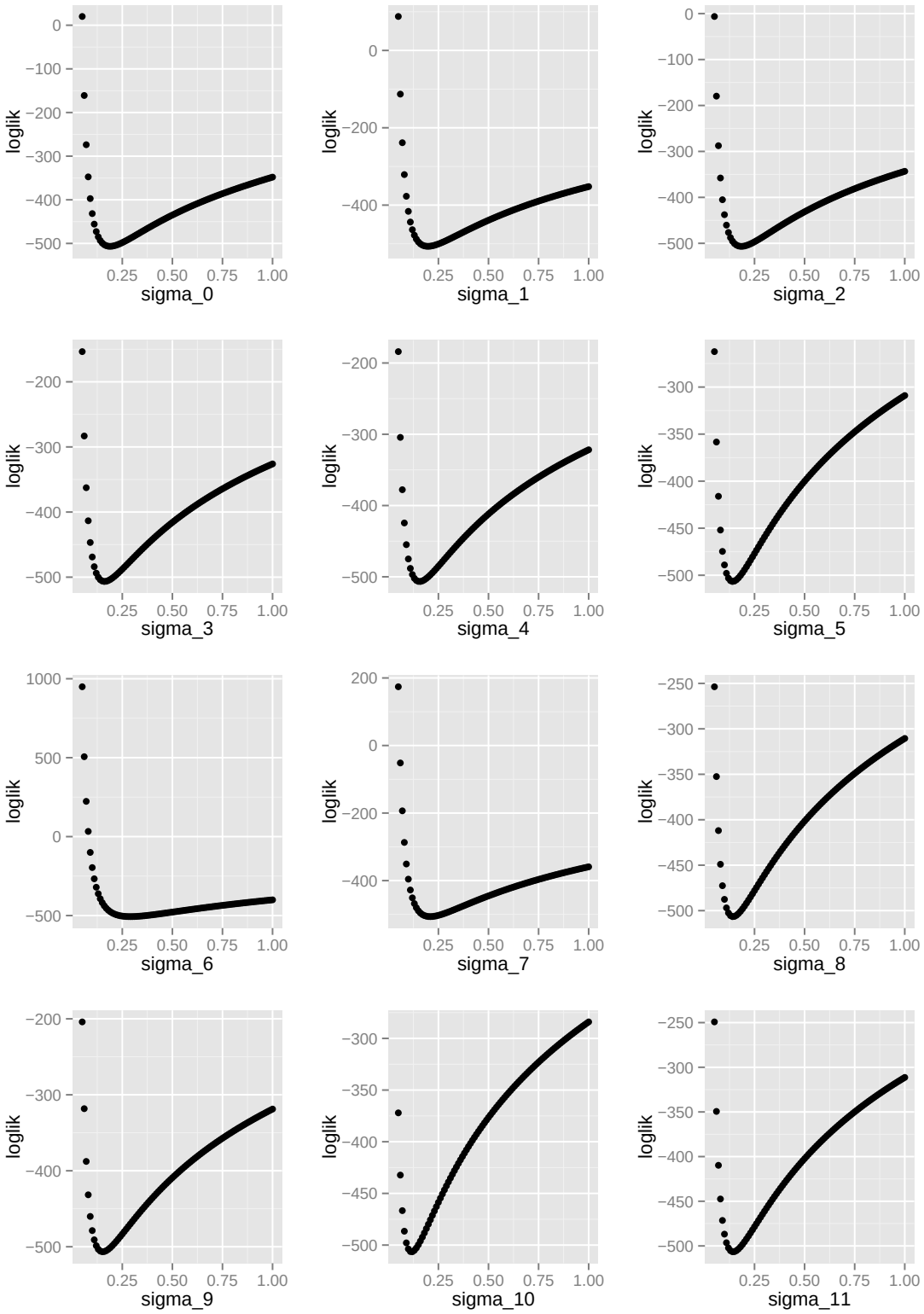


Figure 5.1: $-2\log\{L(\beta, \sigma^2)\}$ as a function of each σ_i, $i = 0, \ldots, 11$, with the remaining parameters fixed.

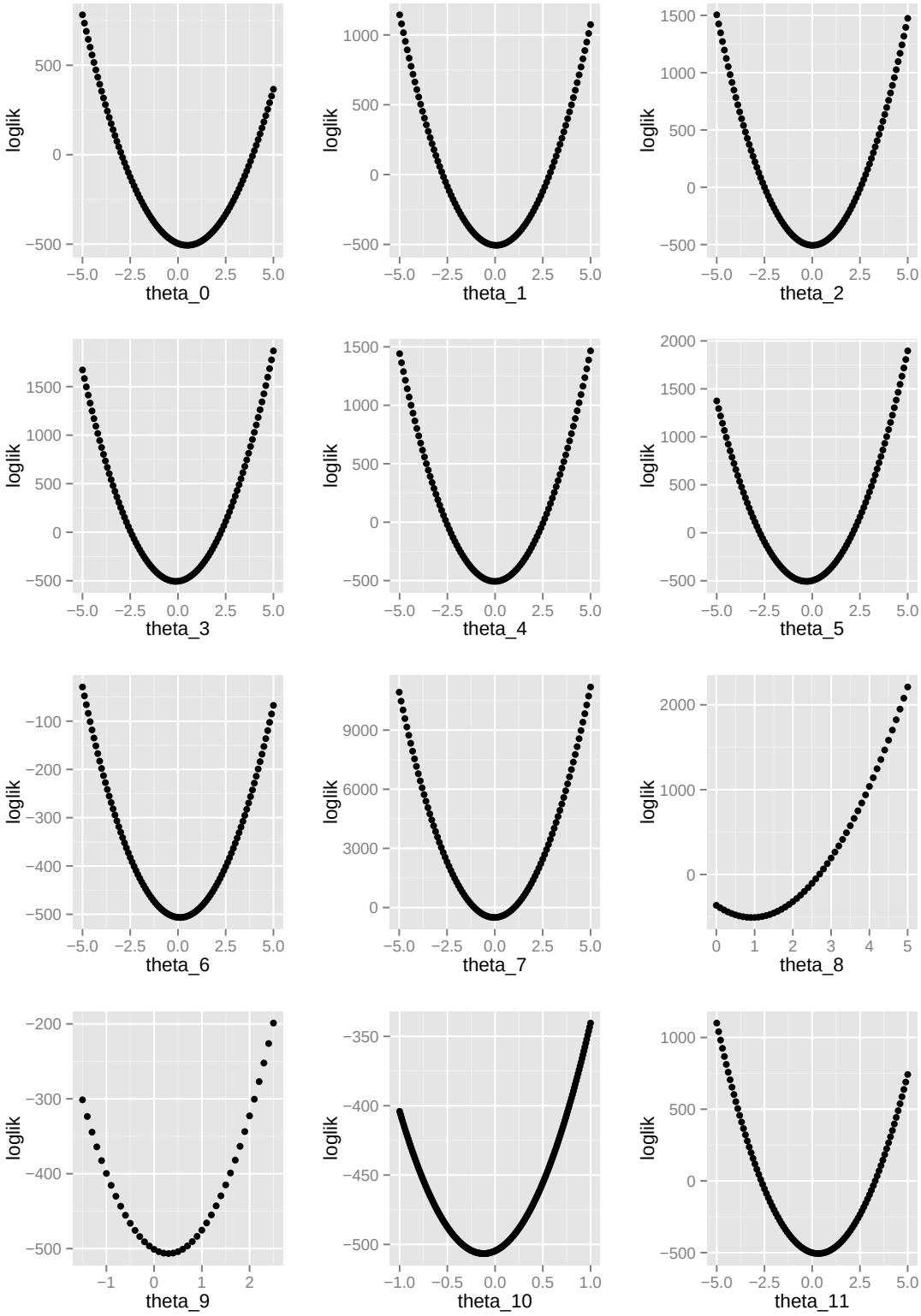


Figure 5.2: $-2\log\{L(\beta, \sigma^2)\}$ as a function of each θ_i, $i = 0, \ldots, 11$, with the remaining parameters fixed.

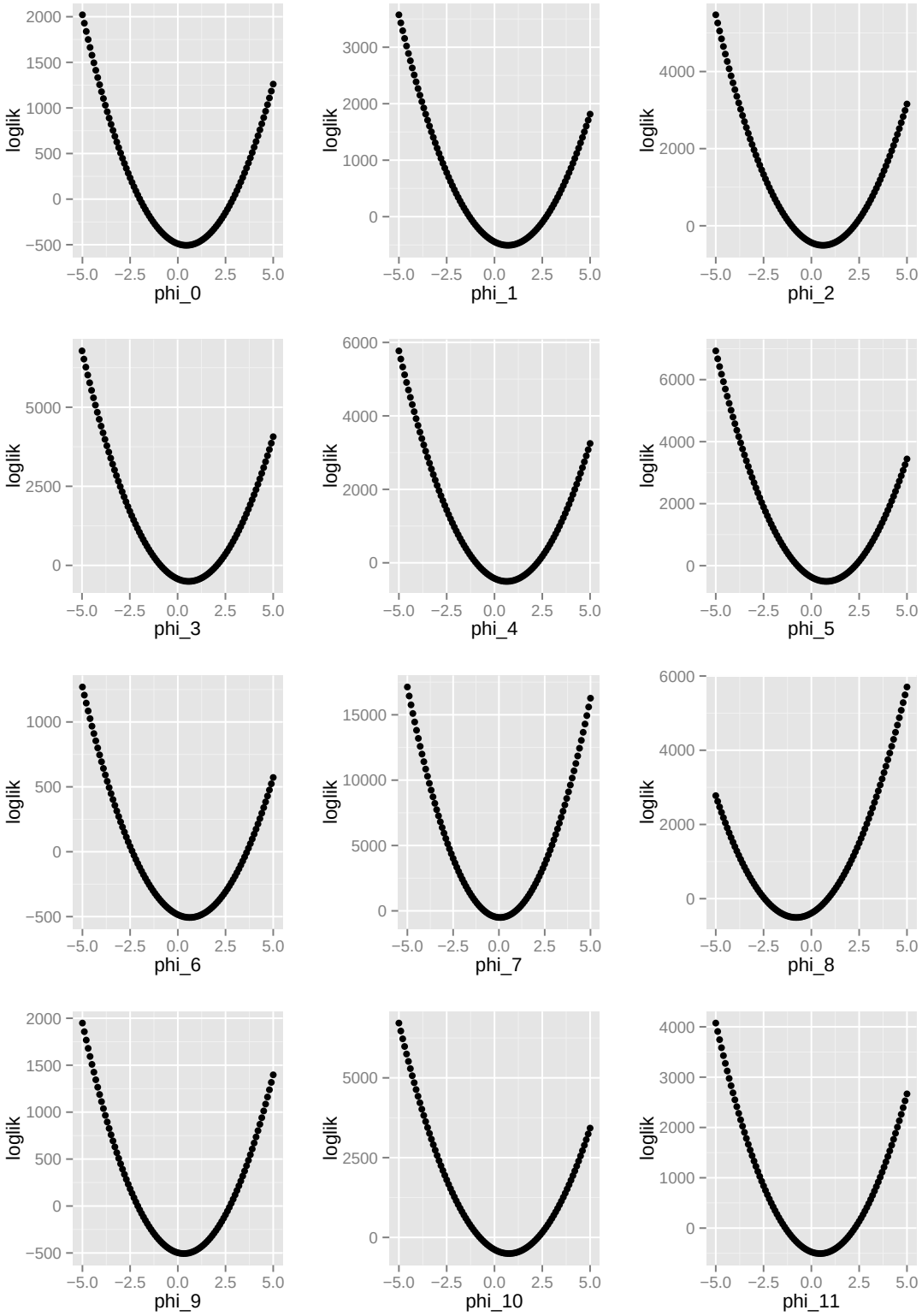


Figure 5.3: $-2\log\{L(\beta, \sigma^2)\}$ as a function of each ϕ_i, $i = 0, \ldots, 11$, with the remaining parameters fixed.

Season	0	1	2	3	4	5
$\hat{\theta}_i$	0.586	0.154	0.119	-0.008	0.085	-0.206
$\hat{\phi}_i$	0.546	0.800	0.709	0.679	0.738	0.884
σ_i^*	0.199	0.208	0.194	0.171	0.166	0.151
Season	6	7	8	9	10	11
$\hat{\theta}_i$	0.204	0.071	1.035	0.398	-0.110	0.415
$\hat{\phi}_i$	0.720	0.162	-0.690	0.417	0.855	0.558
σ_i^*	0.300	0.219	0.153	0.162	0.126	0.154

Table 5.4: MLE by BFGS method, using the values of $\boldsymbol{\sigma}^*$ from Figure 5.1. The resulting MLE value was $-2\log L(\hat{\boldsymbol{\beta}}) = -506.6053$. We take those parameters as our best results in optimization.

5.2 Reduced PARMA$_S(1,1)$ model.

To obtain additional parsimony, it is also advisable to consider simpler models where some statistically insignificant model parameters are set to zero.

5.2.1 Reduced model by asymptotic distribution of $\hat{\psi}_i(\ell)$

In the discussion of a reduced PARMA$_S(1,1)$ model, by (3.0.8),

$$\hat{\phi}_t(1) = \hat{\psi}_t(2)/\hat{\psi}_{t-1}(1) \quad \text{and} \quad \hat{\theta}_t(1) = \hat{\psi}_t(1) - \hat{\phi}_t(1).$$

In order to obtain the estimates of the PARMA$_{12}(1,1)$ model parameters, we need only consider $\hat{\psi}_i(\ell)$ at lag 1 and lag 2. An α-level test statistic rejects the null hypothesis $\big(H_0 : \psi_i(\ell) = 0\big)$ in favor of the alternative $\big(H_a : \psi_i(\ell) \neq 0$, indicating that the model parameter is statistically significantly different from zero) if $|Z| > z_{\alpha/2}$. The p-value for this test is given by (3.0.7), and it gives us a way to determine which coefficients in the identified PARMA model are statistically significantly different from zero (those with a small p-value,

i	0	1	2	3	4	5	6	7	8	9	10	11
θ_i	1.04	0.20	-0.02	-0.03	0.42	-0.40	0.87	0.00	-0.71	0.79	-0.29	0.54
ϕ_i	0.00	0.74	0.71	0.71	0.43	1.02	0.00	0.00	1.00	0.00	0.97	0.41
σ_i	0.16	0.18	0.16	0.15	0.13	0.13	0.26	0.18	0.12	0.14	0.10	0.13

Table 5.5: Parameter estimates for the reduced PARMA model (1.0.1) of average monthly flow series for the Fraser River near Hope BC from October 1912 to September 1982. The resulting negative likelihood value is -351.3939.

say, $p < 0.05$). Coefficients with a higher p-value are shown in bold font in Table 3.1, and we set coefficients $\hat{\psi}_i(\ell) = 0$ in that case. Using (3.0.8), we then list in Table 5.5 the parameter estimates of the reduced $\text{PARMA}_{12}(1,1)$ model, where autoregressive coefficients in season 0, 6, 7 and 9 are set to 0.

Here we apply innovations algorithm to get our model parameters. Later in this chapter we will shown a different method using MLE. Once we obtain the estimates for the reduced $\text{PARMA}_{12}(1,1)$ model parameters, we can carry out forecast procedures similar to the full $\text{PARMA}_{12}(1,1)$ model, as follows. Since autoregressive coefficients in season 0, 6, 7 and 9 are set to 0, the computation is simpler for the W_t process.

- Compute the transformed process (2.1.2) using the reduced model parameters.
- Compute the sample autocovariance of that process by Proposition 2.1.3.
- Apply the innovations algorithm (2.1.9) to get the projection coefficients $\theta_{n,j}$.
- Use (2.1.13) to compute the one-step-ahead predictors $\hat{X}_n$ for $n = 1, 2, \ldots, 840 = 70 \times 12$.
- Apply (2.1.17) to get the forecasts.
- Use the asymptotic formula (2.2.3) to compute 95% prediction bounds, based on the assumption of Gaussian innovations.

The resulting prediction, along with the 95% prediction bands, is shown in Figure 5.4. The actual data (solid line) is also shown for comparison. Note that the forecast (solid line with dots) is in reasonable agreement with the actual data (which were not used in the forecast), and that the actual data lies well within the 95% prediction bands. Furthermore, we give a detailed comparison in Figure 5.5 between full model and the reduced one. As shown in the first two graphs, the prediction results are quite similar. In a further investigation, we check the difference and relative error between the two forecasts, for every step. The differences are small. In fact, when compared to predictions, the relative errors are small. When the step $h \geq 5$, the relative errors are nearly zero. We conclude that the removed autoregressive coefficients in season 0, 6, 7 and 9 are insignificant, and the reduced model is a viable substitute for the full model. We prefer the reduced model, since it has fewer parameters.

5.2.2 Reduced model by asymptotic distribution of MLE

Another method for obtaining a reduced model is to apply the asymptotics of MLE for the PARMA process, which were developed in Basawa and Lund [24]. For a causal and invertible Gaussian PARMA model, with the assumption of $\{\varepsilon_t\}$ being periodic i.i.d. Gaussian noise, Basawa and Lund [24, Theorem 3.1] gives the asymptotic distribution of $\hat{\boldsymbol{\beta}}$, as $N \to \infty$,

$$N^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to \mathrm{N}\left(\mathbf{0}, A^{-1}(\boldsymbol{\beta}, \boldsymbol{\sigma}^{\mathbf{2}})\right), \tag{5.2.1}$$

where

$$A(\boldsymbol{\beta}, \boldsymbol{\sigma}^{\mathbf{2}}) = \sum_{i=0}^{S-1} \sigma_i^{-2} \Gamma_i(\boldsymbol{\beta}, \boldsymbol{\sigma}^{\mathbf{2}}), \tag{5.2.2}$$

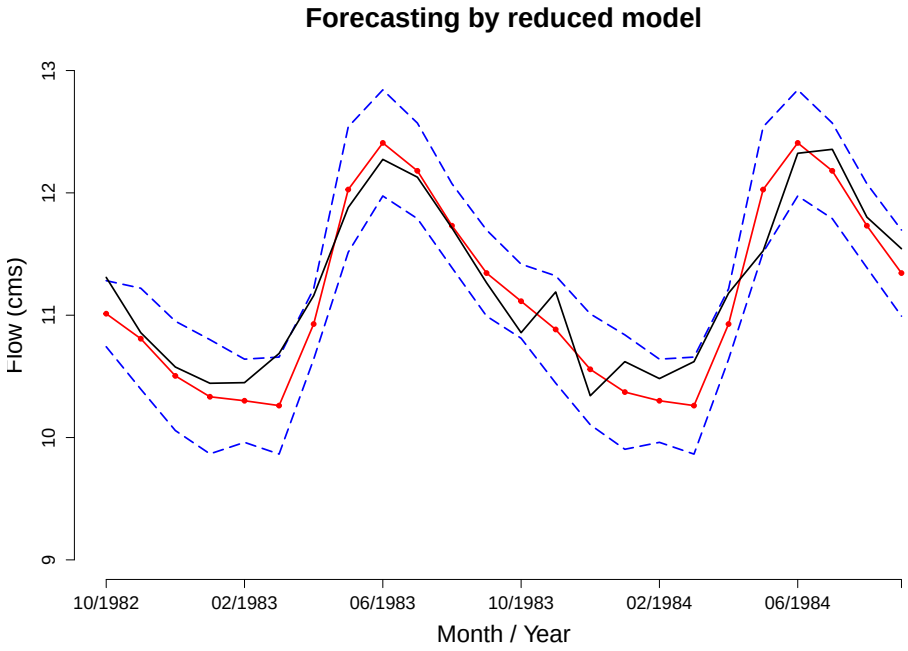


Figure 5.4: A 24-month forecast using the reduced $\text{PARMA}_{12}(1,1)$ model. Solid line is the original data; Solid line with dots is the reduced PARMA(1,1) forecast; The two dotted lines are 95% Gaussian prediction bounds. The 24-month forecast is from October 1982 to September 1984.

and

$$\Gamma_i(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) = \mathrm{E}\left[\left(\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)'\right].$$

Furthermore, Basawa and Lund [24, Remark 3.1] states that if $\{\varepsilon_t\}$ is Gaussian, then the maximum likelihood estimate has the same asymptotic distribution as the weighted least squares estimate. Once we obtain the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$, the MLE of σ_i^2 for $0 \le i \le S-1$ could be computed by

$$\hat{\sigma}_i^2 = \frac{1}{N}\sum_{k=0}^{N-1}\left(X_{kS+i} - \hat{X}_{kS+i}^{(i)}\right)^2 / r_{kS+i} = S_i/N, \tag{5.2.3}$$

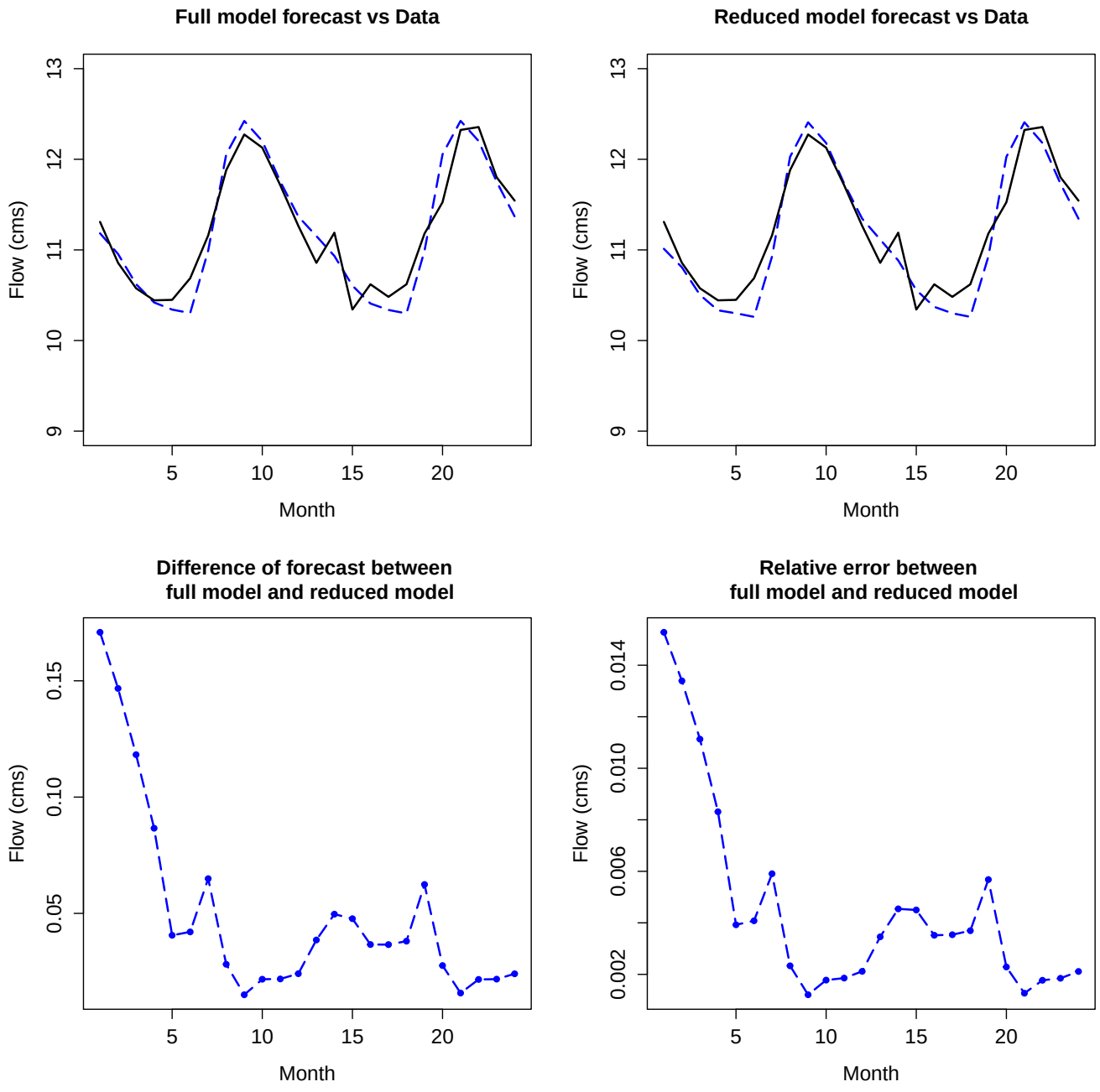


Figure 5.5: A comparison of 24-month forecasts between the full $\text{PARMA}_{12}(1,1)$ model and reduced $\text{PARMA}_{12}(1,1)$ model. In the first two graphs, dotted line is forecast, and solid line is real data.

where $\hat{\boldsymbol{\sigma}}^{\mathbf{2}} = (\hat{\sigma}_0^2, \ldots, \hat{\sigma}_{S-1}^2)'$.

