Qin Shao
Inference for a Class of Periodic Time Series Models and their Applications
(Under the direction of Robert Lund and Ishwar V. Basawa)

This dissertation investigates some topics involving periodic autoregressive moving-average (PARMA) time series models.

Our first research topic studies autocovariance and partial autocorrelation properties of PARMA models. An efficient algorithm to compute PARMA autocovariances is first derived. An Innovations based algorithm to compute partial autocorrelations for a general periodic series is then developed. Periodic moving-averages and periodic autoregressions are characterized as periodically stationary series whose autocovariances and partial autocorrelations, respectively, are zero at all lags that exceed some periodically varying threshold.

Next, techniques for fitting parsimonious periodic time series models are explored. Large sample standard errors for the parameter estimates of a PARMA model under parametric constraints are derived; likelihood ratio statistics are also explored. The techniques are motivated with the analysis of a daily temperature series from Griffin, Georgia.

The dissertation closes by introducing seasonal periodic autoregressive moving-average time series (SPARMA) models. SPARMA models are a hybrid of seasonal autoregressive moving-average models and PARMA models. Some mathematical properties of SPARMA models are derived.

Index words: Periodic Series, PARMA Model, Autocovariances, Partial Autocorrelations, Innovations Algorithm, Fourier Series, Standard Errors, SARMA Model, SPARMA Model

Inference for a Class of Periodic Time Series Models and their Applications

by

Qin Shao

B.S., Nankai University, China 1990

M.S., Nankai University, China 1993

M.S., The University of Georgia, 1999

A Dissertation Submitted to the Graduate Faculty

of The University of Georgia in Partial Fulfillment

of the

Requirements for the Degree

Doctor of Philosophy

Athens, Georgia

2002

Inference for a Class of Periodic Time Series Models and their Applications

by

Qin Shao

Approved:

Major Professor: Robert Lund
Ishwar Basawa

Committee: William P. McCormick
Jaxk H. Reeves
Lynne Seymour
Robert Taylor

Electronic Version Approved:

Gordhan L. Patel
Dean of the Graduate School
The University of Georgia
May 2002.

Acknowledgements

This dissertation could not have been possible without any one of those people who have helped, influenced, and encouraged me during my pursuit of understanding and knowledge in the field of statistics.

First and foremost, sincerest thanks go to my advisors Dr. Lund and Dr. Basawa. It has been truly a pleasure to work under the direction of Dr. Lund. His wise direction has made the completion of this dissertation a successful and rewarding endeavor. Moreover, his enthusiasm, optimism, and confidence will continue to guide me through the good time and bad in the future of my professional life. I am also indebted to Dr. Basawa for bringing many interesting research topics to my attention and giving me time in his busy schedule.

Special thanks are due to my committee members, Dr. Jaxk Reeves, Dr. Robert Taylor, Dr. Lynne Seymour, and Dr. William P. McCormick for enhancing this dissertation and my professional development. I am grateful to have spent five years with the most knowledgable professors and the most friendly staff as well as fellow students, building my solid professional background.

Last, but not the least, my thanks go to my parents and my siblings for their sacrifices and love over the past thirty years. I am extremely lucky to have been raised in a family where education and achievement are considered as important and honorable goals. Thanks Mom and Dad for instilling in me the values which make me who I am today!

Table of Contents

Chapter 1

Introduction

Many time series in climatology, hydrology, sociology, plant physiology, electrical engineering, and economics exhibit periodicity in their autocovariance structure. Such series include quarterly river flows, daily temperatures, quarterly profit margins, monthly airline ticket sales etc. For example, Lund *et al.* (1995) examine monthly ozone concentrations from Arosa, Switzerland, and conclude that the data are well modeled by a time series whose autocovariances vary from month to month within a year, but repeat at the same months in different years.

Loosely speaking, a periodic series is a random sequence in which the first two moments are periodic with period T. Formally, a time series $\{X_t\}$ with bounded second moments satisfying

$$\mathrm{E}[X_{n+T}] = \mathrm{E}[X_n] \tag{1.0.1}$$

and

$$\mathrm{Cov}(X_{n+T}, X_{m+T}) = \mathrm{Cov}(X_n, X_m), \tag{1.0.2}$$

for all m and n, is called periodic (also periodically stationary, or periodically correlated (PC)) with period T. For clarity, T is taken as the smallest positive integer satisfying (1.0.2).