Example 5.2.1. We adopt the causal and invertible first-order $\text{PARMA}_{12}(1,1)$ model from Example 5.1.3, and show how to compute the asymptotic covariance matrix $A^{-1}(\boldsymbol{\beta}, \boldsymbol{\sigma}^{\mathbf{2}})$. The results are shown in Table 5.6, where $se = \dfrac{\text{diag}(A^{-1})}{N^{1/2}}$ stands for the standard error for $\hat{\boldsymbol{\beta}}$. In the following, $p = q = 1$, and $S = 12$ denotes the number of seasons, and

Season	$\sqrt{\hat{\gamma}_i(0)}$	$se(\hat{\theta}_i)$	$se(\hat{\phi}_i)$
0	0.06	0.36	0.30
1	0.07	0.26	0.21
2	0.06	0.23	0.19
3	0.05	0.22	0.18
4	0.05	0.27	0.21
5	0.05	0.24	0.17
6	0.16	0.29	0.19
7	0.08	0.08	0.06
8	0.11	0.25	0.19
9	0.05	0.19	0.09
10	0.03	0.18	0.14
11	0.03	0.24	0.18

Table 5.6: MLE and its standard error.

$i = 0, 1, \ldots, S-1$, and $\langle t \rangle$ denotes the corresponding season index for t, where

$$\langle t \rangle = \begin{cases} t - S[t/S] & \text{if} \quad t \geq 0, \\ S + t - S[t/S + 1] & \text{if} \quad t < 0, \end{cases}$$

and $[\,\cdot\,]$ stands for the greatest integer function. The model is

$$X_t = \phi_t X_{t-1} + \varepsilon_t + \theta_t \varepsilon_{t-1}. \tag{5.2.4}$$

Taking partial derivations in (5.2.4) gives

$$\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -\boldsymbol{e}_{2\langle t\rangle+1} X_{t-1} - \theta_t \frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \boldsymbol{e}_{2\langle t\rangle+2} \varepsilon_{t-1}(\boldsymbol{\beta}), \tag{5.2.5}$$

where $\boldsymbol{e}_j$ denotes a $(p+q)S \times 1$ unit vector whose entries are all zero, except for a 1 in the j^{th} row. Note that in the following X_{t-1}, θ_t and $\varepsilon_{t-1}(\boldsymbol{\beta})$ are scalars, however $\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ and $\boldsymbol{e}_j$ denote the $(p+q)S \times 1$ vectors. Therefore by causality

$$\mathrm{E}\left[\mathrm{X_{t-1}}\varepsilon_{\mathrm{t-1}}(\boldsymbol{\beta})\right] = \mathrm{E}\left[\varepsilon_{\mathrm{t-1}}(\boldsymbol{\beta})\mathrm{X_{t-1}}\right] = \mathrm{E}\left[\varepsilon^2_{\mathrm{t-1}}(\boldsymbol{\beta})\right] = \sigma^2_{\mathrm{t-1}},$$

by Leibniz integral rule, provided that $\varepsilon_{t-1}(\boldsymbol{\beta})$ and $\frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$ are both continuous, we take the derivative outside the expectation,

$$\mathrm{E}\left[\frac{\partial \varepsilon_{\mathrm{t-1}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\mathrm{X_{t-1}}\right] = \frac{\partial}{\partial \boldsymbol{\beta}}\mathrm{E}\left[\varepsilon_{t-1}(\boldsymbol{\beta})X_{t-1}\right] = \frac{\partial}{\partial \boldsymbol{\beta}}\left[\sigma^2_{t-1}\right] = \mathbf{0},$$

where $\mathbf{0}$ is a $(p+q)S \times 1$ zero vector. Similarly

$$\mathrm{E}\left[X_{t-1}\left(\frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)'\right] = \mathbf{0}',$$

and

$$\mathrm{E}\left[\frac{\partial \varepsilon_{\mathrm{t-1}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\varepsilon_{\mathrm{t-1}}(\boldsymbol{\beta})\right] = \frac{\partial}{\partial \boldsymbol{\beta}}\mathrm{E}\left[\varepsilon_{t-1}(\boldsymbol{\beta})\varepsilon_{t-1}(\boldsymbol{\beta})\right] = \frac{\partial}{\partial \boldsymbol{\beta}}\left[\sigma^2_{t-1}\right] = \mathbf{0},$$

and similarly

$$\mathrm{E}\left[\left(\frac{\partial \varepsilon_{\mathrm{t-1}}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)'\varepsilon_{\mathrm{t-1}}(\boldsymbol{\beta})\right] = \mathbf{0}'.$$

Therefore, if we multiply both sides of (5.2.5) by its own transpose and take an expectation,

we have

$$
\begin{aligned}
\mathrm{E}\left[\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\left[\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' &= \mathrm{E}\Big[-\boldsymbol{e}_{2\langle t\rangle+1}X_{t-1} - \theta_t\frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \boldsymbol{e}_{2\langle t\rangle+2}\varepsilon_{t-1}(\boldsymbol{\beta})\Big] \\
&\quad \Big[-\boldsymbol{e}_{2\langle t\rangle+1}X_{t-1} - \theta_t\frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} - \boldsymbol{e}_{2\langle t\rangle+2}\varepsilon_{t-1}(\boldsymbol{\beta})\Big]' \\
&= \theta_t^2\mathrm{E}\left[\frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\left[\frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' + \mathrm{var}(X_{t-1})E_{2\langle t\rangle+1,2\langle t\rangle+1} \\
&\quad + \sigma_{t-1}^2\left[E_{2\langle t\rangle+2,2\langle t\rangle+2} + E_{2\langle t\rangle+2,2\langle t\rangle+1} + E_{2\langle t\rangle+1,2\langle t\rangle+2}\right] \\
&\quad + \theta_t\left(\boldsymbol{e}_{2\langle t\rangle+1}\mathbf{0}' + \mathbf{0}\boldsymbol{e}_{2\langle t\rangle+1}' + \boldsymbol{e}_{2\langle t\rangle+2}\mathbf{0}' + \mathbf{0}\boldsymbol{e}_{2\langle t\rangle+2}'\right) \\
&= \theta_t^2\mathrm{E}\left[\frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\left[\frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' + \mathrm{var}(X_{t-1})E_{2\langle t\rangle+1,2\langle t\rangle+1} \\
&\quad + \sigma_{t-1}^2\left[E_{2\langle t\rangle+2,2\langle t\rangle+2} + E_{2\langle t\rangle+2,2\langle t\rangle+1} + E_{2\langle t\rangle+1,2\langle t\rangle+2}\right] \\
&\quad + O \\
&= \theta_t^2\mathrm{E}\left[\frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\left[\frac{\partial \varepsilon_{t-1}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]' + \mathrm{var}(X_{t-1})E_{2\langle t\rangle+1,2\langle t\rangle+1} \\
&\quad + \sigma_{t-1}^2\left[E_{2\langle t\rangle+2,2\langle t\rangle+2} + E_{2\langle t\rangle+2,2\langle t\rangle+1} + E_{2\langle t\rangle+1,2\langle t\rangle+2}\right],
\end{aligned}
$$

where $\mathrm{var}(X_t) = \gamma_t(0)$ is the variance of X_t, and $E_{i,j} = e_i e_j'$ denotes a $(p+q)S \times (p+q)S$ matrix whose entries are all zero, except for a one in the i^{th} row and j^{th} column, and O is a $(p+q)S \times (p+q)S$ zero matrix. Recall that

$$
\Gamma_t(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) = \mathrm{E}\left[\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]\left[\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right]',
$$

then we obtain

$$
\Gamma_t(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) = \theta_t^2\Gamma_{t-1}(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) + M_t,
$$

where

$$M_t = \sigma^2_{t-1}\left[E_{2\langle t\rangle+2,2\langle t\rangle+2} + E_{2\langle t\rangle+2,2\langle t\rangle+1} + E_{2\langle t\rangle+1,2\langle t\rangle+2}\right] + \mathrm{var}(X_{t-1})E_{2\langle t\rangle+1,2\langle t\rangle+1}.$$

For simplicity we adopt the notation

$$\Gamma_i(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) = \theta_i^2\Gamma_{i-1}(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) + M_i, \tag{5.2.6}$$

with the boundary condition $\Gamma_0(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) = \Gamma_S(\boldsymbol{\beta}, \boldsymbol{\sigma^2})$, hence in this way all the index i are bounded, $0 \le i \le S-1$. The solution to (5.2.6) is

$$\Gamma_i(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) = r^2_{\theta,i}\sum_{k=0}^{i}\frac{M_k}{r^2_{\theta,k}} + (\frac{r^2_{\theta,S-1}}{1-r^2_{\theta,S-1}})r^2_{\theta,i}\sum_{k=0}^{S-1}\frac{M_k}{r^2_{\theta,k}}, \tag{5.2.7}$$

where

$$r_{\theta,t} = \prod_{j=0}^{t}\theta_t, \tag{5.2.8}$$

the proof for (5.2.7) is given as follows. By (5.2.6), $\forall \quad 0 \le i \le S-1$, we could write

recursively:

$$
\begin{aligned}
\Gamma_i &= \theta_i^2 \Gamma_{i-1} + M_i \\
&= \theta_i^2(\theta_{i-1}^2 \Gamma_{i-2} + M_{i-1}) + M_i \\
&= \theta_i^2 \theta_{i-1}^2(\theta_{i-2}^2 \Gamma_{i-3} + M_{i-2}) + \theta_i^2 M_{i-1} + M_i \\
&= \theta_i^2 \theta_{i-1}^2 \theta_{i-2}^2 \Gamma_{i-3} + \theta_i^2 \theta_{i-1}^2 M_{i-2} + \theta_i^2 M_{i-1} + M_i \\
&\dots \\
&= \left(\theta_i \theta_{i-1} \cdots \theta_{i-(S-1)}\right)^2 \Gamma_{i-S} \\
&+ \left(\theta_i \theta_{i-1} \cdots \theta_{i-(S-2)}\right)^2 M_{i-(S-1)} \cdots + \left(\theta_i \theta_{i-1} \cdots \theta_0\right)^2 M_{-1} \\
&+ \left(\theta_i \theta_{i-1} \cdots \theta_1\right)^2 M_0 + \left(\theta_i \theta_{i-1} \cdots \theta_2\right)^2 M_1 + \ldots + (\theta_i \theta_{i-1})^2 M_{i-2} + \theta_i^2 M_{i-1} + M_i \\
&= \left(\theta_i \theta_{i-1} \cdots \theta_{i-(S-1)}\right)^2 \Gamma_{i-S} + \mathrm{I} + \mathrm{II}, \qquad (5.2.9)
\end{aligned}
$$

where we define

$$
\begin{aligned}
\mathrm{I} &= \left(\theta_i \theta_{i-1} \cdots \theta_{i-(S-2)}\right)^2 M_{i-(S-1)} \cdots + \left(\theta_i \theta_{i-1} \cdots \theta_0\right)^2 M_{-1} \\
\mathrm{II} &= \left(\theta_i \theta_{i-1} \cdots \theta_1\right)^2 M_0 + \left(\theta_i \theta_{i-1} \cdots \theta_2\right)^2 M_1 + \ldots + (\theta_i \theta_{i-1})^2 M_{i-2} + \theta_i^2 M_{i-1} + M_i,
\end{aligned}
$$

since $\Gamma_{i-S} = \Gamma_i$ by periodic property, $\theta_i \theta_{i-1} \cdots \theta_{i-(S-1)} = r_{\theta,S-1}$ by definition in

(5.2.8), we could write (5.2.9) as

$$\begin{aligned}\Gamma_i &= \frac{\mathrm{I}+\mathrm{II}}{1-r^2_{\theta,S-1}} \\ &= \frac{\mathrm{II}(1-r^2_{\theta,S-1})+\mathrm{I}+\mathrm{II}(r^2_{\theta,S-1})}{1-r^2_{\theta,S-1}} \\ &= \mathrm{II}+\frac{\mathrm{I}+\mathrm{II}(r^2_{\theta,S-1})}{1-r^2_{\theta,S-1}}. \end{aligned} \tag{5.2.10}$$

Note that

$$\begin{aligned}\mathrm{II} &= \left(\theta_i\theta_{i-1}\cdots\theta_1\right)^2 M_0+\left(\theta_i\theta_{i-1}\cdots\theta_2\right)^2 M_1+\ldots+(\theta_i\theta_{i-1})^2 M_{i-2}+\theta_i^2 M_{i-1}+M_i \\ &= \left(\theta_i\theta_{i-1}\cdots\theta_0\right)^2\left(\frac{M_0}{\theta_0^2}+\frac{M_1}{(\theta_0\theta_1)^2}+\ldots+\frac{M_i}{(\theta_0\theta_1\cdots\theta_i)^2}\right) \\ &= r^2_{\theta,i}\left(\frac{M_0}{r^2_{\theta,0}}+\frac{M_1}{r^2_{\theta,1}}+\ldots+\frac{M_i}{r^2_{\theta,i}}\right) \\ &= r^2_{\theta,i}\sum_{k=0}^{i}\frac{M_k}{r^2_{\theta,k}}. \end{aligned} \tag{5.2.11}$$

On the other hand, we want to prove that

$$\frac{\mathrm{I}+\mathrm{II}(r^2_{\theta,S-1})}{1-r^2_{\theta,S-1}} = \left(\frac{r^2_{\theta,S-1}}{1-r^2_{\theta,S-1}}\right)r^2_{\theta,i}\sum_{k=0}^{S-1}\frac{M_k}{r^2_{\theta,k}}, \tag{5.2.12}$$

in this way we will finish the proof of (5.2.7). Note that

$$
\begin{aligned}
\mathrm{I} + (r^2_{\theta,S-1})\mathrm{II} &= \left(\theta_i\theta_{i-1}\cdots\theta_{i-(S-2)}\right)^2 M_{i-(S-1)}\ldots + \left(\theta_i\theta_{i-1}\cdots\theta_0\right)^2 M_{-1} \\
&+ \left(\theta_{S-1}\theta_{S-2}\cdots\theta_0\right)^2 \left(\theta_i\theta_{i-1}\cdots\theta_0\right)^2 \left(\frac{M_0}{\theta_0^2} + \ldots + \frac{M_i}{(\theta_0\theta_1\cdots\theta_i)^2}\right) \\
&= \left(\theta_i\theta_{i-1}\cdots\theta_{i-(S-2)}\right)^2 M_{i+1}\ldots + \left(\theta_i\theta_{i-1}\cdots\theta_0\right)^2 M_{S-1} \\
&+ r^2_{\theta,S-1} r^2_{\theta,i} \sum_{k=0}^{i} \frac{M_k}{r^2_{\theta,k}} \\
&= \left(\theta_{S-1}\cdots\theta_0\right)^2 \left(\theta_i\cdots\theta_0\right)^2 \left(\frac{M_{i+1}}{(\theta_0\cdots\theta_{i+1})^2} + \ldots + \frac{M_{S-1}}{(\theta_0\cdots\theta_{S-1})^2}\right) \\
&+ r^2_{\theta,S-1} r^2_{\theta,i} \sum_{k=0}^{i} \frac{M_k}{r^2_{\theta,k}} \\
&= r^2_{\theta,S-1} r^2_{\theta,i} \sum_{k=i+1}^{S-1} \frac{M_k}{r^2_{\theta,k}} + r^2_{\theta,S-1} r^2_{\theta,i} \sum_{k=0}^{i} \frac{M_k}{r^2_{\theta,k}} \\
&= r^2_{\theta,S-1} r^2_{\theta,i} \sum_{k=0}^{S-1} \frac{M_k}{r^2_{\theta,k}},
\end{aligned}
$$

dividing both sides by $1 - r^2_{\theta,S-1}$, (5.2.12) is proved. Add (5.2.11) and (5.2.12) into (5.2.10), we complete the proof of (5.2.7).

The invertibility of the model requires that $|r_{\theta,S-1}| < 1$. In the end, $A^{-1}(\boldsymbol{\beta}, \boldsymbol{\sigma^2})$ is a $(p+q)S \times (p+q)S$ matrix, and it could be computed by the inverse of $A(\boldsymbol{\beta}, \boldsymbol{\sigma^2})$ in (5.2.2).

Equation (5.2.1) can be used to produce confidence intervals and hypothesis tests for the model parameters $\boldsymbol{\beta}$. An α-level test statistic rejects the null hypothesis $H_0 : \boldsymbol{\beta}_i = 0$ in favor of the alternative hypothesis $H_a : \boldsymbol{\beta}_i \neq 0$, indicating that $\boldsymbol{\beta}_i$ is statistically significantly

different from zero if $|Z| > z_{\alpha/2}$. The p value for this test is

$$p = P(|Z| > |z|), \tag{5.2.13}$$

where $Z \sim \mathcal{N}(0, 1)$, and

$$Z = (\hat{\boldsymbol{\beta}} - \mathbf{0})/se.$$

After we get the MLE optimization results from Example 5.1.3 and Table 5.4, we do a hypothesis test for the model parameters $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$, where the standard error se is obtained from Table 5.6. Using the level of significance of $\alpha = 5\%$, we get the reduced model parameters, which achieve the goal of parsimony. The results are shown in Table 5.7.

Season	0	1	2	3	4	5	6	7	8	9	10	11
$\hat{\theta}_i$	0.55	-0.01	0.04	-0.02	0.09	-0.26	0.01	0.37	1.89	0.27	-0.08	0.32
p-value	0.13	0.96	0.87	0.93	0.74	0.27	0.98	0.00	0.00	0.16	0.64	0.18
reduced $\hat{\theta}_i$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.37	1.89	0.00	0.00	0.00
$\hat{\phi}_i$	0.37	0.63	0.52	0.54	0.71	0.86	1.18	0.23	-1.41	0.35	0.57	0.37
p-value	0.22	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
reduced $\hat{\phi}_i$	0.00	0.63	0.52	0.54	0.71	0.86	1.18	0.23	-1.41	0.35	0.57	0.37

Table 5.7: Model parameters by MLE optimization and their p-values.

We obtain the forecast in a similar manner as the previous reduced model. The resulting prediction, along with the 95% prediction bands, is shown in Figure 5.6. The actual data (solid line) is also shown for comparison. The forecast (solid line with dots) is in reasonable agreement with the actual data (which were not used in the forecast), and the actual data lies well within the 95% prediction bands. A detailed comparison can be seen in Figure 5.7 between full model and the reduced one. As shown in the first two graphs, the reduced model predicts as well as the full model. In the next two graphs, we check the difference and

relative error between the two forecasts, for every step. The differences are small. In fact, when compared to predictions, the relative errors are quite small. When the step $h \geq 5$, the relative errors are nearly zero. We conclude that the removed autoregressive coefficients in Table 5.7 are insignificant, and the reduced model performs as well as the full model. In summary, we prefer the reduced $\text{PARMA}_{12}(1,1)$ model simply because it is a simpler model with fewer parameters and the same adequacy.

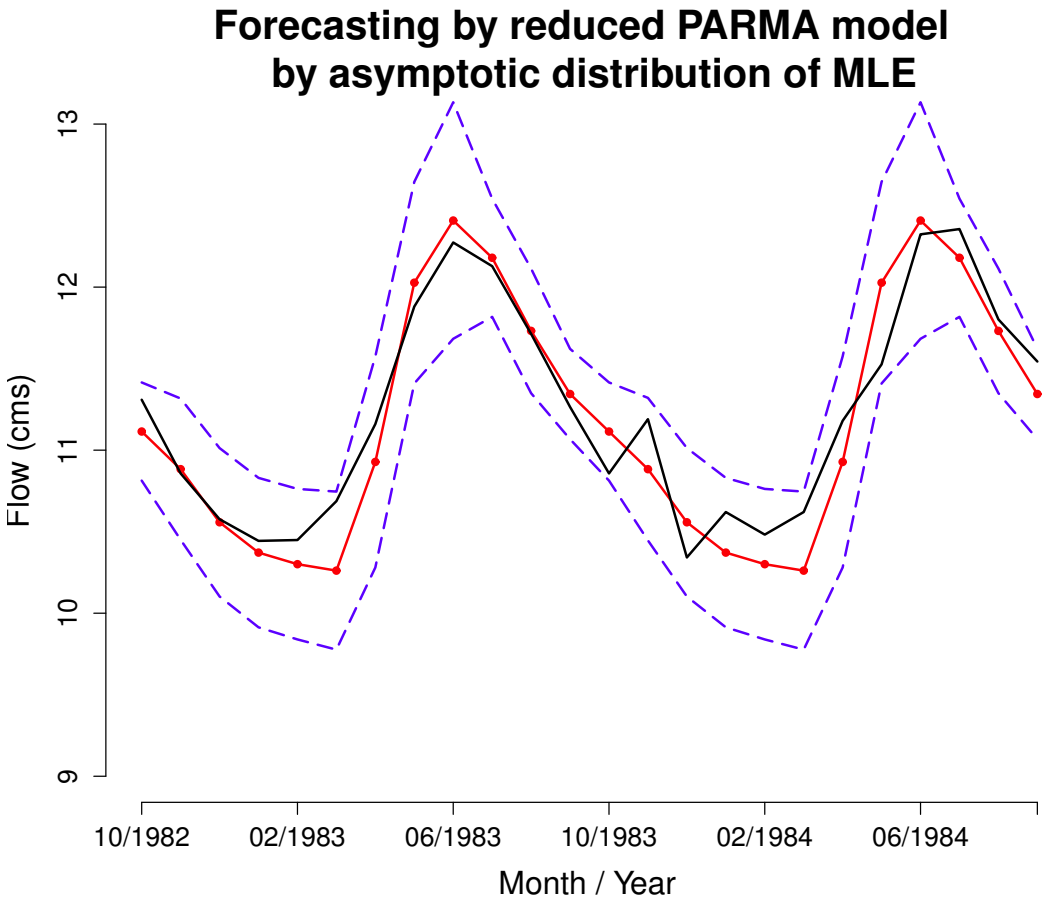


Figure 5.6: A 24-month forecast using the reduced $\text{PARMA}_{12}(1,1)$ model in Table 5.7. Solid line is the original data; Solid line with dots is the reduced PARMA(1,1) forecast; The two dotted lines are 95% Gaussian prediction bounds. The 24-month forecast is from October 1982 to September 1984.

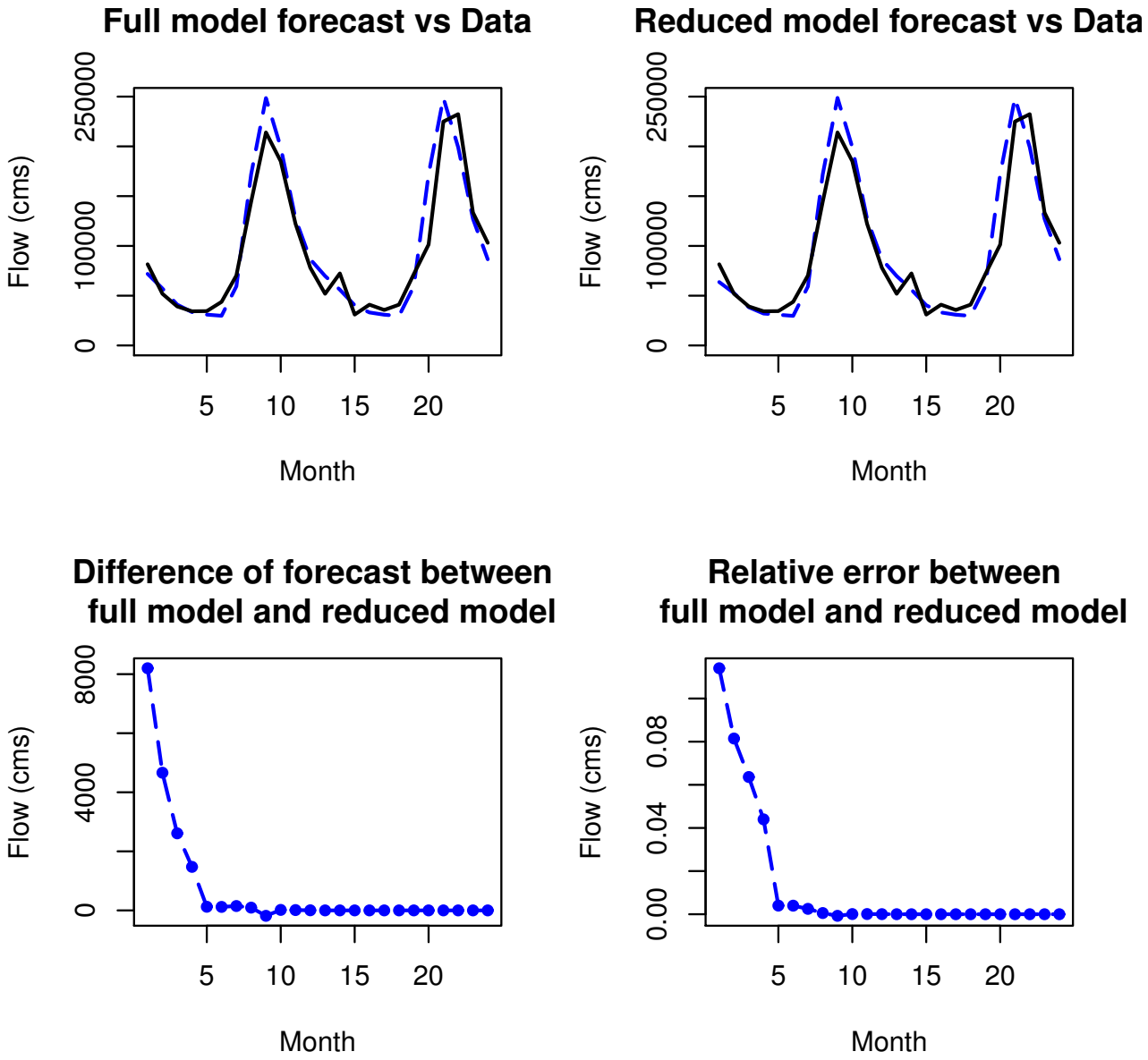


Figure 5.7: A comparison of 24-month forecasts between the full $\text{PARMA}_{12}(1,1)$ model and reduced $\text{PARMA}_{12}(1,1)$ model, by asymptotic distribution of MLE. In the first two graphs, dotted line is forecast, and solid line is real data.

Chapter 6

Asymptotic normality of PARMA model

Let $\mathbf{X}_{n,i} = (X_i, ..., X_{i+n-1})'$ be a causal and invertible PARMA$_S(p,q)$ process (1.0.1), where n is the number of observations and $i = 0, ..., S-1$. For notations, let $\boldsymbol{\phi}_t = (\phi_t(1), \ldots, \phi_t(p))'$ and $\boldsymbol{\theta}_t = (\theta_t(1), \ldots, \theta_t(q))'$ denote the autoregressive and moving-average parameters during season t, respectively. Then we can write the likelihood function as $L(\boldsymbol{\beta})$, where we use $\boldsymbol{\beta} = (\boldsymbol{\phi}_0', \boldsymbol{\theta}_0', \boldsymbol{\phi}_1', \boldsymbol{\theta}_1', \ldots, \boldsymbol{\phi}_{S-1}', \boldsymbol{\theta}_{S-1}')'$ to denote the collection of all PARMA$_S(p,q)$ parameters. The dimension of $\boldsymbol{\beta}$ is $(p+q)S \times 1$. Following the ideas from Basawa and Lund [24], we treat the white noise variances $\boldsymbol{\sigma^2} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{S-1}^2)'$ as nuisance parameters. Then the likelihood function is given by (5.1.13) in Chapter 5. By (5.1.16), once we obtain the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$, the MLE of σ_i^2 for $0 \le i \le S-1$ could be computed by

$$\hat{\sigma}_i^2 = \frac{1}{N} \sum_{k=0}^{N-1} \left(X_{kS+i} - \hat{X}_{kS+i}^{(i)} \right)^2 / r_{kS+i} = S_i/N,$$

where $\hat{\boldsymbol{\sigma}}^{\mathbf{2}} = (\hat{\sigma}_0^2, \ldots, \hat{\sigma}_{S-1}^2)'$. Equivalently, we could minimize the negative log likelihood

$$\begin{aligned}
\ell(\boldsymbol{\beta}) &= -2\log\{L(\boldsymbol{\beta})\} \\
&= n\log(2\pi) + \sum_{j=0}^{n-1} \log(v_{j,i}) + \sum_{j=0}^{n-1} \frac{(X_{i+j} - \hat{X}_{i+j}^{(i)})^2}{v_{j,i}} \\
&= n\log(2\pi) + \sum_{j=0}^{n-1} \log(\sigma_{i+j}^2 r_{j,i}) + \sum_{j=0}^{n-1} \frac{(X_{i+j} - \hat{X}_{i+j}^{(i)})^2}{\sigma_{i+j}^2 r_{j,i}} \\
&= n\log(2\pi) + \sum_{j=0}^{n-1} \log(\sigma_{i+j}^2) + \sum_{j=0}^{n-1} \log(r_{j,i}) + \sum_{j=0}^{n-1} \frac{(X_{i+j} - \hat{X}_{i+j}^{(i)})^2}{\sigma_{i+j}^2 r_{j,i}}.
\end{aligned} \tag{6.0.1}$$

Next we consider the properties of the first derivative $\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_k}$ in a few lemmas.

Lemma 6.0.2. *For k = 1, ..., p+q, there exist constants $C > 0$ and $s \in (0,1)$ such that*

$$\left| \frac{\partial r_{t,i}(\boldsymbol{\beta})}{\partial \beta_k} \right| \leq C(\boldsymbol{\beta}) s^t,$$

where $t \geq 1$, and $1 \leq k \leq p+q$.

Proof. Note that

$$r_{t,i}(\boldsymbol{\beta}) = \mathrm{E}\left(Y_{t+i} - \hat{Y}_{t+i}\right)^2 = \frac{\mathrm{E}\left(X_{t+i} - \hat{X}_{t+i}\right)^2}{\sigma_{t+i}^2},$$

where $\boldsymbol{Y}_{t,i} = (Y_i, \ldots, Y_{t+i-1})'$ is the mean zero PARMA process with period S given by

$$Y_t - \sum_{k=1}^{p} a_t(k) Y_{t-k} = Z_t + \sum_{j=1}^{q} b_t(j) Z_{t-j} \tag{6.0.2}$$

where $\{Z_t\}$ is a sequence of random variables with mean zero and standard deviation 1. Here the autoregressive parameters $a_t(j)$ and the moving average parameters $b_t(j)$ are assumed to be periodic functions of t with the same period $S \geq 1$. By the invertibility assumption, we write

$$Y_{t+i} + \sum_{j=1}^{\infty} Y_{t+i-j}\pi_{t+i}(j) = Z_{t+i}, \tag{6.0.3}$$

where $\pi_{t+i}(j) = \pi_{t+i+kS}(j)$ by the periodic property of PARMA model. In the following, we write $\pi_{t+i}(j) = \pi_{t+i}(j;\boldsymbol{\beta})$ to emphasize the dependence of $\pi_{t+i}(j)$ on $\boldsymbol{\beta}$, and the autocovariance function is

$$\eta_i(h;\boldsymbol{\beta}) = \mathrm{Cov}(Y_i, Y_{i-h}),$$

By Anderson et al. [6], the best linear predictor of Y_{t+i} is defined as

$$\hat{Y}_{t+i}^{(i)} = \phi_{t,1}^{(i)}Y_{t+i-1} + \ldots + \phi_{t,t}^{(i)}Y_i, \ t \geq 1,$$

where the vector of coefficients, $\boldsymbol{\phi}_{t,i} = \left(\phi_{t,1}^{(i)}, \ldots, \phi_{t,t}^{(i)}\right)'$, appears in the prediction equations

$$G_{t,i}\boldsymbol{\phi}_{t,i} = \boldsymbol{\eta}_{t,i},$$

where

$$\boldsymbol{\eta}_{t,i} = \left(\eta_{t+i-1}(1), \eta_{t+i-2}(2), \ldots, \eta_i(t)\right)',$$

and

$$G_{t,i} = \left[\eta_{t+i-\ell}(\ell - j;\boldsymbol{\beta})\right]_{j,\ell=1}^{t}$$

is the covariance matrix of $(Y_{t+i-1}, \ldots, Y_i)'$. Proposition 2.1.1 of Anderson et al. [6] shows

that if $\sigma_i^2 > 0$ for $i = 0, 1, \ldots, S-1$, then for a causal PARMA$_S(p, q)$ process the covariance matrix $G_{t,i}$ is nonsingular for every $t \geq 1$ and i. In this way we can generalize Corollary 5.1.1 in Brockwell and Davis [11] to periodically stationary process,

$$r_{t,i}(\boldsymbol{\beta}) = \eta_{t+i}(0; \boldsymbol{\beta}) - \boldsymbol{\eta}_{t,i}' G_{t,i}^{-1} \boldsymbol{\eta}_{t,i}. \tag{6.0.4}$$

If we multiple Y_{t+i-1} and take expectations on both sides of (6.0.3), we could obtain

$$\eta_{t+i-1}(1; \boldsymbol{\beta}) = -\sum_{j=1}^{\infty} \pi_{t+i}(j; \boldsymbol{\beta}) \eta_{t+i-j}(j-1; \boldsymbol{\beta}),$$

similarly, if we multiple $Y_{t+i-2}, Y_{t+i-3}, \ldots$ individually on both sides of (6.0.3) and take expectations, we have

$$\eta_{t+i-2}(2; \boldsymbol{\beta}) = -\sum_{j=1}^{\infty} \pi_{t+i}(j; \boldsymbol{\beta}) \eta_{t+i-j}(j-2; \boldsymbol{\beta}),$$

$$\eta_{t+i-3}(3; \boldsymbol{\beta}) = -\sum_{j=1}^{\infty} \pi_{t+i}(j; \boldsymbol{\beta}) \eta_{t+i-j}(j-3; \boldsymbol{\beta}),$$

$$\vdots$$

Therefore we could have

$$\boldsymbol{\eta}_{\infty,i} = -G_{\infty,i} \boldsymbol{\pi}_{\infty,i},$$

where

$$\boldsymbol{\eta}_{\infty,i} = \big(\eta_{t+i-1}(1; \boldsymbol{\beta}), \eta_{t+i-2}(2; \boldsymbol{\beta}), \ldots\big)',$$

$$G_{\infty,i} = \Big[\eta_{t+i-\ell}(j-\ell; \boldsymbol{\beta})\Big]_{j,\ell=1}^{\infty},$$

$$\boldsymbol{\pi}_{\infty,i} = \left(\pi_{t+i}(1;\boldsymbol{\beta}), \pi_{t+i}(2;\boldsymbol{\beta}), \ldots\right)'.$$

Additionally, $G^{-1}_{\infty,i}$ can be written as

$$G^{-1}_{i,\infty} = T_{i,\infty} T'_{i,\infty},$$

where

$$T_{i,\infty} = \left[\pi_{t+i}(k-j;\boldsymbol{\beta})\right]_{k,j=1}^{\infty},$$

$\pi_{t+i}(0;\boldsymbol{\beta}) = 1$ and $\pi_{t+i}(j;\boldsymbol{\beta}) = 0$, for $j < 0$. Furthermore, if we multiply Y_{t+i} on both sides of (6.0.3) and take expectation, we have

$$\eta_{t+i}(0;\boldsymbol{\beta}) = \boldsymbol{\pi}'_{\infty,i} G_{\infty,i} \boldsymbol{\pi}_{\infty,i} + 1. \tag{6.0.5}$$

Substituting (6.0.5) into (6.0.4), we have

$$\begin{aligned} r_{t,i}(\boldsymbol{\beta}) &= 1 + \boldsymbol{\pi}'_{\infty,i} G_{\infty,i} \boldsymbol{\pi}_{\infty,i} - \boldsymbol{\eta}'_{t,i} G^{-1}_{t,i} \boldsymbol{\eta}_{t,i} \\ &= 1 + \boldsymbol{\eta}'_{\infty,i} G^{-1}_{\infty,i} \boldsymbol{\eta}_{\infty,i} - \boldsymbol{\eta}'_{t,i} G^{-1}_{t,i} \boldsymbol{\eta}_{t,i} \end{aligned}$$

Therefore

$$\begin{aligned} \frac{\partial r_{t,i}(\boldsymbol{\beta}_0)}{\partial \beta_k} &= 2\frac{\partial \boldsymbol{\eta}'_{\infty,i}}{\partial \beta_k} G^{-1}_{\infty,i} \boldsymbol{\eta}_{\infty,i} + \boldsymbol{\eta}'_{\infty,i} G^{-1}_{\infty,i} \frac{\partial G_{\infty,i}}{\partial \beta_k} G^{-1}_{\infty,i} \boldsymbol{\eta}_{\infty,i} \\ &\quad -2\frac{\partial \boldsymbol{\eta}'_{t,i}}{\partial \beta_k} G^{-1}_{t,i} \boldsymbol{\eta}_{t,i} + \boldsymbol{\eta}'_{t,i} G^{-1}_{t,i} \frac{\partial G_{t,i}}{\partial \beta_k} G^{-1}_{t,i} \boldsymbol{\eta}_{t,i}, \end{aligned}$$

where all the terms on the right hand side are evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, and $\boldsymbol{\beta}_0$ is the truth. By

$G_{t,i}^{-1}\boldsymbol{\eta}_{t,i} = \boldsymbol{\phi}_{t,i}$, the above equation reduces to

$$\begin{aligned}\frac{\partial r_{t,i}(\boldsymbol{\beta}_0)}{\partial \beta_k} &= -2\left[\frac{\partial \boldsymbol{\eta}'_{\infty,i}}{\partial \beta_k}\boldsymbol{\pi}_{\infty,i} + \frac{\partial \boldsymbol{\eta}'_{t,i}}{\partial \beta_k}\boldsymbol{\phi}_{t,i}\right] \\ &\quad - \left[\boldsymbol{\pi}'_{\infty,i}\frac{\partial G_{\infty,i}}{\partial \beta_k}\boldsymbol{\pi}_{\infty,i} - \boldsymbol{\phi}'_{t,i}\frac{\partial G_{t,i}}{\partial \beta_k}\boldsymbol{\phi}_{t,i}\right].\end{aligned}$$

In Anderson et al. [6, Corollary 2.2.4], they show that

$$\phi_{t,k}^{(i)} \to -\pi_{t+i}(k) \; as \; k \to \infty,$$

and

$$\sum_{j=0}^{n-1}\left(\phi_{t,j}^{(i)} + \pi_{t+i}(j)\right)^2 \le \frac{2}{\pi C}\left[\left(\sum_{j\ge n}\left|\pi_{t+i}(j)\right|\right)^2 M\right],$$

where $M = \max\{\eta_{t+i}(0;\boldsymbol{\beta}),\ i = 0,\ldots,S-1\}$. Note that $\left(\sum_{j\ge n}\left|\pi_{t+i}(j)\right|\right)^2$ is bounded in n. Therefore we could have

$$\begin{aligned}\frac{\partial r_{t,i}(\boldsymbol{\beta}_0)}{\partial \beta_k} &\le 2K_1\left(\sum_{j=0}^{n-1}\left|\pi_{t+i}(j) + \phi_{t,j}^{(i)}\right| + \sum_{j\ge n}\left|\pi_{t+i}(j)\right|\right) \\ &\quad + K_1\left(\sum_{j,\ell=1}^{t}\left|\pi_{t+i}(i)\pi_{t+i}(\ell) - \phi_{t,j}^{(i)}\phi_{t,\ell}^{(i)}\right| + 2\sum_{j\ge n}\left|\pi_{t+i}(j)\right|\sum_{\ell}\left|\pi_{t+i}(\ell)\right|\right),\end{aligned}$$

where $K_1 = \frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\partial g(\lambda;\boldsymbol{\beta_0})}{\partial \beta_k}\right| d\lambda$ and $g(\lambda;\boldsymbol{\beta}_0)$ is the spectral density matrix of the equivalent vector ARMA process with length S. Note that $|A|$ is the determinant of A if A is a

matrix. By the Cauchy-Schwarz inequality,

$$\begin{aligned}\left|\frac{\partial r_{t,i}(\boldsymbol{\beta}_0)}{\partial \beta_k}\right| &\leq K_2 t^{1/2} s^{t/2} + K_3 \sum_{j \geq n} \left|\pi_{t+i}(j)\right| + K_4 t s^{t/2} \\ &\leq K_5 s_1^t,\end{aligned} \tag{6.0.6}$$

where K_2, K_3, K_4, K_5 are positive constants, $0 < s < 1$, and $0 < s_1 < 1$. This finishes the proof. □

Lemma 6.0.3. *For k = 1, . . . , p+q,*
$n^{-1/2}\left[\left|\frac{\partial}{\partial \beta_k}\sum_{t=0}^{n-1}\log r_{t,i}\right| + \left|\sum_{t=0}^{n-1}\frac{(X_{t+i}-\hat{X}_t+i)^2}{r_{t,i}^2}\frac{\partial r_{t,i}}{\partial \beta_k}\right|\right]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \xrightarrow{p} 0.$

Proof. Given $\frac{\partial r_{t,i}(\boldsymbol{\beta}_0)}{\partial \beta_k} \leq C(\boldsymbol{\beta})\{S(\boldsymbol{\beta})\}^t$ from Lemma 6.0.2 and the fact $\left|r_{t-1,i}^2\right| \geq 1$, we have

$$\begin{aligned} n^{-1/2}\left|\frac{\partial}{\partial \beta_k}\sum_{t=0}^{n-1}\log r_{t-1,i}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} &\leq n^{-1/2}\sum_{t=0}^{n-1}\frac{1}{r_{t-1,i}^2}\left|\frac{\partial r_{t-1,i}}{\partial \beta_k}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \\ &\leq n^{-1/2}\sum_{t=0}^{n-1}\left|\frac{\partial r_{t-1,i}}{\partial \beta_k}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \\ &\leq C_1 n^{-1/2}\sum_{t=0}^{n-1} s_1^t \\ &\to 0. \end{aligned}$$

Therefore

$$n^{-1/2}\left|\frac{\partial}{\partial \beta_k}\sum_{t=0}^{n-1}\log r_{t-1,i}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \xrightarrow{p} 0.$$

On the other hand,

$$
\begin{aligned}
& n^{-1/2}\mathrm{E}\left(\left|\sum_{t=0}^{n-1}\frac{(X_{t+i}-\hat{X}_{t+i})^2}{r_{t,i}^2}\frac{\partial r_{t,i}}{\partial\beta_k}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} I(A)\right) \\
\leq\ & C_1\mathrm{E}\left(\sum_{t=0}^{n-1}\frac{(X_{t+i}-\hat{X}_{t+i})^2}{r_{t,i}^2}s_1^t I(A)\right) \\
=\ & C_1\left(\sum_{t=0}^{n-1}\frac{\sigma_{t+i}^2 r_{t,i}^2}{r_{t,i}^2}s_1^t I(A)\right) \\
\leq\ & CC_1\sum_{t=0}^{n-1}s_1^t \\
\to\ & 0,
\end{aligned}
$$

where $I(\cdot)$ is the indicator function and $C>0$ is a constant, A is an event with the probability arbitrarily close to 1 and on which $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|$ arbitrarily small for all large n, where we adopt the strong consistency for MLE of a vector ARMA model from Dunsmuir and Hannan [15], and apply it for PARMA model.

$$
n^{-1/2}\left|\sum_{t=0}^{n-1}\frac{(X_{t+i}-\hat{X}_{t+i})^2}{r_{t,i}^2}\frac{\partial r_{t,i}}{\partial\beta_k}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \xrightarrow{p} 0.
$$

□

Remark 6.0.4. From Section 2 of Basawa and Lund [24], the PARMA model (1.0.1) has the $S-$variate vector ARMA representation

$$
\Phi_0\overrightarrow{X}_N-\sum_{k=1}^{p^*}\Phi_k\overrightarrow{X}_{N-k}=\Theta_0\overrightarrow{\varepsilon}_0+\sum_{k=1}^{q^*}\Theta_k\overrightarrow{\varepsilon}_{N-k}, \tag{6.0.7}
$$

where $\{\overrightarrow{X}_N\}$ and $\{\overrightarrow{\varepsilon}_N\}$ are the $S-$variate series $\{\overrightarrow{X}_N\} = \left(X_{NS+1}, \ldots, X_{NS+S}\right)'$ and $\{\overrightarrow{\varepsilon}_N\} = \left(\varepsilon_{NS+1}, \ldots, \varepsilon_{NS+S}\right)'$. The model orders in (6.0.7) are $p^* = [p/S]$ and $q^* = [q/S]$, where $[x]$ denotes the smallest integer greater than or equal to x.

Remark 6.0.5. In Dunsmuir and Hannan [15], the following assumption is needed for the strong consistency of MLE of the vector ARMA model. They restricted the elements of $\Phi_j, 1 \leq j \leq p^*$ and $\Theta_j, 1 \leq j \leq q^*$ to lie in a ball specified by

$$\mathrm{tr}\left[\sum_{1}^{\mathrm{p}^*} \Phi_{\mathrm{j}} \Phi_{\mathrm{j}}{}' + \sum_{1}^{\mathrm{q}^*} \Theta_{\mathrm{j}} \Theta_{\mathrm{j}}{}'\right] < \infty. \tag{6.0.8}$$

With condition (6.0.8), Dunsmuir and Hannan [15, Theorem 3] gave the strong consistency of the MLE of the vector ARMA model.

Remark 6.0.6. By Section 3 of Basawa and Lund [24], we could work in the univariate PARMA setting rather than transform to an $S-$variate vector ARMA model. This is eligible by two reasons. First of all, one has to invert the S-variate ARMA transformation to obtain the individual PARMA model cofficients. Hence the results derived directly in terms of the univariate PARMA model will be more readily usable. Second of all, even though one could obtain a standard vector ARMA model, the covariance matrix of the vector noises and the moving average parameters would still depend both the PARMA autoregressive parameters and the variance of vector ARMA model. For the above reasons, we will work directly in the PARMA setting. Therefore we adopt the strong consistency for MLE of a vector ARMA model from Dunsmuir and Hannan [15], and apply it for PARMA model.

Lemma 6.0.7. *For $k = 1, \ldots, p+q$,*

$$n^{-1/2} \sum_{t=0}^{n-1} \left| \frac{X_{t+i} - \hat{X}_{t+i} + Z_{t+i}}{r_{t,i}} \frac{\partial(\hat{X}_{t+i} + Z_{t+i})}{\partial \beta_k} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \xrightarrow{p} 0,$$

and

$$n^{-1/2} \sum_{t=0}^{n-1} \left| \frac{X_{t+i} - \hat{X}_{t+i} - Z_{t+i}}{r_{t,i}} \frac{\partial(\hat{X}_{t+i} - Z_{t+i})}{\partial \beta_k} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \xrightarrow{p} 0.$$

Proof. We only prove $n^{-1/2} \sum_{t=0}^{n-1} \left| \frac{X_{t+i} - \hat{X}_{t+i} - Z_{t+i}}{r_{t,i}} \frac{\partial(\hat{X}_{t+i} - Z_{t+i})}{\partial \beta_k} \right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \xrightarrow{p} 0$, since the other result could be proved in a similar way. By the invertible assumption of $\text{PARMA}_S(p,q)$ model,

$$Z_t = \sum_{j=0}^{\infty} \pi_t(j) X_{t-j},$$

and from the model set-up in (1.0.1), we can write

$$\boldsymbol{\phi}_t(z) = 1 - \phi_t(1)z - \ldots - \phi_t(p)z^p,$$

and

$$\boldsymbol{\theta}_t(z) = 1 + \theta_t(1)z + \ldots + \theta_t(q)z^q.$$

Additionally,

$$\boldsymbol{\pi}_t(z) = \frac{\boldsymbol{\phi}_t(z)}{\boldsymbol{\theta}_t(z)} = 1 + \sum_{j=1}^{\infty} \pi_j z^j.$$

Notice that $\boldsymbol{\beta} = (\phi_t(1), \phi_t(2), \ldots, \phi_t(p), \theta_t(1), \theta_t(2), \ldots, \theta_t(q))'$, and when $k = 1, \ldots, p$, we

have

$$\frac{z^k}{\boldsymbol{\theta}_t(z)} = \frac{-\partial \boldsymbol{\phi}_t(z)/\partial \beta_k}{\boldsymbol{\theta}_t(z)} = -\frac{\partial \left(\boldsymbol{\phi}_t(z)/\boldsymbol{\theta}_t(z)\right)}{\partial \beta_k} = -\frac{\partial \boldsymbol{\pi}_t(z)}{\partial \beta_k} = -\sum_{j=1}^{\infty} \frac{\partial \pi_t(j)}{\partial \beta_k} z^j,$$

and when $k = 1, 2, \ldots, q$, we have

$$\frac{\partial \left(\boldsymbol{\pi}_t(z)\right)}{\partial \beta_{p+k}} = \sum_{j=1}^{\infty} \frac{\partial \pi_t(j)}{\partial \beta_{p+k}} z^j = \frac{\partial \left(\boldsymbol{\phi}_t(z)/\boldsymbol{\theta}_t(z)\right)}{\partial \beta_{p+k}} = -\frac{\boldsymbol{\phi}_t(z)}{\boldsymbol{\theta}_t^2(z)} \frac{\partial \boldsymbol{\theta}_t(z)}{\partial \beta_{p+k}} = -\frac{\boldsymbol{\phi}_t(z)}{\boldsymbol{\theta}_t^2(z)} z^k.$$

In summary,

$$\begin{cases} \dfrac{z^k}{\boldsymbol{\theta}_t(z)} = -\sum_{j=1}^{\infty} \dfrac{\partial \pi_t(j)}{\partial \beta_k} z^j \\ \dfrac{\boldsymbol{\phi}_t(z)}{\boldsymbol{\theta}_t^2(z)} z^k = -\sum_{j=1}^{\infty} \dfrac{\partial \pi_t(j)}{\partial \beta_{p+k}} z^j. \end{cases} \tag{6.0.9}$$

Then we can prove from (6.0.9) that there exist constants $C > 0$ and $s \in (0, 1)$ such that

$$\left| \frac{\partial \pi_j}{\partial \beta_k} \right| \leq C s^j, \ j \geq 1, \ 1 \leq k \leq p + q.$$

Furthermore, we may expand the Yule-Walker equation $\Gamma_{n,i} \boldsymbol{\phi}_n^{(i)} = \boldsymbol{\gamma}_n^{(i)}$ as

$$\begin{pmatrix} \gamma_{i+n-1}(1) \\ \gamma_{i+n-2}(2) \\ \gamma_{i+n-3}(3) \\ \vdots \\ \gamma_i(n) \end{pmatrix} = \begin{pmatrix} \gamma_{i+n-1}(0) & \gamma_{i+n-1}(-1) & \cdots & \gamma_{i+n-1}(-n+1) \\ \gamma_{i+n-2}(1) & \gamma_{i+n-2}(0) & \cdots & \gamma_{i+n-2}(-n+2) \\ \gamma_{i+n-3}(2) & \gamma_{i+n-3}(1) & \cdots & \gamma_{i+n-3}(-n+3) \\ \vdots & \vdots & \ldots & \vdots \\ \gamma_i(n-1) & \gamma_i(n-2) & \ldots & \gamma_i(0) \end{pmatrix} \begin{pmatrix} \phi_{n,1}^{(i)} \\ \phi_{n,2}^{(i)} \\ \phi_{n,3}^{(i)} \\ \vdots \\ \phi_{n,n}^{(i)} \end{pmatrix},$$

then for $1 \leq j \leq n$, we have

$$\gamma_{i+n-j}(j) = \sum_{\ell=1}^{n} \phi_{n,\ell}^{(i)} \gamma_{i+n-j}(j-\ell),$$

therefore for $1 \leq k \leq p+q$,

$$\frac{\partial \gamma_{i+n-j}(j)}{\partial \beta_k} = \sum_{\ell=1}^{n} \frac{\partial \gamma_{i+n-j}(j-\ell)}{\partial \beta_k} \phi_{n,\ell}^{(i)} + \sum_{\ell=1}^{n} \frac{\partial \phi_{n,\ell}^{(i)}}{\partial \beta_k} \gamma_{i+n-j}(j-\ell). \tag{6.0.10}$$

On the other hand,

$$Z_{n+i} = X_{n+i} + \sum_{\ell=1}^{\infty} \pi_{n+i}(\ell) X_{n+i-\ell},$$

if we multiply X_{n+i-j} on both sides and take expectations, we can obtain

$$\gamma_{n+i-j}(j) = \sum_{\ell=1}^{\infty} \pi_{n+i}(\ell) \gamma_{n+i-j}(j-\ell),$$

then similarly for $1 \leq k \leq p+q$,

$$\frac{\partial \gamma_{n+i-j}(j)}{\partial \beta_k} = \sum_{\ell=1}^{\infty} \frac{\partial \pi_{n+i}(\ell)}{\partial \beta_k} \gamma_{n+i-j}(j-\ell) + \sum_{\ell=1}^{\infty} \pi_{n+i}(\ell) \frac{\partial \gamma_{n+i-j}(j-\ell)}{\partial \beta_k}. \tag{6.0.11}$$

If we subtract (6.0.10) from (6.0.11), we have

$$\begin{aligned} 0 \;=\; & \sum_{\ell=1}^{n} \left(\pi_{i+n}(\ell) - \phi_{n,\ell}^{(i)} \right) \frac{\partial \gamma_{i+n-j}(j-\ell)}{\partial \beta_k} + \sum_{\ell>n} \pi_{i+n}(\ell) \frac{\partial \gamma_{i+n-j}(j-\ell)}{\partial \beta_k} \qquad (6.0.12) \\ & \sum_{\ell=1}^{n} \left(\frac{\partial \pi_{i+n}(\ell)}{\partial \beta_k} - \frac{\partial \phi_{n,\ell}^{(i)}}{\partial \beta_k} \right) \gamma_{i+n-j}(j-\ell) + \sum_{\ell>n} \frac{\partial \pi_{i+n}(\ell)}{\partial \beta_k} \gamma_{i+n-j}(j-\ell). \end{aligned}$$

Furthermore, we may write (6.0.12) as

$$\begin{aligned}
&\sum_{\ell=1}^{n}\left(\frac{\partial\phi_{n,\ell}^{(i)}}{\partial\beta_k}-\frac{\partial\pi_{i+n}(\ell)}{\partial\beta_k}\right)\gamma_{i+n-j}(j-\ell)\\
&=\sum_{\ell=1}^{n}\left(\pi_{i+n}(\ell)-\phi_{n,\ell}^{(i)}\right)\frac{\partial\gamma_{i+n-j}(j-\ell)}{\partial\beta_k}\\
&+\sum_{\ell>n}\left(\pi_{i+n}(\ell)\frac{\partial\gamma_{i+n-j}(j-\ell)}{\partial\beta_k}+\frac{\partial\pi_{i+n}(\ell)}{\partial\beta_k}\gamma_{i+n-j}(j-\ell)\right).
\end{aligned}\tag{6.0.13}$$

Consequently, we write (6.0.13) as

$$\frac{\partial\left(\boldsymbol{\pi}_{n,i}-\boldsymbol{\phi}_{n,i}\right)}{\partial\beta_k}=\Gamma_{n,i}^{-1}\left[\frac{\partial\Gamma_{n,i}}{\partial\beta_k}(\boldsymbol{\phi}_{n,i}-\boldsymbol{\pi}_{n,i})+\boldsymbol{d}_{n,i}\right],$$

where $\boldsymbol{d}_{n,i}$ is an $n\times 1$ vector with

$$\sum_{\ell>n}\left(\pi_{i+n}(\ell)\frac{\partial\gamma_{i+n-j}(j-\ell)}{\partial\beta_k}+\frac{\partial\pi_{i+n}(\ell)}{\partial\beta_k}\gamma_{i+n-j}(j-\ell)\right)$$

as its j-th component. Then under the assumption of $Z_{-t}=X_{-t}=0$ for all $t\geq 0$ we have

$$\begin{aligned}
Z_{t+i} &= \theta(B)^{-1}\phi(B)X_{t+i}\\
&= \pi(B)X_{t+i}\\
&= X_{t+i}+\sum_{j=1}^{\infty}X_{t+i-j}\pi_{t+i}(j)\\
&= X_{t+i}+\sum_{j=0}^{n-1}X_{t+i-j}\pi_{t+i}(j).
\end{aligned}$$

Now it follows from (6.0.13) that

$$
\begin{aligned}
&\mathrm{E}\left(\frac{\partial\left(\hat{X}_{t+i}^{(i)}-Z_{t+i}\right)}{\partial\beta_k}\right)^2\\
=&\mathrm{E}\left(\frac{\partial\left(\phi_{t,1}^{(i)}X_{t+i-1}+\ldots+\phi_{t,t}^{(i)}X_i-X_{t+i}-\sum_{j=0}^{n-1}X_{t+i-j}\pi_{t+i}(j)\right)}{\partial\beta_k}\right)^2\\
=&\mathrm{E}\left(\sum_{j=0}^{n-1}\frac{\partial\left(\phi_{t,j}^{(i)}-\pi_{t+i}(j)\right)}{\partial\beta_k}X_{t+i-j}\right)^2\\
=&\frac{\partial\left(\boldsymbol{\phi}_{t,i}-\boldsymbol{\pi}_{t,i}\right)'}{\partial\beta_k}\Gamma_{t,i}\frac{\partial\left(\boldsymbol{\phi}_{t,i}-\boldsymbol{\pi}_{t,i}\right)}{\partial\beta_k}\\
=&\left[\frac{\partial\Gamma_{t,i}}{\partial\beta_k}\left(\boldsymbol{\phi}_{t,i}-\boldsymbol{\pi}_{t,i}\right)+\boldsymbol{d}_{n,i}\right]'\Gamma_{t,i}^{-1}\left[\frac{\partial\Gamma_{t,i}}{\partial\beta_k}\left(\boldsymbol{\phi}_{t,i}-\boldsymbol{\pi}_{t,i}\right)+\boldsymbol{d}_{t,i}\right]\\
\leq&\frac{2}{\lambda_{\min}(\Gamma_{t,i})}\left[\|\frac{\partial\Gamma_{t,i}}{\partial\beta_k}\left(\boldsymbol{\phi}_{t,i}-\boldsymbol{\pi}_{t,i}\right)\|^2+\|\boldsymbol{d}_{t,i}\|^2\right]\\
\leq&\frac{2\max\{\alpha,\gamma_i(0)\}t}{\lambda_{\min}(\Gamma_{t,i})}\left[\|\boldsymbol{\phi}_{t,i}-\boldsymbol{\pi}_{t,i}\|^2+\left(\sum_{\ell>t}\left|\pi_{t+i}(\ell)\right|+\left|\frac{\partial\pi_{t+i}(\ell)}{\partial\beta_k}\right|\right)^2\right]\\
\leq&C_1 s_1^t,
\end{aligned}
$$

where $\lambda_{\min}(\Gamma_{t,i})>0$ denotes the minimum eigenvalue of $\Gamma_{t,i}$, and

$$
\alpha\triangleq\left|\int_{-\pi}^{\pi}\frac{\partial\left|\theta(e^{iw})/\phi(e^{iw})\right|^2}{\partial\beta_k}dw\right|\geq\left|\frac{\partial\gamma_i(j)}{\partial\beta_k}\right|,
$$

for $j\geq 1$ and $1\leq k\leq p+q$, and $C_1>0$ and $s_1\in(0,1)$ are some constants, which depend on $\beta\in B$ continuously. Additionally, a detailed discussion of spectral density for PARMA

model could be seen in Wyłomańska [56]. Next using the similar argument as in the proof of Lemma 6.0.3, we may show that

$$\mathrm{E}\left[\frac{\partial(\hat{X}_{t+i}-Z_{t+i})}{\partial\beta_k}I(A)\right]^2_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} \leq C_2 s_2^t,\ t\geq 1,\ 1\leq k\leq p+q,$$

where A is an event with the probability arbitrarily close to 1 and on which $\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|$ arbitrarily small for all large n, and $C_2>0$ and $S_2\in(0,1)$ are some constants. Hence

$$\begin{aligned}
& n^{-1/2}\textstyle\sum_{t=0}^{n-1}\left(\mathrm{E}\left|\frac{(X_{t+i}-\hat{X}_{t+i}-Z_{t+i})}{r_{t,i}}\frac{\partial(\hat{X}_{t+i}-Z_{t+i})}{\partial\beta_k}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} I(A)\right) \\
\leq\ & n^{-1/2}\textstyle\sum_{t=0}^{n-1}\left(\mathrm{E}\left[X_{t+i}-\hat{X}_{t+i}(\boldsymbol{\beta})-Z_{t+i}(\boldsymbol{\beta})\right]^2\mathrm{E}\left[\frac{\partial(\hat{\mathrm{X}}_{\mathrm{t+i}}-\mathrm{Z}_{\mathrm{t+i}})}{\partial\beta_{\mathrm{k}}}\right]^2_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} I(A)\right)^{1/2} \\
\leq\ & Cn^{-1/2}\textstyle\sum_{t=0}^{n-1} s_2^{t/2} \\
\to\ & 0,
\end{aligned}$$

where $C>0$ is a constant. Thus

$$n^{-1/2}\sum_{t=0}^{n-1}\left|\frac{(X_{t+i}-\hat{X}_{t+i}-Z_{t+i})}{r_{t,i}}\frac{\partial(\hat{X}_{t+i}-Z_{t+i})}{\partial\beta_k}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\to 0.$$

In a similar manner, we could prove

$$n^{-1/2}\sum_{t=0}^{n-1}\left|\frac{X_{t+i}-\hat{X}_{t+i}+Z_{t+i}}{r_{t,i}}\frac{\partial(\hat{X}_{t+i}+Z_{t+i})}{\partial\beta_k}\right|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}\xrightarrow{p}0,$$

and this ends the proof. □

We will show that the asymptotic variance of the estimated periodic AR and MA coeffi-

cient vector can be nicely represented in terms of the two PAR models. Define

$$\phi_t(B)\xi_t = W_t,$$

and

$$\theta_t(B)\varsigma_t = W_t,$$

where $\{W_t\} \sim \text{WN}(0,1)$ is a white noise process with mean 0 and variance 1. Let $\boldsymbol{\xi} = \left(\xi_{-1}, \xi_{-2}, \ldots, \xi_{-p}, \varsigma_{-1}, \ldots, \varsigma_{-q}\right)'$, and

$$\mathbf{W}(\boldsymbol{\beta}) = \mathbf{W}(\boldsymbol{\phi_t}, \boldsymbol{\theta_t}) = \{\text{Var}(\boldsymbol{\xi})\}^{-1}.$$

Before we state our next lemma, we need to clarify notation. By the invertible assumption of PARMA model (1.0.1), we may write

$$\varepsilon_{t+i} = \varepsilon_{t+i}(\boldsymbol{\beta}) = X_{t+i} - \phi_t(i)X_{t+i-1} - \ldots \phi_t(p)X_{t+i-p} - \theta_t(1)\varepsilon_{t+i-1} \ldots - \theta_t(q)\varepsilon_{t+i-q}.$$

Then for $k = 1, \ldots .p$, we may write

$$U_{tk}^{(i)} = -\frac{\partial \varepsilon_{t+i}}{\partial \beta_k},$$

and for $k = 1, 2, \ldots, q$,

$$V_{tk}^{(i)} = -\frac{\partial \varepsilon_{t+i}}{\partial \beta_{p+k}}.$$

Let $\chi = (\boldsymbol{X}, \mathcal{Z})$ and $\tau = (\boldsymbol{U}, \boldsymbol{V})$, where $\boldsymbol{X}$ and $\boldsymbol{U}$ are $n \times p$ matrices with $X_{i+j-\ell}$ and $U_{j\ell}^{(i)}$ as their (j,ℓ)-th elements, respectively, and $\mathcal{Z}$ and $\boldsymbol{V}$ are $n \times q$ matrices with

$\varepsilon_{i+j-\ell}$ and $V_{j\ell}^{(i)}$ as their (j,ℓ)-th elements respectively. Denote R as the diagonal matrix $\text{diag}\left(r_{0,i}(\boldsymbol{\beta_0}),\ldots,r_{n-1,i}(\boldsymbol{\beta_0})\right)$.

Lemma 6.0.8. $n^{-1}\tau' R^{-1}\tau \xrightarrow{P} \mathbf{W}^*(\boldsymbol{\beta}) = \{\text{Var}(\boldsymbol{\sigma}_t\boldsymbol{\xi}_t)\}^{-1}$.

Proof. In the following proof, all $U_{tk}^{(i)}$,$V_{tk}^{(i)}$,Z_{t+i}, and $r_{t,i}$ are evaluated at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. We adopt the notations that

$$B^k U_{tj}^{(i)} = U_{t-k,j}^{(i)},$$

and

$$B^k V_{tj}^{(i)} = V_{t-k,j}^{(i)}.$$

From the invertible representation, we have

$$U_{tj}^{(i)} = \theta_t^{-1}(B)X_{t+i-j},$$

and

$$V_{tj}^{(i)} = \theta_t^{-1}(B)Z_{t+i-j} = \phi_t(B)\theta_t^{-2}(B)X_{t+i-j},$$

assuming that $X_{-t} = Z_{-t} = 0$ for all $t \geq 0$. Let

$$\begin{cases} 1/\theta_t(z) = 1 - \sum_{j\geq k}\psi_t(j)z^j \\ \phi_t(z)/\theta_t^2(z) = 1 - \sum_{j\geq 1}\eta_j z^j, \end{cases}$$

then

$$\begin{aligned} U_{tj}^{(i)} =& X_{t+i-j} - \sum_{k=1}^{t+i-j-1} \psi_t(j) X_{t+i-j-k} \\ =& \phi_t(B)^{-1} Z_{t+i-j} + \sum_{k \geq t+i-j} \psi_t(k) X_{t+i-j-k} \\ =& \tilde{U_{tj}^{(i)}} + u_{tj}^{(i)}, \end{aligned} \tag{6.0.14}$$

and

$$\begin{aligned} V_{tj}^{(i)} =& X_{t+i-j} - \sum_{k=1}^{t+i-j-1} \eta_k X_{t+i-j-k} \\ =& \theta_t(B)^{-1} Z_{t+i-j} + \sum_{k \geq t+i-j} \eta_k X_{t+i-j-k} \\ =& \tilde{V_{tj}^{(i)}} + v_{tj}^{(i)}. \end{aligned} \tag{6.0.15}$$

Note that

$$\begin{aligned} \mathrm{E}\left(u_{tj}^{(i)}\right)^2 &= \mathrm{E}\left(\sum_{k \geq t+i-j}^{\infty} \psi_t(k) X_{t+i-j-k}\right)^2 \\ &\leq \sum_{\ell=1, k \geq t+i-j}^{\infty} \gamma_{t+i-j-k}(k-\ell) \psi_t(k) \psi_t(\ell) \\ &\leq M \sum_{k=1}^{\infty} \psi_{t+i-j-k}^2 \\ &\leq C s^{t-j}, \end{aligned} \tag{6.0.16}$$

for $t \geq 1$, and $1 \leq j \leq p$, where $0 < s < 1$ and $M = \max\{\gamma_i(0) : i = 0, 1, \ldots, S-1\}$.

Similarly, we could have

$$\mathrm{E}\left(v_{tj}^{(i)}\right)^2 \leq Cs^{t-j},$$

for $t \geq 1$, and $1 \leq j \leq q$. Notice that the $(j,\ell)-$th element of $n^{-1}\tau' R^{-1}\tau$ is

$$\frac{1}{n}\sum_{t=0}^{n-1} U_{tj}^{(i)}U_{t\ell}^{(i)}/r_{t,i} = \frac{1}{n}\sum_{t=0}^{n-1}\left(\tilde{U}_{tj}^{(i)}\tilde{U}_{t\ell}^{(i)} + \tilde{U}_{tj}^{(i)}u_{t\ell}^{(i)} + u_{tj}^{(i)}\tilde{U}_{t\ell}^{(i)} + u_{tj}^{(i)}u_{t\ell}^{(i)}\right)/r_{t,i}.$$

By the ergodic theorem, under the assumption of a measurable density function for PARMA process,

$$\begin{aligned}
\frac{1}{n}\sum_{t=0}^{n-1}\tilde{U}_{tj}^{(i)}\tilde{U}_{t\ell}^{(i)} &= \frac{1}{n}\sum_{t=0}^{n-1}[\phi_t^{-1}(B)\varepsilon_{t+i-j}][\phi_t^{-1}(B)\varepsilon_{t+i-\ell}] \\
&\overset{a.s.}{\to} \mathrm{Cov}(\tilde{U}_{tj}^{(i)}\tilde{U}_{t\ell}^{(i)}) \\
&= \mathrm{Cov}\left(\phi_t^{-1}(B)Z_{t+i-j}, \phi_t^{-1}(B)Z_{t+i-\ell}\right) \\
&= (\sigma_{t+i-j}\sigma_{t+i-\ell})\mathrm{Cov}(\xi_{t+i-j}, \xi_{t+i-\ell}).
\end{aligned}$$

By Corollary 5.1.2 $|r_{t,i} - 1| \to 0$ as $t \to \infty$, we have

$$\frac{1}{n}\sum_{t=0}^{n-1}\tilde{U}_{tj}^{(i)}\tilde{U}_{t\ell}^{(i)}/r_{t,i} \overset{a.s.}{\to} (\sigma_{t+i-j}\sigma_{t+i-\ell})\mathrm{Cov}(\xi_{t+i-j}, \xi_{t+i-\ell}).$$

By Cauchy-Schwartz inequality and (6.0.16),

$$\begin{aligned}\frac{1}{n}\sum_{t=0}^{n-1}\mathrm{E}\left|u_{tj}^{(i)}u_{t\ell}^{(i)}\right|/r_{t,i} &\le \frac{1}{n}\sum_{t=0}^{n-1}\mathrm{E}\left|u_{tj}^{(i)}u_{t\ell}^{(i)}\right| \\ &\le \frac{1}{n}\sum_{t=0}^{n-1}\left[\mathrm{E}\left(u_{tj}^{(i)}\right)^2\mathrm{E}\left(u_{t\ell}^{(i)}\right)^2\right]^{1/2} \\ &\le \frac{C}{n}\sum_{t=0}^{n-1}s^{2t-j-\ell} \\ &\to 0.\end{aligned} \tag{6.0.17}$$

Then

$$\frac{1}{n}\sum_{t=0}^{n-1}u_{tj}^{(i)}u_{t\ell}^{(i)}/r_{t,i}\xrightarrow{p}0.$$

With a similar method, we may also show that

$$\frac{1}{n}\sum_{t=0}^{n-1}\tilde{U_{tj}^{(i)}}u_{t\ell}^{(i)}/r_{t,i}\xrightarrow{p}0,$$

and

$$\frac{1}{n}\sum_{t=0}^{n-1}u_{tj}^{(i)}\tilde{U_{t\ell}^{(i)}}/r_{t,i}\xrightarrow{p}0.$$

Therefore for $1\le j,\ell\le p$,

$$\frac{1}{n}\sum_{t=0}^{n-1}U_{tj}^{(i)}U_{t\ell}^{(i)}/r_{t,i}\xrightarrow{p}(\sigma_{t+i-j}\sigma_{t+i-\ell})\mathrm{Cov}(\xi_{t+i-j},\xi_{t+i-\ell}).$$

We can prove in a similar manner that for $1\le j\le p$ and $1\le\ell\le q$,

$$\frac{1}{n}\sum_{t=0}^{n-1}U_{tj}^{(i)}V_{t\ell}^{(i)}/r_{t,i}\xrightarrow{p}(\sigma_{t+i-j}\sigma_{t+i-\ell})\mathrm{Cov}(\xi_{t+i-j},\varsigma_{t+i-\ell}),$$

and $1 \leq j, \ell \leq q$,

$$\frac{1}{n}\sum_{t=0}^{n-1} V_{tj}^{(i)} V_{t\ell}^{(i)} / r_{t,i} \xrightarrow{p} (\sigma_{t+i-j}\sigma_{t+i-\ell})\mathrm{Cov}(\varsigma_{t+i-j}, \varsigma_{t+i-\ell}).$$

Putting the above three equations together, we have completed the proof. □

Lemma 6.0.9. *Let $\tau = (\boldsymbol{U}, \boldsymbol{V})$, $\boldsymbol{U}$ is an $n \times p$ matrix with $U_{j\ell}^{(i)}$ as its (j,ℓ)-th elements, respectively, and $\boldsymbol{V}$ is an $n \times q$ matrix with $V_{j\ell}^{(i)}$ as their (j,ℓ)-th elements. Both $\boldsymbol{U}$ and $\boldsymbol{V}$ follow the assumption of second finite moment. R denotes the diagonal matrix* $\mathrm{diag}\left(r_{0,i}(\boldsymbol{\beta_0}), \ldots, r_{n-1,i}(\boldsymbol{\beta_0})\right)$, *and*

$$\mathcal{Z} = \left(Z_{t+i}(\boldsymbol{\beta_0}), Z_{t+i+1}(\boldsymbol{\beta_0}) \ldots, Z_{t+n-1}(\boldsymbol{\beta_0})\right)',$$

then

$$n^{-1/2}\tau' R^{-1}\mathcal{Z} \xrightarrow{D} N\left(0, \sigma_t^2 W^*(\boldsymbol{\beta}_0)^{-1}\right).$$

Proof. Define

$$\tilde{U}_t^{(i)} = \left(\tilde{U}_{t1}^{(i)}, \ldots, \tilde{U}_{tp}^{(i)}, \tilde{V}_{t1}^{(i)}, \ldots, \tilde{V}_{tp}^{(i)}\right)'$$

and

$$u_t^{(i)} = \left(u_{t1}^{(i)}, \ldots, u_{tp}^{(i)}, v_{t1}^{(i)}, \ldots, v_{tq}^{(i)}\right)',$$

where $\tilde{U}_{tj}^{(i)}, \tilde{V}_{tj}^{(i)}, u_{tj}^{(i)}$ and $v_{tj}^{(i)}$ are defined in (6.0.14) and (6.0.15). From the model invertible assumption, we may write $Z_t = \varepsilon_t + z_t$, where $z_t = -\sum_{j \geq t} \pi_t(j) X_{t-j}$, and $\boldsymbol{\pi}_t(z) = \frac{\boldsymbol{\phi}_t(z)}{\boldsymbol{\theta}_t(z)} =$

$1+\sum_{j=1}^{\infty}\pi_j z^j$. then

$$
\begin{aligned}
n^{-1/2}{}_{\tau}{}' R^{-1}\mathcal{Z} &= n^{-1/2}\sum_{t=0}^{n-1}\left(\tilde{U}_t^{(i)}+u_t^{(i)}\right)\frac{\varepsilon_t+z_t}{r_{t,i}} \\
&= n^{-1/2}\sum_{t=0}^{n-1}\frac{\tilde{U}_t^{(i)}\varepsilon_t+\tilde{U}_t^{(i)}z_t+u_t^{(i)}\varepsilon_t+u_t^{(i)}z_t}{r_{t,i}}.
\end{aligned}
\tag{6.0.18}
$$

Using similar method as (6.0.16), we may obtain that for all $t \geq 1$,

$$
\mathrm{E}\left[z_t^2\right] \leq Cs^t,
$$

where $C > 0$, and $s \in (0,1)$ are some constants. Additionally based on the same argument as (6.0.17), we may obtain that

$$
n^{-1/2}\sum_{t=0}^{n-1}\frac{\tilde{U}_t^{(i)}z_t+u_t^{(i)}\varepsilon_t+u_t^{(i)}z_t}{r_{t,i}} \xrightarrow{P} 0. \tag{6.0.19}
$$

Let $F_t^{(i)}$ be the $\sigma-$algebra generated by $\{\varepsilon_{t+i-k}, k \geq 0\}$, then $\{\alpha' \tilde{U}_t^{(i)}\varepsilon_t/r_{t,i}\}$ are martingale differences with respect to $\{F_t^{(i)}\}$, for any $\alpha \in R^{p+q}$. Furthermore, $\alpha' \tilde{U}_t^{(i)}\varepsilon_t/r_{t,i}$ is $F_t^{(i)}-$measurable and

$$
\mathrm{E}\left[\left(\alpha' \tilde{U}_t^{(i)}\varepsilon_t/r_{t,i}\right)|F_{t-1}^{(i)}\right] = \left(\alpha' \tilde{U}_t^{(i)}/r_{t,i}\right)\mathrm{E}\varepsilon_t = 0.
$$

Further for any $\varepsilon > 0$,

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=0}^{n-1} \mathrm{E}\left[\left(\alpha' \tilde{U}_t^{(i)} \varepsilon_t / r_{t,i}\right)^2 I\left(\left|\alpha' \tilde{U}_t^{(i)} \varepsilon_t / r_{t,i}\right|\right) | F_{t-1}^{(i)}\right] \\
\leq &\frac{1}{n}\sum_{t=0}^{n-1} \mathrm{E}\Bigg[\left(\alpha' \tilde{U}_t^{(i)} \varepsilon_t\right)^2 I\left(\left|\alpha' \tilde{U}_t^{(i)} \varepsilon_t\right| > n^{1/2}\varepsilon\right) \\
&\left(I\left(\left|\alpha' \tilde{U}_t^{(i)}\right| > \log n\right) + I\left(\left|\alpha' \tilde{U}_t^{(i)}\right| \leq \log n\right)\right) | F_{t-1}^{(i)}\Bigg] \\
\leq &\frac{1}{n}\sum_{t=0}^{n-1} \left[\sigma_t^2 \left(\alpha' \tilde{U}_t^{(i)}\right)^2 I\left(\left|\alpha' \tilde{U}_t^{(i)}\right| > \log n\right) + \left(\alpha' \tilde{U}_t^{(i)}\right)^2 \mathrm{E}\left[\varepsilon_t^2 I\left(|\varepsilon_t| > \frac{n^{1/2}\varepsilon}{\log n}\right)\right]\right] \\
\sim &\sigma_t^2 \mathrm{E}\left[\left(\alpha' \tilde{U}_1^{(i)}\right)^2 I\left(\left|\alpha' \tilde{U}_1^{(i)}\right| > \log n\right)\right] + \mathrm{E}\left(\alpha' \tilde{U}_1^{(i)}\right)^2 \mathrm{E}\left[\varepsilon_1^2 I\left(|\varepsilon_1| > \frac{n^{1/2}\varepsilon}{\log n}\right)\right] \\
\to &0.
\end{aligned}
$$

The last limit follows from the fact that both ε_t and $\alpha' \tilde{U}_1^{(i)}$ have finite second moments.

Note that since $r_{t,i} \to 1$ as $t \to \infty$,

$$
\begin{aligned}
\frac{1}{n}\sum_{t=0}^{n-1} \left(\alpha' \tilde{U}_t^{(i)} \varepsilon_t / r_{t,i}\right)^2 &\sim \frac{1}{n}\sum_{t=0}^{n-1} \left(\alpha' \tilde{U}_t^{(i)} \varepsilon_t\right)^2 \\
&\overset{a.s.}{\to} \mathrm{E}\left(\alpha' \tilde{U}_t^{(i)} \varepsilon_t\right)^2 \\
&= \sigma_t^2 \mathrm{E}\left(\alpha' \tilde{U}_t^{(i)}\right)^2 \\
&= \sigma_t^4 \alpha' W(\boldsymbol{\beta}_0)^{-1} \alpha \\
&= \sigma_t^2 \alpha' W^*(\boldsymbol{\beta}_0)^{-1} \alpha.
\end{aligned}
$$

Then it follows from Theorem 4 of p. 511 of Shiryaev [40] that

$$n^{-1/2}\sum_{t=0}^{n-1}\left(\alpha'\tilde{U}_t^{(i)}\varepsilon_t/r_{t,i}\right)\xrightarrow{d}N\left(0,\sigma_t^2\alpha' W^*(\boldsymbol{\beta}_0)^{-1}\alpha\right), \tag{6.0.20}$$

for any $\alpha\in R^{p+q}$. Now the limit

$$n^{-1/2}\sum_{t=0}^{n-1}\tau' R^{-1}\mathcal{Z}\xrightarrow{D}N\left(0,\sigma_t^2 W^*(\boldsymbol{\beta}_0)^{-1}\right)$$

follows from (6.0.20), (6.0.18) and (6.0.19). □

Theorem 6.0.10. *Let X_t be the $PARMA_S(p,q)$ process defined by* (1.0.1)*, and suppose that the vector of true parameter values $\boldsymbol{\beta}_0\in\mathcal{B}$, where $\mathcal{B}$ is the parameter space containing all $\boldsymbol{\beta}$. Then as $n\to\infty$,*

$$n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)\xrightarrow{D}N\left(0,W^*(\boldsymbol{\beta}_0)\right).$$

Proof. Let $\hat{X_n}=\left(\hat{X}_i,\ldots,\hat{X}_{i+n-1}\right)$, then $X_n=H^{(i)}\left(X_n-\hat{X_n}\right)$, where $H^{(i)}$ is given as

$$H^{(i)}=\begin{pmatrix}1 & 0 & 0 & \ldots & 0\\ h_{11}^{(i)} & 1 & 0 & \ldots & 0\\ h_{22}^{(i)} & h_{21}^{(i)} & 1 & \ldots & 0\\ h_{33}^{(i)} & h_{32}^{(i)} & h_{31}^{(i)} & \ldots & 0\\ h_{n-1,n-1}^{(i)} & h_{n-1,n-2}^{(i)} & h_{n-1,n-3}^{(i)} & \ldots & 1\end{pmatrix},$$

where the coefficients $h_{22}^{(i)}$ depend on i. Notice that Consequently,

$$X_n^{'}\Sigma(\beta)^{-1}X_n = \sum_{t=0}^{n-1} \frac{\left(X_{t+i} - \hat{X}_{t+i}^{(i)}\right)^2}{\sigma_t^2 r_{t-1,i}},$$

where $|\Sigma(\beta)| = \left(\prod_{i=0}^{S-1} \sigma_i^2\right)^N$, with $N = n/S$. Now it follows from (6.0.1) that

$$\ell(\boldsymbol{\beta}) = -2\log L(\boldsymbol{\beta}) = n\log(2\pi) + n + \sum_{j=0}^{n-1} \log(r_{j,i}) + N\sum_{i=0}^{S-1} \log(S_i/N),$$

where

$$S_i = \sum_{k=0}^{N-1} \frac{\left(X_{kS+i} - \hat{X}_{kS+i}^{(i)}\right)^2}{r_{kS+i}}.$$

Note that $\hat{\boldsymbol{\beta}}$ is the solution of the equation $\frac{\partial}{\partial\beta}\ell(\boldsymbol{\beta}) = 0$, and for $1 \le k \le p$, the equality $\frac{\partial}{\partial\beta}\ell(\boldsymbol{\beta})|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0$ leads to

$$\begin{aligned} 0 &= \sum_{t=0}^{n-1} \frac{Z_t(\hat{\boldsymbol{\beta}})U_{tk}^{(i)}(\hat{\boldsymbol{\beta}})}{r_{t,i}(\hat{\boldsymbol{\beta}})} + \delta_k^{(i)} \qquad (6.0.21) \\ &= \sum_{t=0}^{n-1} \left[X_t - \sum_{j=1}^{p} \hat{\boldsymbol{\beta}}_t(j)X_{t-j} - \sum_{i=1}^{q} \hat{\boldsymbol{\beta}}_t(p+i)Z_{t-i}(\boldsymbol{\beta}_0)\right] \frac{\hat{U}_{tk}^{(i)}(\boldsymbol{\beta}_0)}{r_{t,i}} + \boldsymbol{\eta}_k^{'}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \delta_k^{(i)}, \end{aligned}$$

where $X_{-j} = Z_{-j} = 0$ for all $j \geq 0$, and

$$\begin{aligned}\delta_k^{(i)} =& \left(\frac{\sigma_t^2}{2} \frac{\partial}{\partial \beta_k} \sum_{t=0}^{n-1} \log r_{t,i} - \frac{1}{2} \sum_{t=0}^{n-1} \frac{(X_t - \hat{X}_t)^2}{r_{t,i}^2} \frac{\partial r_{t,i}}{\partial \beta_k} \right. \\ & \left. - \frac{1}{2} \sum_{t=0}^{n-1} \left[\frac{X_t - \hat{X}_t + Z_t}{r_{t,i}} \frac{\partial (\hat{X}_t + Z_t)}{\partial \beta_k} + \frac{X_t - \hat{X}_t - Z_t}{r_{t,i}} \frac{\partial (\hat{X}_t - Z_t)}{\partial \beta_k} \right] \right) \Big|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}},\end{aligned} \tag{6.0.22}$$

and

$$\begin{aligned}\boldsymbol{\eta}_k =& \sum_{t=0}^{n-1} \frac{U_{tk}^{(i)}(\boldsymbol{\beta}_n)}{r_{t-1}(\boldsymbol{\beta}_n)} \sum_{j=1}^{q} \hat{\boldsymbol{\beta}}_{p+j} U_{t-j}^{(i)}(\boldsymbol{\beta}_n) \\ & + \sum_{t=0}^{n-1} \left[X_t - \sum_{j=1}^{p} \hat{\beta}_j X_{t-j} - \sum_{j=1}^{q} \hat{\boldsymbol{\beta}}_{p+j} Z_{t-i}(\boldsymbol{\beta}_n) \right] \frac{\partial}{\partial \beta} \left(\frac{U_{tk}^{(i)}}{r_{t-1}} \right)_{\boldsymbol{\beta} = \boldsymbol{\beta}_n} \\ =& \sum_{t=0}^{n-1} \frac{U_{tk}^{(i)}(\boldsymbol{\beta}_0)}{r_{t-1}(\boldsymbol{\beta}_0)} \sum_{j=1}^{q} \boldsymbol{\beta}_{p+j} U_{t-j}^{(i)}(\boldsymbol{\beta}_0) + \sum_{t=0}^{n-1} Z_t(\boldsymbol{\beta}_0) \frac{\partial}{\partial \beta} \left(\frac{U_{tk}^{(i)}}{r_{t-1}} \right)_{\boldsymbol{\beta} = \boldsymbol{\beta}_0} \\ & + O_p(n \| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \|).\end{aligned}$$

In the above expression, $\boldsymbol{U}_t = \left(U_{t1}^{(i)}, \ldots, U_{tp}^{(i)}, V_{t1}^{(i)}, \ldots, V_{tq}^{(i)} \right)$, and $\boldsymbol{\beta}_n$ is always between $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}_0$. Similarly, the equation $\frac{\partial}{\partial \beta_{p+j}} \ell(\boldsymbol{\beta})|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}} = 0$ $(1 \leq k \leq q)$ leads to

$$\begin{aligned}0 =& \sum_{t=0}^{n-1} \left[X_t - \sum_{j=1}^{p} \hat{\boldsymbol{\beta}}_j X_{t-j} - \sum_{j=1}^{q} \hat{\boldsymbol{\beta}}_{p+j} Z_{t-i}(\boldsymbol{\beta}_0) \right] V_{tk}^{(i)}(\boldsymbol{\beta}_0) r_{t-1}(\boldsymbol{\beta}_0) \\ & + \boldsymbol{\eta}_{p+k}^{(i)} {}^{(i)\prime} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \delta_{p+k}^{(i)},\end{aligned}$$

where

$$\delta_{p+k}^{(i)} = \left(\frac{\sigma_t^2}{2} \frac{\partial}{\partial \beta_{p+k}} \sum_{t=0}^{n-1} \log r_{t,i} - \frac{1}{2} \sum_{t=0}^{n-1} \frac{(X_t - \hat{X}_t)^2}{r_{t,i}^2} \frac{\partial r_{t,i}}{\partial \beta_{p+k}} \right.$$
$$\left. - \frac{1}{2} \sum_{t=0}^{n-1} \left[\frac{X_t - \hat{X}_t + Z_t}{r_{t,i}} \frac{\partial (\hat{X}_t + Z_t)}{\partial \beta_{p+k}} + \frac{X_t - \hat{X}_t - Z_t}{r_{t,i}} \frac{\partial (\hat{X}_t - Z_t)}{\partial \beta_{p+k}} \right] \right) \Big|_{\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}}, \tag{6.0.23}$$

and

$$\boldsymbol{\eta}_{p+k}^{(i)} = \sum_{t=0}^{n-1} \frac{U_{tk}^{(i)}(\boldsymbol{\beta}_0)}{r_{t-1}(\boldsymbol{\beta}_0)} \sum_{j=1}^{q} \boldsymbol{\beta}_{p+j} U_{t-j}^{(i)}(\boldsymbol{\beta}_0) + \sum_{t=0}^{n-1} Z_t(\boldsymbol{\beta}_0) \frac{\partial}{\partial \beta} \left(\frac{V_{tk}^{(i)}}{r_{t-1}} \right)_{\boldsymbol{\beta} = \boldsymbol{\beta}_0}$$
$$+ O_p(n \| \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \|).$$

It follows from (6.0.21) and (6.0.23) that

$$\boldsymbol{U}' R^{-1} \boldsymbol{X} \hat{\boldsymbol{\beta}} = \boldsymbol{U}' R^{-1} \boldsymbol{Y} + \boldsymbol{A}' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + \boldsymbol{\delta^{(i)}}, \tag{6.0.24}$$

where

$$\boldsymbol{\delta^{(i)}} = (\delta_1^{(i)}, \ldots, \delta_{p+q}^{(i)})',$$

and $\boldsymbol{A}$ is the $(p+q) \times (p+q)$ matrix with $\boldsymbol{\eta}_k^{(i)}$ as its k-th column. Note that $\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}_0 = \mathcal{Z}$ and

$$\boldsymbol{U} = \boldsymbol{X} - \sum_{j=1}^{q} \boldsymbol{\beta}_{p+j} \begin{pmatrix} \boldsymbol{U}_{-j}(\boldsymbol{\beta}_0)' \\ \vdots \\ \boldsymbol{U}_{n-1-j}(\boldsymbol{\beta}_0)' \end{pmatrix}.$$

By (6.0.21), (6.0.23) and (6.0.24), we could obtain

$$\boldsymbol{U}R^{-1}\boldsymbol{U}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)=\boldsymbol{U}R^{-1}\boldsymbol{\mathcal{Z}}+\left[\boldsymbol{B}^{(i)}\right]'(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0)+\boldsymbol{\delta^{(i)}},$$

where $\boldsymbol{B}^{(i)}$ is the $(p+q)\times(p+q)$ matrix, and the sum of last two terms on the right hand side of (6.0.23) is the $(p+k)$-th term for $\boldsymbol{B}^{(i)}$, with $k=1,\ldots,q$. Therefore,

$$\begin{aligned} n^{1/2}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0) &= \left[\frac{\boldsymbol{U}R^{-1}\boldsymbol{U}}{n}-\frac{\boldsymbol{B}^{(i)\prime}}{n}\right]^{-1}\frac{\boldsymbol{U}R^{-1}\boldsymbol{\mathcal{Z}}-\boldsymbol{\delta}^{(i)}}{n^{1/2}} \\ &= \left[\frac{\boldsymbol{U}R^{-1}\boldsymbol{U}}{n}\right]^{-1}\frac{\boldsymbol{U}R^{-1}\boldsymbol{\mathcal{Z}}}{n^{1/2}}+o_p(1), \end{aligned}$$

where the last equality follows from Lemma 6.0.3 and Lemma 6.0.7, and the fact that $\frac{\boldsymbol{B}^{(i)}}{n}\xrightarrow{p}0$, with a similar proof as (6.0.17). Then the theorem follows from Lemma 6.0.9. □

Chapter 7

Periodic AIC for PARMA$_S(p, q)$ model

7.1 Kullback-Liebler (K-L) Information

The development of the AIC is predicted on the Kullback-Liebler (K-L) information between two probability density functions f and g, where K-L information is defined to be

$$I(f, g) = \int f(t) \ln \left(\frac{f(t)}{g(t)} \right) dt.$$

The notation $I(f, g)$ denotes a measure of the information lost when g is used to approximate f, which is also the expectation of the logarithmic difference between the two density functions f and g.

Example 7.1.1. Suppose we approximate the normal distribution given by $f(t|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-\mu}{\sigma})^2}$ with $g(t|\xi, \tau^2) = \frac{1}{\tau\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{t-\xi}{\tau})^2}$. Then

$$\ln \left(\frac{f(t)}{g(t)} \right) = \frac{1}{2} \{ \ln \frac{\tau^2}{\sigma^2} - (\frac{t-\mu}{\sigma})^2 + (\frac{t-\xi}{\tau})^2 \},$$

and the K-L information is

$$\begin{aligned} I(f,g) &= \mathrm{E}_f\left[\ln\left(\frac{f(t)}{g(t)}\right)\right] = \frac{1}{2}\left[\ln\frac{\tau^2}{\sigma^2} - \mathrm{E}(\frac{t-\mu}{\sigma})^2 + \mathrm{E}(\frac{t-\xi}{\tau})^2\right] \\ &= \frac{1}{2}\left[\ln\frac{\tau^2}{\sigma^2} - 1 + \frac{\sigma^2+(\mu-\xi)^2}{\tau^2}\right]. \end{aligned}$$

If the true distribution f is the standard normal and $g \sim \mathrm{N}(0.1, 1.5)$, then

$$I(f,g) = \frac{1}{2}\{\ln\frac{1.5}{1} - 1 + \frac{1+(0-0.1)^2}{1.5}\} = 0.0394.$$

Proposition 7.1.2. $I(f,g) \geq 0$.

Proof. For a convex function, $\mathcal{C}$, Jensen's Inequality from Durrett [16, Theorem 1.5.1] asserts that $\mathcal{C}\left[\mathrm{E}(X)\right] \leq \mathrm{E}\left(\mathcal{C}(X)\right)$. Therefore by letting $\mathcal{C} = -\ln(t)$, we write

$$\begin{aligned} I(f,g) &= \int f(t)\ln\left(\frac{f(t)}{g(t)}\right)dt \\ &= -\int f(t)\ln\left(\frac{g(t)}{f(t)}\right)dt \\ &= \int f(t)\mathcal{C}\left(\frac{g(t)}{f(t)}\right)dt \\ &= \mathrm{E}_f\left[\mathcal{C}\left(\frac{g(t)}{f(t)}\right)\right] \\ &\geq \mathcal{C}\left[\mathrm{E}_f\left(\frac{g(t)}{f(t)}\right)\right] \\ &= -\ln\int f(t)\left(\frac{g(t)}{f(t)}\right)dt \\ &= -\ln\int g(t)dt \\ &= -\ln 1 \\ &= 0. \end{aligned}$$

□

7.2 Derivation of periodic AIC for PARMA$_S(p,q)$ process

Following the ideas from Basawa and Lund [24], we treat the white noise variances $\boldsymbol{\sigma^2} = (\sigma_0^2, \sigma_1^2, \ldots, \sigma_{S-1}^2)'$ as nuisance parameters. We use $\boldsymbol{\beta} = (\boldsymbol{\phi}_0', \boldsymbol{\theta}_0', \boldsymbol{\phi}_1', \boldsymbol{\theta}_1', \ldots, \boldsymbol{\phi}_{S-1}', \boldsymbol{\theta}_{S-1}')'$ to denote the collection of all PARMA$_S(p,q)$ parameters. The dimension of $\boldsymbol{\beta}$ is $(p+q)S \times 1$. Then the likelihood function is given by (5.1.13), where $v_{j,i}$ and $\hat{X}_{i+j}^{(i)}$ depend on $\boldsymbol{\beta}$. We also need the asymptotic results of MLE for PARMA process in Basawa and Lund [24]. For a causal and invertible Gaussian PARMA model, with the assumption of $\{\varepsilon_t\}$ being periodic i.i.d. Gaussian noise, Theorem 3.1 in Basawa and Lund [24] gave the asymptotic distribution of $\hat{\boldsymbol{\beta}}$,

$$N^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to \mathrm{N}\left(\mathbf{0}, A^{-1}(\boldsymbol{\beta}, \boldsymbol{\sigma^2})\right), \tag{7.2.1}$$

where

$$A(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) = \sum_{i=0}^{S-1} \sigma_i^{-2} \Gamma_i(\boldsymbol{\beta}, \boldsymbol{\sigma^2}), \tag{7.2.2}$$

and

$$\Gamma_i(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) = \mathrm{E}\left[\left(\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)\left(\frac{\partial \varepsilon_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)'\right],$$

where the right hand side also depends on $\boldsymbol{\beta}$ and $\boldsymbol{\sigma^2}$ by (5.2.7). However, their proof for the asymptotic distribution is based on the asymptotic equivalence of lease square estimator and MLE. In Chapter 6 of my thesis, I gave a direct proof in Theorem 6.0.10.

Once we obtain the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$, the MLE of σ_i^2 for $0 \le i \le S-1$

could be computed from (5.1.16), where

$$\hat{\sigma}_i^2 = \frac{1}{N}\sum_{k=0}^{N-1}\left(X_{kS+i} - \hat{X}_{kS+i}\right)^2 / r_{kS+i} = S_i/N,$$

where $\hat{X}_{kS+i}$ and r_{kS+i} come from the innovations algorithm applied to the model.

Now we are ready to discuss the derivation of AIC. In the K-L information, f represents the true probability distribution and g represents a model distribution that estimates f. In the context of PARMA time series modeling, we assume that the truth f and all approximating alternatives g are Gaussian. Additionally, define it as $g(t|\boldsymbol{\beta_0})$. Suppose $\boldsymbol{X}$ is an n-length PARMA$_S(p,q)$ time series whose probability density is given by $f(t) = g(t|\boldsymbol{\beta_0})$, where $\boldsymbol{X} = (X_i, X_{i+1}, \ldots, X_{i+n-1})'$. To see how we use the K-L information to determine which model, $g(t|\boldsymbol{\beta})$, best fits the truth f, so that it would minimize $I(f,g)$, we write

$$\begin{aligned} I(f,g) &= \int f(t)\ln(\frac{f(t)}{g(t|\boldsymbol{\beta})})dt \\ &= \int f(t)\ln(f(t))dt - \int f(t)\ln(g(t|\boldsymbol{\beta}))dt. \end{aligned} \tag{7.2.3}$$

For all models g, the first integral on the right-hand side of (7.2.3) is a constant, so it suffices to maximize $\int f(t)\ln(g(t|\boldsymbol{\beta}))dt$. Given that we have data $\boldsymbol{Y} = (Y_i, Y_{i+1}, \ldots, Y_{i+n-1})'$ as a sample from the same truth $f(t)$, the logical step would be to find the MLE $\hat{\boldsymbol{\beta}}(\boldsymbol{Y})$, since $\hat{\boldsymbol{\beta}}(\boldsymbol{Y})$ approximates $\hat{\boldsymbol{\beta}}_{\mathbf{0}}$ that minimizes the K-L information. Then we compute an estimate of $I(f, g(t|\boldsymbol{\beta_0}))$ as

$$I(f, g(t|\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))) = \int f(t)\ln(\frac{f(t)}{g(t|\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))})dt.$$

Since $f = g(t|\boldsymbol{\beta_0})$, any value of $\hat{\boldsymbol{\beta}}(\boldsymbol{Y})$ other than $\boldsymbol{\beta_0}$ results in $I(f, g(t|\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))) >$

$I(f, g(t|\boldsymbol{\beta_0}))$.

Consider the method of repeated sampling as a guide to inference, and minimize the K-L discrepancy $\mathrm{E}_{\boldsymbol{Y}}[I(f, g(t|\hat{\boldsymbol{\beta}}(\boldsymbol{Y})))]$. We therefore want to select the model g that minimizes

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{Y}}[I(f, g(t|\hat{\boldsymbol{\beta}}(\boldsymbol{Y})))] &= \int f(t)\ln(f(t))dt - \mathrm{E}_{\boldsymbol{Y}}\left[\int f(t)\ln\left(g(t|\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))\right)dt\right] \\
&= \text{constant} - \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}\left[\ln\left(g(t|\hat{\boldsymbol{\beta}}(\boldsymbol{Y})\right)\right] \\
&= \text{constant} - T,
\end{aligned}$$

where all expectations are taken under the assumption that f is the density of X and Y, and X and Y are independent. Hence, the K-L information criterion for selecting the best model, g, is to maximize the objective function denoted by

$$T = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}\left[\ln\left(g(t|\hat{\boldsymbol{\beta}}(\boldsymbol{Y})\right)\right]. \tag{7.2.4}$$

We have assumed that $f(t)$ has the true $\mathrm{PARMA}_S(p,q)$ model structure, and $g\left(t|\hat{\boldsymbol{\beta}}(\boldsymbol{Y})\right)$ is an estimate of $f(t)$. By (7.2.1), $\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) \to \boldsymbol{\beta}$ as $N \to \infty$. Note that $\hat{\boldsymbol{\beta}}$ is not necessarily equal to $\hat{\boldsymbol{\beta}}_{\boldsymbol{0}}$ or even of the same dimension. Here $\hat{\boldsymbol{\beta}}$ is the parameter vector for the PARMA model under consideration with the smallest K-L discrepancy from the true model.

Without loss of generality we take the likelihood of $\boldsymbol{\beta}(\boldsymbol{X})$ as $g(t|\boldsymbol{\beta}(\boldsymbol{X})) = L_{\boldsymbol{X}}(\boldsymbol{\beta}(\boldsymbol{X}))$, by simply then interpreting g as a function of $\boldsymbol{\beta}(\boldsymbol{X})$ given $\boldsymbol{X}$, which equals the likelihood under data $\boldsymbol{X}$. Similarly, under data $\boldsymbol{Y}$ we write $g(t|\boldsymbol{\beta}(\boldsymbol{Y})) = L_{\boldsymbol{Y}}(\boldsymbol{\beta}(\boldsymbol{Y}))$. In the following we will show that our estimate of $T = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{Y}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y})))]$ would be $\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))$,

where the data $\boldsymbol{X} = (X_i, X_{i+1}, \ldots, X_{i+n-1})'$ are given. The bias of this estimate is

$$\text{bias} = \mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - T. \tag{7.2.5}$$

We will obtain a first order estimate of the bias, and then remove this. Ultimately, AIC is defined to be

$$AIC = -2\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X}))) + 2 \text{ bias}. \tag{7.2.6}$$

Next we will state a few prerequisite results for our main theorem. Define

$$S(\boldsymbol{\beta}) = \sum_{j=0}^{n-1} \frac{\varepsilon_{i+j}^2(\boldsymbol{\beta})}{\sigma_{i+j}^2} = \sum_{j=0}^{N-1}\sum_{k=0}^{S-1} \frac{\varepsilon_{jS+k+i}^2(\boldsymbol{\beta})}{\sigma_{k+i}^2}, \tag{7.2.7}$$

where $n = NS$, and

$$\varepsilon_t = X_t - \sum_{k=1}^{p} \phi_t(k) X_{t-k} - \sum_{j=1}^{q} \theta_t(j)\varepsilon_{t-j},$$

are the model residuals. Then we can have the following result, where we write the residual ε_t as $\varepsilon_t(\boldsymbol{\beta})$ to emphasize explicit dependence of ε_t on $\boldsymbol{\beta}$, and $\boldsymbol{\beta}$ needs to be estimated.

Proposition 7.2.1. *Under the assumption of finite second moment for the causal and invertible PARMA process, as $N \to \infty$, we have*

$$\frac{1}{N}\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \to 2A(\boldsymbol{\beta}, \boldsymbol{\sigma}^{\mathbf{2}}) \quad \textit{in probability,}$$

where $A(\boldsymbol{\beta}, \boldsymbol{\sigma}^{\mathbf{2}})$ is given in (7.2.2).

Proof. In the following, $\boldsymbol{\beta}$ is a $(p+q)S \times 1$ vector, $\dfrac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}$ is a $(p+q)S \times (p+q)S$ dimensional

matrix.

$$\begin{aligned}
\frac{1}{N}\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} &= \frac{1}{N}\frac{\partial^2(\sum_{j=0}^{N-1}\sum_{k=0}^{S-1}\frac{\varepsilon^2_{jS+k+i}(\boldsymbol{\beta})}{\sigma^2_{k+i}})}{\partial \boldsymbol{\beta}^2} \\
&= \frac{2}{N}\frac{\partial(\sum_{j=0}^{N-1}\sum_{k=0}^{S-1}\frac{\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\sigma^2_{k+i}}\frac{\partial\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}'} \\
&= \frac{2}{N}\sum_{j=0}^{N-1}\sum_{k=0}^{S-1}\sigma^{-2}_{jS+k+i}(\frac{\partial\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}})(\frac{\partial\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}})' \\
&+ \frac{2}{N}\sum_{j=0}^{N-1}\sum_{k=0}^{S-1}\frac{\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\sigma^2_{k+i}}\frac{\partial^2\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^2} \\
&= 2\sum_{k=0}^{S-1}\sigma^{-2}_{i+k}\left[\frac{1}{N}\sum_{j=0}^{N-1}(\frac{\partial\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}})(\frac{\partial\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}})'\right] \\
&+ 2\sum_{k=0}^{S-1}\left[\frac{1}{N}\sum_{j=0}^{N-1}\frac{\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\sigma^2_{k+i}}\frac{\partial^2\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^2}\right],
\end{aligned}$$

As $N \to \infty$, by equation (3.13) in Basawa and Lund [24],

$$\frac{1}{N}\sum_{j=0}^{N-1}(\frac{\partial\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}})(\frac{\partial\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}})' \to \Gamma_{i+k}(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) \quad \text{in probability,}$$

and by equation (3.14) in Basawa and Lund [24],

$$\frac{1}{N}\sum_{j=0}^{N-1}\frac{\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\sigma^2_{k+i}}\frac{\partial^2\varepsilon_{jS+k+i}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^2} \to \mathbf{0} \quad \text{in probability,}$$

therefore as $N \to \infty$,

$$\begin{aligned} \frac{1}{N}\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} &\to 2\sum_{k=0}^{S-1} \sigma_{i+k}^{-2}\Gamma_{i+k}(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) + \mathbf{0} \quad \text{in probability} \\ &= 2A(\boldsymbol{\beta}, \boldsymbol{\sigma^2}). \end{aligned}$$

□

Lemma 7.2.2. *Given X a random variable, if $\mathrm{E}|X| < \infty$, then as $x \to \infty$,*

$$\mathrm{E}\left(|X| I_{\{|X|>x\}}\right) \to 0.$$

Thus every random variable X such that $\mathrm{E}|X| < \infty$ is by itself uniform integrable.

Proof. $0 \leq |X| I_{\{|X|>x\}}$ is monotone increasing in x to $|X|$, and therefore using the monotone convergence theorem yields

$$\mathrm{E}\left[|X| I_{\{|X|\leq x\}}\right] \to \mathrm{E}|X|, \quad \text{as} \quad x \to \infty.$$

Notice that

$$\mathrm{E}|X| = \mathrm{E}\left[|X| I_{\{|X|\leq x\}}\right] + \mathrm{E}\left[|X| I_{\{|X|>x\}}\right],$$

by assumption $\mathrm{E}|X| < \infty$, we conclude that

$$\mathrm{E}\left(|X| I_{\{|X|>x\}}\right) \to 0,$$

and X is by definition uniform integrable. □

Lemma 7.2.3. *If $X_n \to X$ in probability, the following statements are equivalent: (i) $\{X_n : n \geq 0\}$ is uniform integrable. (ii) $X_n \to X$ in L^1. (iii) $\mathrm{E}|X_n| \to \mathrm{E}|X| < \infty$.*

Proof. See Durrett [16, pp. 221-222]. □

In the following, we rewrite the second derivatives of log likelihood function, to simplify notation. Define

$$\Omega(\boldsymbol{\beta}) = \frac{\partial^2 \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^2},$$

then we have the following lemma, which was given in Lund et al. [25, Equation 10] without proof. A few necessary assumptions are needed for Lemma 7.2.4 and Theorem 7.2.5, and they guarantee that $\frac{1}{N}\frac{\partial^2 S^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}$, $\frac{1}{N}\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}$ and $\frac{1}{N}(-2\Omega(\boldsymbol{\beta}))$ are not far apart. If we could prove the strong consistency of $\hat{\boldsymbol{\beta}}$ for PARMA model, some of the assumptions below may be reduced. This would be the direction of our further investigation. Currently we adopt the strong consistency for MLE of a vector ARMA model from Section 3 of Basawa and Lund [24], and apply it for PARMA model.

- A1. $\Omega(\boldsymbol{\beta})$ is bounded and continuous on the parameter space $\mathcal{B}$.
- A2. $\frac{1}{N}(\frac{\partial^2 S^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} - \frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}) = o_p(1)$.
- A3. $\frac{1}{N}\left(\Omega(\boldsymbol{\beta}) + \frac{1}{2}\frac{\partial^2 S^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}\right) = o_p(1)$.
- A4. $N\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'$ is uniform integrable.

Lemma 7.2.4.

$$-\frac{\Omega(\boldsymbol{\beta})}{N} \to A(\boldsymbol{\beta}, \boldsymbol{\sigma}^2) \quad \textit{in probability as} \quad N \to \infty.$$

Proof. By Proposition 7.2.1, as $N \to \infty$,

$$\frac{1}{2N}\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \to A(\boldsymbol{\beta}, \boldsymbol{\sigma^2}) \quad \text{in probability,}$$

then by assumption A2 and A3 we can write, as $N \to \infty$,

$$\begin{aligned} -\frac{\Omega(\boldsymbol{\beta})}{N} &= -\frac{1}{N}\left(\Omega(\boldsymbol{\beta}) + \frac{1}{2}\frac{\partial^2 S^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}\right) + \frac{1}{N}\left(\frac{1}{2}\frac{\partial^2 S^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} - \frac{1}{2}\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}\right) + \frac{1}{2N}\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \\ &= -o_p(1) + \frac{1}{2}o_p(1) + \frac{1}{2N}\frac{\partial^2 S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2} \\ &\to A(\boldsymbol{\beta},\boldsymbol{\sigma^2}) \quad \text{in probability,} \end{aligned}$$

where the last equality is by Brockwell and Davis [11, Proposition 6.1.3]. □

Now we are ready to state our main theorem.

Theorem 7.2.5. *Suppose that X_t is a causal and invertible Gaussian $PARMA_S(p,q)$ process, such that assumptions $A1 - A4$ hold. Let $f(t)$ be the joint probability density function of $\boldsymbol{X} = (X_i, X_{i+1}, \ldots, X_{i+n-1})'$, and $\hat{\boldsymbol{\beta}}(\boldsymbol{X})$ denote the maximum likelihood estimates given data $\boldsymbol{X}$, suppose that we are given an independent realization of the same process $\boldsymbol{Y} = (Y_i, Y_{i+1}, \ldots, Y_{i+n-1})'$, with $\hat{\boldsymbol{\beta}}(\boldsymbol{Y})$ as its MLE. Let $T = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{Y}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y})))]$, then*

$$T = \mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - (p+q)S + o(1).$$

Proof. In the following, both the expectations $\mathrm{E}_{\boldsymbol{X}}$ and $\mathrm{E}_{\boldsymbol{Y}}$ are with respect to f, where the samples $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent. Let true parameter values be $\boldsymbol{\beta}_0 \in \mathcal{B}$, where $\mathcal{B}$ is the

parameter space containing all $\boldsymbol{\beta}$. From (7.2.5),

$$\begin{aligned}
&\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - T \\
=&\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))] \\
=&\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X}))) - \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))] \\
=&\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X}))) - \ln\left(L_{\boldsymbol{X}}(\boldsymbol{\beta_0})\right)] + \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0})) - \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))] \\
=&\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X}))) - \ln\left(L_{\boldsymbol{X}}(\boldsymbol{\beta_0})\right)] + \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0})) - \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))]
\end{aligned} \tag{7.2.8}$$

We will prove in the following that $\mathrm{E}_{\boldsymbol{X}}\left[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X}))) - \ln\left(L_{\boldsymbol{X}}(\boldsymbol{\beta_0})\right)\right] = \frac{1}{2}(p+q)S + o(1)$ and $\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0})) - \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))] = \frac{1}{2}(p+q)S + o(1)$, respectively.

First of all, we compute $\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X}))) - \ln\left(L_{\boldsymbol{X}}(\boldsymbol{\beta_0})\right)]$ in (7.2.8). Apply a Taylor series expansion to $\ln\left(L_{\boldsymbol{X}}(\boldsymbol{\beta_0})\right)$ about the MLE $\hat{\boldsymbol{\beta}}(\boldsymbol{X})$ for a sample of data $\boldsymbol{X}$ yielding

$$\begin{aligned}
\ln L_{\boldsymbol{X}}(\boldsymbol{\beta_0}) \;=\; & \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X}))) + [\frac{\partial \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))}{\partial \boldsymbol{\beta}}]'(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) \\
& + \frac{1}{2}(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))'[\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))}{\partial \boldsymbol{\beta}^2}](\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) + \mathrm{Re},
\end{aligned} \tag{7.2.9}$$

where $\left[\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))}{\partial \boldsymbol{\beta}^2}\right]$ is a $(p+q)S \times (p+q)S$ matrix, $\left(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})\right)$ is a $1 \times (p+q)S$ column vector, and Re represents the exact remainder term for the quadratic Taylor series expansion and assume it is uniformly integrable. Note that $\mathrm{Re} = \mathrm{o_p}(\frac{1}{\mathrm{n}})$, and the convergence of Re in probability could be inferred from Lemma 6.0.7 and Brockwell and Davis [11, Proposition 6.1.5].

Since $\hat{\boldsymbol{\beta}}(\boldsymbol{X})$ is the MLE from data $\boldsymbol{X}$, then

$$[\frac{\partial \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))}{\partial \boldsymbol{\beta}}] = [\frac{\partial \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}}]_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}} = 0.$$

Taking expectations on both sides of (7.2.9),

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{X}}[\ln L_{\boldsymbol{X}}(\boldsymbol{\beta_0})] &= \mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] \\
&+ \frac{1}{2}\mathrm{E}_{\boldsymbol{X}}[(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \frac{\partial^2 \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))}{\partial \boldsymbol{\beta}^2}(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))] + \mathrm{E}_{\boldsymbol{X}}[\mathrm{Re}] \\
&= \mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] \\
&+ \frac{1}{2}\mathrm{tr}\{\mathrm{E}_{\boldsymbol{X}}[\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))}{\partial \boldsymbol{\beta}^2}](\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))'\} + \mathrm{E}_{\boldsymbol{X}}[\mathrm{Re}],
\end{aligned}$$

where $\mathrm{Re} = o_p(\frac{1}{n})$ by Lemma 6.0.7 and Brockwell and Davis [11, Proposition 6.1.5]. Therefore the expectation of the remainder term is negligible for large sample sizes if we assume the uniform integrability, i.e. $\lim_{N\to\infty} \mathrm{E}_{\boldsymbol{X}}[\mathrm{Re}] = 0$.

Additionally, because $\hat{\boldsymbol{\beta}}$ is the MLE under $L_{\boldsymbol{X}}(\boldsymbol{\beta_0})$, by (7.2.1), $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta_0}$ in probability as $N \to \infty$. By assumption A1 and [11, Proposition 6.1.4], we have

$$\Omega(\hat{\boldsymbol{\beta}}) \to \Omega(\boldsymbol{\beta_0}) \quad \text{in probability as} \quad N \to \infty.$$

By assumption A1 and Lemma 7.2.2, $\Omega(\hat{\boldsymbol{\beta}})$ is also uniformly integrable. Then by Lemma 7.2.3 we can get:

$$\lim_{N\to\infty} \mathrm{E}_{\boldsymbol{X}}[\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))}{\partial \boldsymbol{\beta}^2}] = \mathrm{E}_{\boldsymbol{X}}[\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^2}].$$

Hence we may write

$$
\begin{aligned}
&(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \frac{\partial^2 \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))}{\partial \boldsymbol{\beta}^2} (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) \\
&= (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \Omega(\hat{\boldsymbol{\beta}}) (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) \\
&= (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \Omega(\boldsymbol{\beta_0}) (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) \\
&+ (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \left(\Omega(\hat{\boldsymbol{\beta}}) - \Omega(\boldsymbol{\beta_0}) \right) (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) \\
&= (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \Omega(\boldsymbol{\beta_0}) (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) + o_p(1),
\end{aligned}
$$

since $(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) = o_p(1)$ so that

$$
\left(\Omega(\hat{\boldsymbol{\beta}}) - \Omega(\boldsymbol{\beta_0}) \right) = o_p(1),
$$

and

$$
(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \left(\Omega(\hat{\boldsymbol{\beta}}) - \Omega(\boldsymbol{\beta_0}) \right) (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) = o_p(1)
$$

by Brockwell and Davis [11, Proposition 6.1.1]. Therefore

$$
(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \Omega(\hat{\boldsymbol{\beta}}) (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) - (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \Omega(\boldsymbol{\beta_0}) (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) = o_p(1). \quad (7.2.10)
$$

Now by assumption A1 and A4,

$$
\mathrm{E}_{\boldsymbol{X}} \left[(\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \Omega(\hat{\boldsymbol{\beta}}) (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) - (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X}))' \Omega(\boldsymbol{\beta_0}) (\boldsymbol{\beta_0} - \hat{\boldsymbol{\beta}}(\boldsymbol{X})) \right] = o(1).
$$

Therefore

$$
\begin{aligned}
&\mathrm{E}_{\boldsymbol{X}}\left[(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))'\Omega(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))\right]\\
&=\mathrm{E}_{\boldsymbol{X}}\left[(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))'\Omega(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))-(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))'\Omega(\boldsymbol{\beta_0})(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))\right]\\
&+\mathrm{E}_{\boldsymbol{X}}\left[(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))'\Omega(\boldsymbol{\beta_0})(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))\right]\\
&=o(1)+\mathrm{E}_{\boldsymbol{X}}\left[(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))'\Omega(\boldsymbol{\beta_0})(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))\right]\\
&=o(1)+\mathrm{tr}\left\{\mathrm{E}_{\boldsymbol{X}}\left[\Omega(\boldsymbol{\beta_0})\left(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X})\right)\left(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X})\right)'\right]\right\}.
\end{aligned}
$$

Then

$$
\begin{aligned}
\lim_{N\to\infty}\mathrm{E}_{\boldsymbol{X}}[\ln L_{\boldsymbol{X}}(\boldsymbol{\beta_0})] &= \lim_{N\to\infty}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))]\\
&+ \frac{1}{2}\lim_{N\to\infty}\mathrm{E}_{\boldsymbol{X}}\left[(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))'\Omega(\hat{\boldsymbol{\beta}})(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X}))\right]\\
&+ \lim_{N\to\infty}\mathrm{E}_{\boldsymbol{X}}[\mathrm{Re}]\\
&= \lim_{N\to\infty}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))]\\
&+ +\frac{1}{2}\lim_{N\to\infty}\mathrm{tr}\{\mathrm{E}_{\boldsymbol{X}}\left[\Omega(\boldsymbol{\beta_0})\left(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X})\right)\left(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X})\right)'\right]\}\\
&= \mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))]-\frac{1}{2}\lim_{N\to\infty}\mathrm{tr}\{\mathrm{E}_{\boldsymbol{X}}\left[A_N W_N\right]\},
\end{aligned}
$$

where we define $A_N=-\frac{\Omega(\boldsymbol{\beta_0})}{N}$, and $W_N=N\left(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X})\right)\left(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X})\right)'$. Additionally we define $W_N=Z_N Z_N^{'}$, where $Z_N=\sqrt{N}\left(\boldsymbol{\beta_0}-\hat{\boldsymbol{\beta}}(\boldsymbol{X})\right)\to Z$ as $N\to\infty$ and $Z\sim \mathrm{N}(\mathbf{0},A^{-1}(\boldsymbol{\beta},\boldsymbol{\sigma^2}))$ by (7.2.1). Next we will show that

$$
\lim_{N\to\infty}\mathrm{E}_{\boldsymbol{X}}\left[A_N W_N\right]=I_{(p+q)S}, \tag{7.2.11}
$$

where $I_{(p+q)S}$ is the identity matrix. Notice that $A_N \to A(\boldsymbol{\beta},\boldsymbol{\sigma^2})$ in probability as $N \to \infty$, by Lemma 7.2.4, and $W_N \to W = ZZ'$, then $\mathrm{E}W_N \to \mathrm{E}W = \mathrm{E}\left(ZZ'\right) = A^{-1}(\boldsymbol{\beta},\boldsymbol{\sigma^2})$, where the uniform integrability is guaranteed in assumption A4. Hence

$$\begin{aligned}\mathrm{E}_{\boldsymbol{X}}\left[A_N W_N\right] = \mathrm{E}_{\boldsymbol{X}}\left[A_N\left(W_N - W\right)\right] + \mathrm{E}_{\boldsymbol{X}}\left[\left(A_N - A(\boldsymbol{\beta},\boldsymbol{\sigma^2})\right)W\right] \\ + \mathrm{E}_{\boldsymbol{X}}\left[A(\boldsymbol{\beta},\boldsymbol{\sigma^2})W\right].\end{aligned} \tag{7.2.12}$$

By assumption A1, $|A_N| < M$, where $M < \infty$, therefore as $N \to \infty$,

$$0 \le |\mathrm{E}_{\boldsymbol{X}}\left[A_N\left(W_N - W\right)\right]| \le |\mathrm{E}_{\boldsymbol{X}}\left[|A_N|\left(W_N - W\right)\right]| \le M|\mathrm{E}_{\boldsymbol{X}}\left[\left(W_N - W\right)\right]| \to 0,$$

then

$$\mathrm{E}_{\boldsymbol{X}}\left[A_N\left(W_N - W\right)\right] \to 0. \tag{7.2.13}$$

Similarly,

$$\mathrm{E}_{\boldsymbol{X}}\left[\left(A_N - A(\boldsymbol{\beta},\boldsymbol{\sigma^2})\right)W\right] \to 0, \tag{7.2.14}$$

and

$$\mathrm{E}_{\boldsymbol{X}}\left[A(\boldsymbol{\beta},\boldsymbol{\sigma^2})W\right] = A(\boldsymbol{\beta},\boldsymbol{\sigma^2})\mathrm{E}_{\boldsymbol{X}}\left[W\right] = A(\boldsymbol{\beta},\boldsymbol{\sigma^2})A^{-1}(\boldsymbol{\beta},\boldsymbol{\sigma^2}) = I_{(p+q)S}. \tag{7.2.15}$$

Substitute (7.2.13), (7.2.14) and (7.2.15) into (7.2.12), and let $N \to \infty$, to arrive at (7.2.11). Then

$$\begin{aligned}\lim_{N\to\infty} \mathrm{E}_{\boldsymbol{X}}[\ln L_{\boldsymbol{X}}(\boldsymbol{\beta_0})] &= \mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - \frac{1}{2}\lim_{N\to\infty} \mathrm{tr}\{I_{(p+q)S}\} \\ &= \mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - \frac{1}{2}(p+q)S.\end{aligned}$$

So we obtain

$$\lim_{N\to\infty}\{\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - \mathrm{E}_{\boldsymbol{X}}[\ln L_{\boldsymbol{X}}(\boldsymbol{\beta_0})]\} = \frac{1}{2}(p+q)S,$$

or, in other words,

$$\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - \mathrm{E}_{\boldsymbol{X}}[\ln L_{\boldsymbol{X}}(\boldsymbol{\beta_0})] = \frac{1}{2}(p+q)S + o(1). \tag{7.2.16}$$

Now let us consider $\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0})) - \ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))]$, the remaining term in the last line of (7.2.8). Similarly, apply the Taylor expansion to $\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))$ around $\boldsymbol{\beta_0}$ for any given data $\boldsymbol{X}$ yielding

$$\begin{aligned}\ln L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y})) &= \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0})) + [\frac{\partial \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}}]'(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0}) \\ &\quad + \frac{1}{2}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0})'[\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^2}](\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0}) + \mathrm{Re}.\end{aligned}$$

Taking expectations with respect to X yields

$$\begin{aligned}\mathrm{E}_{\boldsymbol{X}}[\ln L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))] &= \mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] + \mathrm{E}_{\boldsymbol{X}}[\frac{\partial \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}}]'(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0}) \\ &\quad + \frac{1}{2}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0})'\mathrm{E}_{\boldsymbol{X}}[\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^2}](\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0}) \\ &\quad + \mathrm{E}_{\boldsymbol{X}}[\mathrm{Re}]\end{aligned} \tag{7.2.17}$$

where $\boldsymbol{Y}$ is independent of $\boldsymbol{X}$.

Taking expectations of (7.2.17) with respect to $\boldsymbol{Y}$ yields

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))] &= \mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] + \mathrm{E}_{\boldsymbol{X}}[\frac{\partial \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}}]'\mathrm{E}_{\boldsymbol{Y}}\left[\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0}\right] \\
&+ \frac{1}{2}\mathrm{E}_{\boldsymbol{Y}}[(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0})'\mathrm{E}_{\boldsymbol{X}}[\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^2}](\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0})] \\
&+ \mathrm{E}_{\boldsymbol{X}}[\mathrm{Re}],
\end{aligned}$$

letting $N \to \infty$ on both sides, the linear terms vanishes since $\mathrm{E}_{\boldsymbol{Y}}\left[\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0}\right] \to 0$ by

(7.2.1), then the above equations becomes

$$
\begin{aligned}
&\lim_{N\to\infty} \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))] \\
=&\lim_{N\to\infty} \mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] + \lim_{N\to\infty} \mathrm{E}_{\boldsymbol{X}}[\frac{\partial \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}}]' \mathrm{E}_{\boldsymbol{Y}}\left[\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0}\right] \\
&+\frac{1}{2}\lim_{N\to\infty} \mathrm{E}_{\boldsymbol{Y}}[(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0})' \mathrm{E}_{\boldsymbol{X}}[\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^2}](\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0})] \\
&+\lim_{\mathrm{N}\to\infty} \mathrm{E}_{\boldsymbol{X}}[\mathrm{Re}] \\
=&\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] + \mathrm{E}_{\boldsymbol{X}}[\frac{\partial \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}}]' \lim_{N\to\infty} \mathrm{E}_{\boldsymbol{Y}}\left[\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0}\right] \\
&+\frac{1}{2}\lim_{N\to\infty} \mathrm{tr}\{\mathrm{E}_{\boldsymbol{X}}[\frac{1}{N}\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^2}]\mathrm{E}_{\boldsymbol{Y}}[N^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0})N^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0})']\} \\
=&\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] \\
&+\frac{1}{2}\mathrm{tr}\{\lim_{N\to\infty} \mathrm{E}_{\boldsymbol{X}}[\frac{1}{N}\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^2}]\mathrm{E}_{\boldsymbol{Y}}[N^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0})N^{\frac{1}{2}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}) - \boldsymbol{\beta_0})']\} \\
=&\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] + \frac{1}{2}\mathrm{tr}\{\mathrm{E}_{\boldsymbol{X}} \lim_{N\to\infty}[\frac{1}{N}\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^2}]A(\boldsymbol{\beta}, \boldsymbol{\sigma^2})^{-1}\} \\
=&\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] \\
&+\frac{1}{2}\mathrm{tr}\{\mathrm{E}_{\boldsymbol{X}} \lim_{N\to\infty}[\frac{1}{N}\frac{\partial^2 \ln(L_{\boldsymbol{X}}(\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}^2} + \frac{1}{2}\frac{\partial^2 S^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}]A(\boldsymbol{\beta}, \boldsymbol{\sigma^2})^{-1}\} \\
&-\frac{1}{2}\mathrm{tr}\{\mathrm{E}_{\boldsymbol{X}} \lim_{N\to\infty}[\frac{1}{2}\frac{\partial^2 S^*(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}]A(\boldsymbol{\beta}, \boldsymbol{\sigma^2})^{-1}\} \\
=&\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] + 0 - \frac{1}{4}\mathrm{tr}\{[2A(\boldsymbol{\beta}, \boldsymbol{\sigma^2})A(\boldsymbol{\beta}, \boldsymbol{\sigma^2})^{-1}]\} \\
=&\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] - \frac{1}{2}\mathrm{tr}\{I_{(p+q)S}\} \\
=&\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] - \frac{1}{2}(p+q)S,
\end{aligned}
$$

Hence

$$
\lim_{N\to\infty}\{\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] - \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))]\} = \frac{1}{2}(p+q)S,
$$

or, in other words,

$$\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\boldsymbol{\beta_0}))] - \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y}))] = \frac{1}{2}(p+q)S + o(1). \tag{7.2.18}$$

Add (7.2.16) and (7.2.18) into (7.2.8), we have

$$\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - T = (p+q)S + o(1)$$

which completes the proof. □

Akaike [2] defined an information criterion (AIC) by multiplying $\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))$ by -2, to get

$$AIC = -2\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X}))) + 2K,$$

where K is the bias term for maximum log-likelihood as an estimator for

$$T = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{Y}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y})))],$$

which is equal to the number of estimable parameters in the model. This has become known as Akaike's information criterion or AIC. For a PARMA$_S(p,q)$ model, if we treat the variance as nuisance parameters, there are $k = (p+q)S$ estimable parameters. It is also proved in Theorem 7.2.5, where

$$\lim_{N\to\infty} \text{bias} = \lim_{N\to\infty} \{\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - T\} = (p+q)S,$$

so that

$$\text{bias} = (p+q)S + o(1).$$

Therefore an asymptotically unbiased estimator of $T = \mathrm{E}_{\boldsymbol{Y}}\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{Y}}(\hat{\boldsymbol{\beta}}(\boldsymbol{Y})))]$ is

$$\mathrm{E}_{\boldsymbol{X}}[\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X})))] - (p+q)S.$$

From (7.2.6) the AIC for a PARMA model is obtained by

$$AIC = -2\ln(L_{\boldsymbol{X}}(\hat{\boldsymbol{\beta}}(\boldsymbol{X}))) + 2(p+q)S, \tag{7.2.19}$$

7.2.1 Application to model selection for the Fraser River

Besides the compare of forecast plots in Chapter 5, we can also compute the value of AIC for each candidate model, and the one yielding the minimum AIC is the best model. The results are shown in Table 7.1. Within all full model candidates, the full PARMA$(1,1)$ model fitted by MLE in Table 5.4 has the minimum AIC, and so it is the best full model. For reduced model, the model in in Table 5.7, obtained by asymptotic distribution of MLE, has the minimum AIC, and there are only 13 estimable parameters. Lastly, for $\mathrm{PAR}_S(p)$ model, we tried two different approaches. First approach is by removing all $\hat{\theta}_i$ in Table 5.7, since there are only three of them. In this way, we obtain a $\mathrm{PAR}_{12}(1)$ model, and there are only 12 estimable parameters. However, the value of AIC turns out to be very large. The second approach is done in a more rigorous way, using the *pear* package in *R* to do automatic model selection for $\mathrm{PAR}_{12}(p)$ model. The *pear* package was developed by A.I. McLeod and Mehmet Balcilar, for estimating periodic autoregressive models, and they provided a built-in

data set for the historical flows of Fraser river. The best model selected is shown in Table 7.2, which is a $\text{PAR}_{12}(3)$ model with 19 estimable parameters. We compute its AIC, shown in Table 7.1. Additionally, Table 7.3 demonstrate the AIC values for the logarithm of the data, where the model parameters have more impacts on AIC. Note that both full and reduced $\text{PARMA}_{12}(1,1)$ models perform better than PAR models. Therefore the $\text{PARMA}_{12}(1,1)$ model is a better model for the Fraser river flows.

Model	Number of parameters	AIC
Full $\text{PARMA}_{12}(1,1)$ model in Table 5.3	24	18524.25
Full $\text{PARMA}_{12}(1,1)$ model in Table 5.4	24	18476.47
Reduced $\text{PARMA}_{12}(1,1)$ model in Table 5.5	19	18768.79
Reduced $\text{PARMA}_{12}(1,1)$ model in Table 5.7	13	18528.13
$\text{PAR}_{12}(1)$ model, removing all $\hat{\theta}_i$ in Table 5.7	12	18754.09
$\text{PAR}_{12}(3)$ model in Table 7.2	19	17714.33

Table 7.1: Comparison of AIC values for different models

Season i	$\hat{\boldsymbol{\phi}}_i(1)$	$\hat{\boldsymbol{\phi}}_i(2)$	$\hat{\boldsymbol{\phi}}_i(3)$	$\hat{\boldsymbol{\sigma}}_i$
0	0.527	0.000	0.000	6325.612
1	0.779	-0.231	0.189	5004.514
2	0.764	0.000	0.000	5326.131
3	1.188	0.000	0.000	17540.533
4	0.647	0.000	0.000	37180.017
5	0.411	-1.237	1.562	38084.824
6	0.545	0.000	0.000	34809.521
7	0.517	0.000	0.000	17666.091
8	0.661	-0.127	0.000	13518.955
9	0.890	-0.434	0.165	14685.378
10	0.631	0.000	0.000	12434.951
11	0.543	0.000	0.000	8535.279

Table 7.2: The model parameters in $\text{PAR}_{12}(p)$ automatic model fitting by *pear* package in *R*, where the number of estimable parameters is 19, assuming $\hat{\boldsymbol{\sigma}}_i$ as nuisance parameters in the AIC computation.

Model	Number of parameters	AIC
Full $\text{PARMA}_{12}(1,1)$ model in Table 5.3	24	-415.8586
Full $\text{PARMA}_{12}(1,1)$ model in Table 5.4	24	-506.6053
Reduced $\text{PARMA}_{12}(1,1)$ model in Table 5.5	19	-313.3939
Reduced $\text{PARMA}_{12}(1,1)$ model in Table 5.7	14	-297.9929
$\text{PAR}_{12}(3)$ model in Table 7.2	19	520.9146

Table 7.3: A compare of AIC values for different models, after taking the log of the data. Note that both full and reduced $\text{PARMA}_{12}(1,1)$ models perform better than PAR models.

7.3 Future Research

In this section I would like to list out the open problems in my research, and this would also be helpful for researchers who are interested in studying in this topic deeply.

- Strong consistency of MLE. This could follow the result for ARMA model from Yao and Brockwell [57]. Clear details applying to PARMA model should be carefully generalized.

- A stronger condition for Theorem 7.2.5, with fewer assumptions. The uniform integrability of $N\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)'$ is waiting for a complete proof, and the property of the second derivative of likelihood function in PARMA model needs to be studied in details.

- I will work on cleaning up my R code, and add the forecasting tool in perARMA package. This would be very useful for researchers would work on hydrology and periodic stationary time series prediction.

APPENDIX

APPENDIX

```
# R code for the paper  "Forecasting for periodic ARMA models"
# by Anderson, Paul; Meerschaert, Mark; Zhang, Kai

# Please put the file "frazierc.txt" under your work directory

########################################
#   Seasonal Sample Mean for (4.1)
########################################

data <- read.table("frazierc.txt")
XMEAN <- array(0,c(12))
# XMEAN is a vector of sample means for 12 seasons
for(I in 1:12)
{
 XMEAN[I] <- 0
 for(T in 0:71)
 {
 XMEAN[I] <- XMEAN[I] + data[T*12+I,1]
 }
 XMEAN[I] <- XMEAN[I]/72
```

```
}

#####################################

# Sample Autocovariance for (4.2)

#####################################


# season i = 0, 1, ..., 11

# LAG = 0, 1, ..., 124


COVAR <- array(0,c(12,125))

for(L in 1:125)

{

 LAG <- L-1

 for (I in 1:12)

 {

  i <- (I-1)

  COVAR[I,L] <- 0

  J <- as.integer((i+LAG)/12)

  K <- ((I+LAG)-(12*J))

  for (T in 0:(71-J))

   {

   COVAR[I,L] <- COVAR[I,L]

   +(data[T*12+I,1]-XMEAN[I])*(data[T*12+I+LAG,1]-XMEAN[K])
```

```
   }
 COVAR[I,L] <- COVAR[I,L]/(72-J)
 }
}

####################################
# Sample Autocorrelation for (4.3)
####################################
rho <- array(0,c(12,125))
for (I in 1:12)
{
for (L in 1:125)
{
R <- (I+L-2)%%12 + 1
rho[I,L] <- COVAR[I,L]/((COVAR[I,1]*COVAR[R,1])^.5)
}
}

##############################################
#  Innovations Algorithm for X_t process
##############################################

##  I = i + 1
```

```
##  J = j + 1

##  K = k + 1

##  L = ell + 1

##  N = n + 1

##  COVAR(I,L) = gamma_i(ell)

##  V(N,I) = v_{n,i}

##  THETA(N,M,I) = theta_{n,m}^{(i)} WHERE m = M + 1

##  NOTE THAT:  theta_{n,n-k}^{(i)} = THETA(N,N-K+1,I)

##         theta_{k,k-j}^{(i)} = THETA(K,K-J+1,I)

##         gamma_k(n-k)  = COVAR(K,N-K+1)


V<- array(0,c(50,12))

THETA <- array(1,c(50,50,12))

for (I in 1:12)

{

 V[1,I] <- COVAR[I,1]

 for (N in 2:50)

 {

  for (K in 1:(N-1))

  {

    S <- 0

    if (K == 1)

    {
```

```
  K0 <- as.integer((I+1-2)/12)
  K1 <- (I+K-1)-12*K0
  THETA[N,N,I] <- (COVAR[K1,N]-S)/V[1,I]
  }

  else
  {
   for (J in 1:(K-1))
    {
    S <- S+THETA[K,K-J+1,I]*THETA[N,N-J+1,I]*V[J,I]
    K0 <- as.integer((I+K-2)/12)
    K1 <- (I+K-1)-12*K0
    THETA[N,N-K+1,I] <- (COVAR[K1,N-K+1]-S)/V[K,I]
    }
  }
}
   R <- 0
   for(J in 1:(N-1))
   {
   R <- R+V[J,I]*(THETA[N,N-J+1,I])^2
   N0 <- as.integer((I+N-2)/12)
   N1 <- (I+N-1)-12*N0
   V[N,I] <- COVAR[N1,1]-R
```

```
     }
 }
}

# At k = 20 iterations, get the convergence of THETA and V
psi1 <- array(0,c(12,12))
for ( I in 1:12 )
{
R <- 0
for (J in 1:12)
{
R <- (I-20-1)%%12+1
psi1[I,J] <- THETA[21,J,R]
}
}

sigma_square <- array(0,c(12))
for (I in 1:12)
{
S <- 0
S <- (I-20-1)%%12+1
sigma_square[I] <- V[21,S]
}
```

```
##############################################
# Get model parameter estimates by (4.4)
##############################################

phi <- array(0,c(12))
sigma <- array(0,c(12))
theta <- array(0,c(12,864))
for (I in 1:12)
{
R <- ((I-1)-1)%%12+1
phi[I] <- psi1[I,3]/psi1[R,2]
}
for (I in 1:12)
{
theta[I,1] <- -1
}
for (I in 1:12)
{
theta[I,2] <- psi1[I,2]-phi[I]
}
for (I in 1:12)
{
```

```
sigma[I] <- (sigma_square[I])^.5
}

# A simple output of model estimates
phi
theta[,2]
sigma

# reduced model
# phi[1] <- 0
# phi[5] <- 0
# phi[7] <- 0
# phi[8] <- 0
# phi[10] <- 0

#######################################
# The following is for prediction
#######################################

# Autocovariances K(J,L) for W_t process
# K(J,L) = C in (2.6)

K <- array(0,c(865,865))
```

```
for (I in 1:12)     # I is season
{
  K[I,I] <- COVAR[I,1]      # when J = L = I



for (J in 1:12)
{
  for (L in 1:13)
  {
    if ( J <= I && L == (I+1) )
    {
     s1 <- (J-1)%%12+1
     l1 <- (L-1)%%12+1
     K[J,L] <- COVAR[s1,abs(J-L)+1]
     - phi[l1]*COVAR[s1,abs(L-1-J)+1]
    }
  }
}
}

for (I in 1:12)
{
```

```
for (J in 1:865)
{
  for (L in 1:865)
  {
  if (min(J,L) >= (I+1) && abs(J-L) <= 1)
  {
  s1 <- (J-1)%%12+1
  n1 <- (L-1)%%12+1
  q1 <- (J-2)%%12+1
  K[J,L] <- theta[s1,1]*theta[n1,abs(J-L)+1]*(sigma[s1])^2
  + theta[s1,2]*theta[n1,(abs(1+J-L))+1]*(sigma[q1])^2
  }
  }
}
}

##########################################
# Innovations algorithm for W_t process
# The computation in (2.6)
##########################################

V<- array(0,c(865,12))
THETA <- array(0,c(865,865,12))
```

```
for (I in 1:12)

{

for (J in 1:865)

{

THETA[J,1,I] <- 1

}

}


for (I in 1:12)

{

 V[1,I] <- K[I,I]


 for (N in 2:(865-I))

 {

     THETA[N,2,I] <- K[I+N,I+N-1]/V[N-1,I]

     V[N,I] <- K[I+N,I+N]- V[N-1,I]*(THETA[N,2,I])^2

 }

}



#####################################

# Computation of \hat{X} in (2.7)

#####################################
```

```
# Subtract seasonal mean

fraser <- t(data)

X <- array(0,c(864))

for(I in 1:12)

{

 for(T in 0:71)

 {

 X[T*12+I] <- fraser[T*12+I]-XMEAN[I]

 }

}


# Xhat = \hat{X} in (2.7)

Xhat <- array(0,c(12,880))

for (I in 1:12)

{

Xhat[I,1+I] <- 0

Xhat[I,2+I] <- THETA[2,2,I]*(X[I+1]-Xhat[I,I+1])

for (N in 3:(840-I))

{

   s1 <- (I+N-1)%%12+1

   Xhat[I,N+I] <- phi[s1]*X[I+N-1]+ THETA[N,2,I]*(X[I+N-1]-Xhat[I,I+N-1])

}
```

```
}

##############################################
# h-step prediction in (2.8) and (2.9)
##############################################
for (I in 1:12)
{
s1 <- (841-1)%%12+1
Xhat[I,841] <- phi[s1]*X[840] + THETA[841-I,2,I]*(X[840]-Xhat[I,840])
for (h in 2:24)
{
m1 <- (840+h-1)%%12+1
Xhat[I,840+h] <- phi[m1]*Xhat[I,840+h-1]
}
}

##############################
# Forecast error in (3.3)
##############################

# Calculation of Casual represention for PARMA(1,1)
# The casual coefficient psi is periodic in S
psi <- array(0,c(12,24))    # I = 12, h =24
```

```
for (I in 1:12)
{
psi[I,1] <- 1
psi[I,2] <- (phi[I]+theta[I,2])
  S <- 1
  for (k in 3:24)
  {
    for (j in 0:(k-3))
  {
  j0 <- (I-j-1)%%12+1
  S <- S*phi[j0]
  }
   j1 <- (I-(k-1)-1)%%12+1
   psi[I,k] <- S*(phi[j1]+theta[j1,2])
  }
}


# h-step prediction error
# Use this error for confidence band in Corr.3.2
sigma_h2 <- array(0,c(12,24))
for (I in 1:12)
{
  for (h in 1:24)
```

```
  {
    R <- 0
    for (J in 1:h)
    {
    s0 <- (840+h-1)%%12+1
    s1 <- (840+h-J-1)%%12+1
    R <- R + (psi[s0,J])^2*(sigma[s1])^2
    }
    sigma_h2[I,h] <- R
  }
}

# Add seasonal mean to Xhat
Yhat <- array(0,c(880))
for(I in 1:12)
{
 for(T in 0:71)
 {
 Yhat[T*12+I] <- Xhat[1,T*12+I]+XMEAN[I]
 }
}

# Computation of residuals
```

```
res <- X - Yhat[1:864]

for (I in 1:12)

{

for (T in 0:71)

{

res[T*12+I] <- res[T*12+I]/sigma[I]

}

}


################################################

# 95% Prediction bounds for h-step prediction in Figure 3

################################################

CI_low <- array(0,c(24))

CI_up <- array(0,c(24))

for (I in 1:12)

{

  for (h in 1:24)

  {

  CI_low[h] <- Yhat[840+h]-1.96*sqrt(sigma_h2[I,h])

  CI_up[h] <- Yhat[840+h]+1.96*sqrt(sigma_h2[I,h])

  }

}
```

```
##########################################
# 95% Confidence Intervals for sample mean
##########################################
CI_low_mean <- array(0,c(12))
CI_up_mean <- array(0,c(12))
for (I in 1:12)
{
  CI_low_mean[I] <- XMEAN[I]-1.96*(COVAR[I,1])^.5
  CI_up_mean[I] <- XMEAN[I]+1.96*(COVAR[I,1])^.5
}


########################################
# 95% Confidence Intervals for gamma_0
########################################
error <- array(0,c(12))    ## define error as sqrt{[(V_00)_ellell]/72}
for (I in 1:12)
{
S <- 0
for (L in 0:10)
{
S <- S + 4*((COVAR[I,12*L+1])^2)
}
error[I] <- ((S - 2*((COVAR[I,1])^2))/72)^.5
```

```
## subtract the exact one from niu = 0
}

CI_low_gamma0 <- array(0,c(12))
CI_up_gamma0 <- array(0,c(12))

for (I in 1:12)
{
  CI_low_gamma0[I] <- COVAR[I,1]-1.96*error[I]
  CI_up_gamma0[I] <- COVAR[I,1]+1.96*error[I]
}

########################################
# 95% Confidence Intervals for rho_1
########################################

W_11 <- array(0,c(12))
for (I in 1:12)
{
S <- 0
R <- ((I-1)%%12)+1
## S is the sum at n = 0, niu = 12
S <- rho[I,1]*rho[R,1] + rho[I,2]*rho[R,2] - rho[I,2]*(rho[I,1]*rho[I,2]+
```

```
rho[R,2]*rho[R,1]) - rho[I,2]*(rho[I,1]*rho[R,2]+rho[I,2]*rho[R,1]) +
.5*rho[I,2]^2*(rho[I,1]^2+rho[I,2]^2+rho[R,2]^2+rho[R,1]^2)
## Next loop is sum from n = -10 to 10
for (L in 1:10)
{
S <- S + (rho[I,L*12+1]*rho[R,L*12+1] + rho[I,L*12+2]*rho[R,L*12] -
rho[I,2]*(rho[I,L*12+1]*rho[I,L*12+2]+rho[R,L*12]*rho[R,L*12+1]) -
rho[I,2]*(rho[I,L*12+1]*rho[R,L*12]+rho[I,L*12+2]*rho[R,L*12+1]) +
.5*rho[I,2]^2*(rho[I,L*12+1]^2
+rho[I,L*12+2]^2+rho[R,L*12]^2+rho[R,L*12+1]^2))
}
P <- 0
for (L in (-10):(-1))
{
P <- P + (rho[I,abs(L*12)+1]*rho[R,abs(L*12)+1] +
rho[I,abs(L*12+1)+1]*rho[R,abs(L*12-1)+1] -
rho[I,2]*(rho[I,abs(L*12)+1]*rho[I,abs(L*12+1)+1]+
rho[R,abs(L*12-1)+1]*rho[R,abs(L*12)+1]) -
rho[I,2]*(rho[I,abs(L*12)+1]*rho[R,abs(L*12-1)+1]+
rho[I,abs(L*12+1)+1]*rho[R,abs(L*12)+1]) +
.5*rho[I,2]^2*(rho[I,abs(L*12)+1]^2+rho[I,abs(L*12+1)+1]^2
+rho[R,abs(L*12-1)+1]^2+rho[R,abs(L*12)+1]^2))
}
```

```
W_11[I] <- P+S
}

CI_low_rho1 <- array(0,c(12))
CI_up_rho1 <- array(0,c(12))

for (I in 1:12)
{
  CI_low_rho1[I] <- rho[I,2]-1.96*((W_11[I]/72)^.5)
  CI_up_rho1[I] <- rho[I,2]+1.96*((W_11[I]/72)^.5)
}

########################################
# 95% Confidence Intervals for rho_2
########################################
W_22 <- array(0,c(12))
for (I in 1:12)
{
S <- 0
R <- ((I)%%12)+1
## S is the sum at n = 0, niu = 12
S <- rho[I,1]*rho[R,1] + rho[I,3]*rho[R,3] -
rho[I,3]*(rho[I,1]*rho[I,3]+rho[R,3]*rho[R,1]) -
```

```
rho[I,3]*(rho[I,1]*rho[R,3]+rho[I,3]*rho[R,1]) +
.5*rho[I,3]^2*(rho[I,1]^2+rho[I,3]^2+rho[R,3]^2+rho[R,1]^2)
## Next loop is sum from n = -10 to 10
for (L in 1:10)
{
S <- S + (rho[I,L*12+1]*rho[R,L*12+1] +
rho[I,L*12+3]*rho[R,L*12-1]
- rho[I,3]*(rho[I,L*12+1]*rho[I,L*12+3]+
rho[R,L*12-1]*rho[R,L*12+1])
- rho[I,3]*(rho[I,L*12+1]*rho[R,L*12-1]+
rho[I,L*12+3]*rho[R,L*12+1]) + .5*rho[I,3]^2*(rho[I,L*12+1]^2+
rho[I,L*12+3]^2+rho[R,L*12-1]^2+rho[R,L*12+1]^2))
}
P <- 0
for (L in (-10):(-1))
{
P <- P + (rho[I,abs(L*12)+1]*rho[R,abs(L*12)+1] +
rho[I,abs(L*12+2)+1]*rho[R,abs(L*12-2)+1] -
rho[I,3]*(rho[I,abs(L*12)+1]*rho[I,abs(L*12+2)+1]+
rho[R,abs(L*12-2)+1]*rho[R,abs(L*12)+1]) -
rho[I,3]*(rho[I,abs(L*12)+1]*rho[R,abs(L*12-2)+1]+
rho[I,abs(L*12+2)+1]*rho[R,abs(L*12)+1]) +
.5*rho[I,3]^2*(rho[I,abs(L*12)+1]^2+rho[I,abs(L*12+2)+1]^2+
```

```
rho[R,abs(L*12-2)+1]^2+rho[R,abs(L*12)+1]^2))
}
W_22[I] <- P+S
}

CI_low_rho2 <- array(0,c(12))
CI_up_rho2 <- array(0,c(12))

for (I in 1:12)
{
  CI_low_rho2[I] <- rho[I,3]-1.96*((W_22[I]/72)^.5)
  CI_up_rho2[I] <- rho[I,3]+1.96*((W_22[I]/72)^.5)
}



#############
# Output
#############

# Please remove "#" if you want to generate output files

write(CI_low,file="CI_low_prediction.txt",ncolumns=1)
write(CI_up,file="CI_up_prediction.txt",ncolumns=1)
```

```
write(fraser,file="Data of Fraser River.txt",ncolumns=1)
write(Yhat[1:864],file="24-month Predictions.txt",ncolumns=1)

# write(XMEAN,file="Sample Mean.txt",ncolumns=1)
# write(CI_low_mean,file="CI_low_mean.txt",ncolumns=1)
# write(CI_up_mean,file="CI_up_mean.txt",ncolumns=1)

# write(COVAR[,1],file="Sample Variance.txt",ncolumns=1)
# write(CI_low_gamma0,file="CI_low_variance.txt",ncolumns=1)
# write(CI_up_gamma0,file="CI_up_variance.txt",ncolumns=1)

# write(rho[,2],file="rho_1.txt",ncolumns=1)
# write(CI_low_rho1,file="CI_low_rho1.txt",ncolumns=1)
# write(CI_up_rho1,file="CI_up_rho1.txt",ncolumns=1)

# write(rho[,3],file="rho_2.txt",ncolumns=1)
# write(CI_low_rho2,file="CI_low_rho2.txt",ncolumns=1)
# write(CI_up_rho2,file="CI_up_rho2.txt",ncolumns=1)

# write(res,file="Residuals.txt",ncolumns=1)

############################
#         Plots
```

```
############################

# The plots in this paper were producted by Minitab
# The following plots are provided as drafts for reference

############
# Figure 1
g_range <- range(0, data)
plot(data[1:180,1], col = "black", ylab = "Flow (cms)",
axes=FALSE, type = "l", ylim=g_range, xlab="Month / Year",
 cex.lab=1.5, lty = 5)
axis(1,at = c(1,37,73,109,145,180),lab= c("10/1912",
"10/1915","10/1918","10/1921","10/1924","09/1927"))
axis(2, at = c(1,50000,100000,150000,200000,250000,
300000,350000,400000), lab =c("0","50000","100000",
"150000","200000","250000","300000","350000","400000") )

############
# Figure 3
############
g_range <- range(0, CI_up)
plot(CI_low, col = "blue", ylab = "Flow (cms)", lwd = 2,
axes=FALSE, type = "l", ylim=g_range, xlab="Month / Year",
```

```
cex.main = 2, cex.lab=1.5, lty = 5)

axis(1,at = c(1,5,9,13,17,21,24),lab= c("10/1982","02/1983",

"06/1983","10/1983","02/1984","06/1984",""))

axis(2)

lines(CI_up, col = "blue", type="l", lty=5, lwd = 2)

lines(Yhat[841:864], col = "red", type="o", pch=20, lty=1, lwd = 2)

lines(data[841:864,1],col = "black", type="l", lty=1, lwd = 2)


############

# Figure 4

############

g_range <- range((CI_low-c(XMEAN,XMEAN)), (CI_up-c(XMEAN,XMEAN)))

plot((CI_low-c(XMEAN,XMEAN)), col = "blue", type = "l",

ylab = "Width of prediction bounds", ylim=g_range, xlab="Month",

main="Width of prediction bounds (mean subtracted)", lty = 5)

lines(Yhat[841:864]-c(XMEAN,XMEAN), col = "red", type="o", pch=20, lty=1)

lines((CI_up-c(XMEAN,XMEAN)), col = "blue", type="l", lty=5)


############

# Figure 2

############

par(mfrow = c(2,2))

g_range <- range(0, (CI_up_mean))
```

```
plot(XMEAN, col = "black", ylab = "Sample mean (cms)", lwd =2,
ylim=g_range,axes=FALSE, xlab="Season",
main="(a) Sample Means", lty = 5, cex.main = 2.5, cex.lab=1.5)
axis(1,at = 1:12,lab= c("0","1","2","3","4","5","6","7","8","9","10","11"))
axis(2)
lines(XMEAN, type="o",lwd = 2,pch = 20)
lines((CI_low_mean), col = "red", type="l", lty=5,lwd =2)
lines((CI_up_mean), col = "red", type="l", lty=5,lwd =2)

g_range <- range(0, (CI_up_gamma0)^.5 )
plot((COVAR[,1])^.5, col = "black", ylab = "Sample sd (cms)",
axes=FALSE, lwd=2, ylim=g_range, xlab="Season",
main="(b) Sample Standard Deviations", lty = 5,cex.main = 2.2, cex.lab=1.5)
axis(1,at = 1:12,lab= c("0","1","2","3","4","5","6","7","8","9","10","11"))
axis(2)
lines((COVAR[,1])^.5, type="o",lwd = 2,pch = 20)
lines((CI_low_gamma0)^.5, col = "red", type="l", lty=5,lwd=2)
lines((CI_up_gamma0)^.5, col = "red", type="l", lty=5,lwd=2)

g_range <- range(0, 1)
plot(rho[,2], col = "black", ylab = "Autocorrelations", axes=FALSE,
ylim=g_range, xlab="Season", main="(c) Sample Autocorrelations : lag = 1",
lty = 5,cex.main = 1.9, cex.lab=1.5)
```

```
axis(1,at = 1:12,lab= c("0","1","2","3","4","5","6","7","8","9","10","11"))
axis(2)
lines(rho[,2], type="o",lwd = 2,pch = 20)
lines(CI_low_rho1, col = "red", type="l", lty=5,lwd=2)
lines(CI_up_rho1, col = "red", type="l", lty=5,lwd=2)

g_range <- range(-.5, 1)
plot(rho[,3], col = "black", ylab = "Autocorrelations", axes=FALSE,lwd=2,
ylim=g_range, xlab="Season", main="(d) Sample Autocorrelations : lag = 2",
lty = 5,cex.main = 1.9, cex.lab=1.5)
axis(1,at = 1:12,lab= c("0","1","2","3","4","5","6","7","8","9","10","11"))
axis(2)
lines(rho[,3], type="o",lwd = 2,pch = 20)
lines(CI_low_rho2, col = "red", type="l", lty=5,lwd=2)
lines(CI_up_rho2, col = "red", type="l", lty=5,lwd=2)

################################################
# Computation of PARMA Autocovariances in Chapter 1.1
################################################

# Model is PARMA_12(1,1)  i.e. there are 12 seasons
# Model is based on table 5 in Tesfaye, Meerschaert and Anderson (2006)

# phi_1 <- c(.198,.568,.560,.565,.321,.956,1.254,.636,-1.942,-.092,.662,.355)
```

```
phi_1 <- c(0,.568,.560,.565,0,.956,0,0,-1.942,0,.662,.355)

theta_1 <- c(.687,.056,-.052,-.05,.47,-.389,-.178,-.114,2.393,.71,-.213,.322)

sigma <- c(11875.479,11598.254,7311.452,5940.845,4160.214,4610.209,

15232.867,31114.514,32824.370,29712.190,15511.187,12077.991)

theta_0 <- array(1,c(12))

psi_0 <- array(1,c(12))

psi_1 <- phi_1 + theta_1

 # By (16) in Tesfaye, Meerschaert and Anderson (2006)


# Set AX = b, then solve X, where X is a vector

# X gives ACVF = gamma_i(h), when h <= max(p,q), i = 0, 1, ... 11

# In PARMA_12(1,1), p = q = 1


A <- array(0,c(24,24))

for (i in 1:12)

{

A[i,i] <- 1

A[i+12,(i-2)%%12+1+12] <- 1

A[i,(i-2)%%12+1+12] <- (-phi_1[i])

A[i+12,(i-2)%%12+1] <- (-phi_1[i])

}


b <- array(0,c(24))
```

```
for (i in 1:12)
{
b[i] <- theta_0[i] * psi_0[i] * (sigma[i])^2 +
theta_1[i] * psi_1[i] * (sigma[(i-2)%%12+1])^2
b[i+12] <- theta_1[i] * psi_0[(i-2)%%12+1] * (sigma[(i-2)%%12+1])^2
}

X <- solve(A,b)
X

matrix(X, ncol = 2)

# COVAR is autocovariance function
# COVAR[I,H] = \gamma_{i}(h)
# I = i + 1, i is season, i = 0, 1, 2, ... 11
# H = h + 1, h is lag, h = 0, 1, 2, ... 79
COVAR <- array(0,c(12,80))

COVAR[,1:2] <- X  # Read X into first columns of COVAR
                  # A good way to get rid of the for loop below

#for (I in 1:12)
# This loop reads X into COVAR, for h <= max(p,q); here h = 0,1
```

```
# {
# COVAR[I,1] <- X[I]
# COVAR[I,2] <- X[I+12]
# }

for (H in 3:80)            # This loop computes ACVF for h > max(p,q)
{
for (I in 1:12)
{
COVAR[(I-H)%%12+1,H] <- phi_1[I]*COVAR[(I-H)%%12+1,H-1]
#this one is right!
# (I-2)%%12+1 represents season I-1
# COVAR[I,H] <- phi_1[I]*COVAR[(I-2)%%12+1,H-1]
}
}

################################
# Innovations Algorithm
################################

V<- array(0,c(50,12))
THETA <- array(0,c(50,50,12))
```

```
for (I in 1:12)          # I = i+1, i is season
{


 V[1,I] <- COVAR[I,1]
 for (N in 2:50)         # N = n+1, n is number of iterations
 {
  for (K in 1:(N-1))
  {
    S <- 0
    if (K == 1)
    {
    K0 <- as.integer((I+1-2)/12)
    K1 <- (I+K-1)-12*K0
    THETA[N,N,I] <- (COVAR[K1,N]-S)/V[1,I]
    }


    else
    {
     for (J in 1:(K-1))
      {
      S <- S+THETA[K,K-J+1,I]*THETA[N,N-J+1,I]*V[J,I]
      K0 <- as.integer((I+K-2)/12)
      K1 <- (I+K-1)-12*K0
```

```
      THETA[N,N-K+1,I] <- (COVAR[K1,N-K+1]-S)/V[K,I]
      THETA[N,1,I] <- 1      # This defines \theta_{n,0}(i) = 1
      }
    }
  }
     R <- 0
     for(J in 1:(N-1))
     {
     R <- R+V[J,I]*(THETA[N,N-J+1,I])^2
     N0 <- as.integer((I+N-2)/12)
     N1 <- (I+N-1)-12*N0
     V[N,I] <- COVAR[N1,1]-R
     }
 }
}

#################################
# convergence of theta to psi
#################################

psi_k <- array(0,c(50,12,50))         # psi_k is \psi_{i}(\ell)
for (K in 1:50)                       # K = k+1, k is number of iterations
{
```

```
for ( I in 1:12 )                    # I = i+1, i is season
{
R <- 0
for (J in 1:50)
{
R <- (I-K)%%12+1
psi_k[K,I,J] <- THETA[K,J,R]
}
}
}
# Test output for lag = 1
# This matches values of psi_1 = phi_1 + theta_1, after 5 iterations
psi_k[,,2]
psi_1

# Test output for lag = 2
# psi_k[,,3]

# This shows the error between psi_k[,,2] and psi_1
error <- array(0,c(50,12))
for (I in 1:12)
{
for (K in 1:50)
```

```
{
error[K,I] <- psi_k[K,I,2]-psi_1[I]
}
}

max_error <- array(0,c(49))
# for all season, for lag =1 only
for (K in 1:49)
 for (I in 1:12)
{
{
max_error[K] <- max(abs(error[K,I]))
}
}
plot(max_error,ylab = "Value of convergence error",
xlab="Number of iterations", main="error",cex.main = 2, cex.lab=1.8)
lines(max_error, type="o", pch=20, lty=1, col="red")



########################################
# convergence of v to sigma^2
########################################
```

```
sigma_square <- array(0,c(50,12))
for (K in 1:50)
{
for (I in 1:12)
{
S <- 0
S <- (I-K)%%12+1
sigma_square[K,I] <- V[K,S]
}
}

# Test output, which matches sigma, after 5 iterations
sigma_square^.5
sigma

# This gives the error between sigma_square and sigma^2
error2 <- array(0,c(50,12))
for (I in 1:12)
{
for (K in 1:50)
{
error2[K,I] <- (sigma_square[K,I]-(sigma[I])^2)
}
```

```
}
```

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Adams, G.J and G.C. Goodwin, (1995), Parameter estimation for periodically ARMA models. *Journal of Time Series Analysis, 16*, 127–145.

[2] Akaike, H. (1974), A new look at the statistical model identification. *IEEE Transactions of Automatic Control, 19*, 716–723.

[3] Anderson, P.L. and A.V. Vecchia (1993), Asymptotic results for periodic autoregressive moving -average processes. *Journal of Time Series Analysis, 14*, 1–18.

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

[5] Anderson, P.L. and M.M. Meerschaert (1998), Modeling river flows with heavy tails. *Water Resources Research* , *34*(9), 2271–2280.

[6] Anderson, P.L., M.M. Meerschaert and A.V. Veccia (1999), Innovations algorithm for periodically stationary time series. *Stochastic Processes and their Applications, 83*, 149–169.

[7] Anderson, P.L. and M.M. Meerschaert (2005), Parameter estimates for periodically stationary time series. *Journal of Time Series Analysis, 26*, 489–518.

[8] Anderson, P.L., M.M. Meerschaert, Y.G. Tesfaye (2007), Fourier-PARMA models and their application to modeling of river flows. *Journal of Hydrologic Engineering*, *12*(5), 462–472.

[9] Ansley, C.F. (1979), An algorithm for the exact likelihood of a mixed autoregressive-moving average process. *Biometrika*, *66*(1), 59–65.

[10] Box, G.E., and G.M. Jenkins (1976), *Time Series Analysis: Forcasting and Control*, 2nd edition, San Francisco: Holden-Day.

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

[12] Broyden, C.G. (1970), The convergence of a class of double-rank minimization algorithms, *Journal of the Institute of Mathematics and Its Applications 6*, 7690.

[13] Burnham, K.P. and Anderson, D.R. (2010), *Model selection and multi-model inference: a practical information theoretic approach*, 1st ed. Springer.

[14] Chen, H.-L. and A.R. Rao (2002), Testing hydrologic time series for stationarity. *Journal of Hydrologic Engineering*,*7*(2), 129–136.

[15] Dunsmuir, W. and Hannan E. J. (1976), Vector linear time series models. *Advances in Applied Probability*, *8*(2), 339 - 364.

[16] Durrett, R. (2010), *Probability: Theory and Examples*, 4th ed. Cambridge University Press.

[17] Fletcher, R. (1970), A New Approach to Variable Metric Algorithms, *Computer Journal 13*(3): 317322.

[18] Gladyshev, E.G. (1961), Periodically correlated random sequences. *Soviet Mathematics, 2*, 385–388.

[19] Goldfarb, D. (1970), A Family of Variable Metric Updates Derived by Variational Means, *Mathematics of Computation 24*(109): 2326.

[20] Jones, R.H. and W.M. Brelsford (1967), Times series with periodic structure. *Biometrika, 54*, 403–408.

[21] Lehmann, E.L. and Caselle, G. (1998), *Theory of Point Estimation*, 2nd ed. Springer.

[22] Lund, R.B. and I.V. Basawa (1999), Modeling and inference for periodically correlated time series. *Asymptotics, Nonparameterics, and Time Sereis*(ed. S. Ghosh), 37–62. Marcel Dekker, New York.

[23] Lund, R.B. and I.V. Basawa (2000), Recursive prediction and likelihood evaluation for periodic ARMA models. *Journal of Time Series Analysis*, *20*(1), 75–93.

[24] Basawa, I.V. and Lund, R.B. (2001), Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis*, 22(6), 651–663.

[25] Lund, R.B., Shao Q. and I.V. Basawa (2006), Parsimonious periodic time series modeling. *Australian & New Zealand Journal*, 48(1), 33–47.

[26] Makagon, A., Miamee, A. G. and Salehi, H. Periodically correlated processes and their spectrum(1991). *Nonstationary stochastic processes and their applications*, 147-164.

[27] Makagon, A., Miamee, A. G. and Salehi, H. Continuous time periodically correlated processes: spectrum and prediction(1994). *Stochastic Process. Appl.*, (49)2, 277-295.

[28] Makagon, A., Miamee, A. G., Salehi, H. and Soltani, A. R. On spectral domain of periodically correlated processes(2008). *Theory Probab. Appl.*, 52(2), 353-361.

[29] Makagon, A. and Salehi, H. Structure of periodically distributed stochastic sequences(1993). *Stochastic processes*, 245-251.

[30] Mcleod, A.I. (1994), Diagnostic checking of periodic autoregression models with application. *Journal of Time Series Analysis*, 15(2), 221-233.

[31] Noakes, D.J., Mcleod, A.I. and Hipel, K.W. (1985), Forecasting monthly riverflow time series. *International Journal of Forecasting, 1*, 179–190.

[32] Pagano, M. (1978), On periodic and multiple autoregressions. *Annals of Statistics*, *6*(6), 1310–1317.

[33] R Development Core Team (2008) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, `http://www.R-project.org`.

[34] Salas, J.D, D.C. Boes and R.A. Smith (1982), Estimation of ARMA models with seasonal parameters. *Water Resources Research*, *18*, 1006–1010.

[35] Salas, J.D, G.Q. Tabios III and P. Bartolini (1985), Approaches to multivariate modeling of water resources time series. *Water Resources Bulletin*, *21*, 683–708.

[36] Salas, J.D. and J.T.B Obeysekera (1992), Conceptual basis of seasonal streamflow time series models. *Journal of Hydraulic Engineering*, *118*(8), 1186–1194.

[37] Schwarz,G. (1978), Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.

[38] Shanno, D.F. (1970), Conditioning of quasi-Newton methods for function minimization, *Math. Comput. 24*(111): 647656.

[39] Shao, Q. and R.B. Lund (2004), Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. *Journal of Time Series Analysis*, *25*(3), 359–372.

[40] Shiryaev, A.N. (1984), *Probability*. Springer-Verlag, New York.

[41] Tesfaye, Y.G. (2005), Seasonal Time Series Models and Their Application to the Modeling of River Flows. PhD. Dissertation, Uinversity of Nevada, Reno, Nevada.

[42] Tesfaye, Y.G., M.M. Meerschaert and P.L. Anderson (2006), Identification of PARMA Models and Their Application to the Modeling of River Flows. *Water Resources Research*, *42*(1), W01419.

[43] Y.G. Tesfaye, P.L. Anderson, and M.M. Meerschaert (2011), Asymptotic results for Fourier-PARMA time series. *Journal of Time Series Analysis*, *32*(2), 157–174.

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

[45] Tjøstheim, D. and J. Paulsen (1982), Empirical identification of multiple time series.*Journal of Time Series Analysis , 3*, 265–282.

[46] Thompstone,R.M., K.W. Hipel and A.I. McLeod (1985), Forecasting quarter-monthly riverflow. *Water Resources Bulletin*, *25*(5),731–741.

[47] Troutman, B.M. (1979), Some results in periodic autoregression. *Biometrika, 6*, 219–228.

[48] Ula, T.A. (1990), Periodic covariance stationarity of multivariate periodic autoregressive moving average processess. *Water Resources Research*, *26*(5), 855–861.

[49] Ula, T.A. (1993), Forcasting of Multivariate periodic autoregressive moving average processes. *Journal of Time Series Analysis, 14*, 645–657.

[50] Ula, T.A. and A.A Smadi (1997), Periodic stationarity conditions for periodic autoregressive moving average processess as eigenvalue problems. *Water Resources Research*, *33*(8), 1929–1934.

[51] Ula, T.A. and A.A Smadi (2003), Identification of periodic moving average models. *Communications in Statistics: Theory and Methods, 32*(12), 2465–2475.

[52] Ursu, E. and Turkman K.F. (2012), Periodic autoregression model identification using genetic algorithms. *Journal of Time Series Analysis*, doi: 10.1111/j.1467-9892.2011.00772.x.

[53] Vecchia , A.V. (1985a), Periodic autoregressive-moving average (PARMA) modelling with applications to water resources. *Water Resources Bulletin 21*, 721–730.

[54] Vecchia, A.V. (1985b), Maximum likelihood estimation for periodic moving average models. *Technometrics*, *27*(4), 375–384.

[55] Vecchia , A.V. and R. Ballerini (1991), Testing for periodic autocorrelations in seasonal time series data. *Biometrika*, *78*(1), 53–63.

[56] Wyłomańska, A. (2008), Spectral measures of PARMA sequences. *Journal of Time Series Analysis*, *29*(1), 113.

[57] Yao, Q. and Brockwell, P.J. (2006), Gaussian maximum likelihood estimation for ARMA models I: time series. *Journal of time series analysis*, *27*(6), 857-875.