For notation in a general periodic series, we define

$$\mu_\nu = \mathrm{E}[X_{nT+\nu}] \tag{1.0.3}$$

and the autocovariance function (ACVF) of $\{X_t\}$ at season ν and lag $h \geq 0$, as

$$\gamma_h(\nu) = \text{Cov}(X_{nT+\nu}, X_{nT+\nu-h}), \tag{1.0.4}$$

and the corresponding autocorrelation function (ACF) as

$$\rho_h(\nu) = \frac{\gamma_h(\nu)}{\sqrt{\gamma_0(\nu)\gamma_0(\nu-h)}}.$$

The autocovariance $\text{Cov}(X_{nT+\nu}, X_{nT+\nu-h})$ of a periodic series does not depend on the cycle n, but does depend on both lag h and season ν. The periodic notation $\{X_{nT+\nu}\}$ denotes the series during the νth season of the nth cycle and is preferred to non-periodic notation $\{X_t\}$ when emphasis is on periodicity. A season here refers to any one discrete time point within a period, specifically, a value of ν satisfying $1 \leq \nu \leq T$.

Periodically stationary time series are not stationary unless $T = 1$. Many periodic time series cannot be transformed to stationary series; therefore, some well-established statistical inference techniques for stationary time series are not appropriate in the periodic setting. As we will see, the analysis of periodic time series is much more complicated than that for stationary series.

Before doing any statistical analysis on a given series, it may be necessary to confirm that the series is periodic in its first two moments. Gladyšev (1961) gives a necessary and sufficient condition for a series to be periodic, establishing a connection to T-variate stationary series. Later, Vecchia and Ballerini (1991) and Anderson and Vecchia (1993) developed a formal test for periodicity based on the asymptotic properties of the Fourier transform of the estimated periodic autocorrelation function. These results are theoretically useful but practically inconvenient. Hurd and Gerr (1991), and Lund and Basawa (1999) study average squared coherence methods, easily computed from the discrete Fourier transformation of the series, to detect periodicity. By graphing the average squared coherences of the series, one can usually conclude visually if a given series is periodic or not. We discuss this procedure further in Chapter 2 and apply it to a temperature series in Chapter 4.

Perhaps the most useful class of periodic time series models is comprised of periodic ARMA (PARMA) models. PARMA series satisfy a difference equation similar to that for ARMA models, except that model parameters and white noise variances are allowed to change periodically with season.

PARMA models have been studied extensively in both time and frequency domain literatures because of their wide scope of applicability. Many noteworthy results and methods, including model identification, parameter estimation, and diagnostic checking, have been studied (cf. Hannan 1955, Jones and Brelsford 1967, Tiao and Grupe 1980, Vecchia 1985a and 1985b, Bartolini *et al.* 1988, Lund *et al.* 1995, Ghysels *et al.* 1996, Rasmussen *et al.* 1996, Lund and Basawa 1999, among many others). This knowledge has been successfully applied in analyzing periodic data from a wide range of fields. PARMA models have become the most powerful and useful modeling vehicle for time series with periodically varying second moments. We will highlight the development and major results of PARMA modeling in the next chapter.

A desirable property of fitted time series models is parsimony. The fitted model should fit the available data adequately using the smallest number of model parameters. For ARMA models, parsimony is usually achieved by eliminating unnecessary autoregressive and moving-average model parameters. However, in PARMA models, the number of parameters depends not only on the autoregressive and the moving-average orders, but also on the season. The number of parameters in a PARMA model can be extremely large when the period is large even for a model with small autoregressive and moving-average orders. For example, if the period is a year ($T = 365$), the number of autoregressive model parameters in a first order autoregressive model is 365. Some *ad hoc* parameter reduction techniques are given by Thompstone *et al.* (1985), Bartolini *et al.* (1988), Bloomfield *et al.* (1994), Lund *et al.* (1995), and Rasmussen *et al.* (1996). In Chapter 4, we will use Fourier techniques to reduce and consolidate PARMA model parameters for a daily temperature series.

Several methods for PARMA parameter estimation have been explored, including Yule-Walker moment methods, least squares methods (LS), and Gaussian maximum likelihood (ML) methods. For PAR(p) models, the Yule-Walker estimates are perhaps the easiest to compute. For a Gaussian PARMA series, Yule-Walker estimates are known to be asymptotically most efficient. For Gaussian PARMA series with large T, the calculation of Yule-Walker estimates can be quite involved. Hence, Gaussian maximum likelihood is a better way to estimate PARMA model parameters. Many papers exist on Gaussian likelihood evaluation (cf. Jones and Brelsford (1967), Pagano (1978), Li and Hui (1988), Vecchia (1983, 1985a, 1985b), Adams and Goodwin (1993), and Lund and Basawa (2000)). Lund and Basawa (2000) present an Innovations based algorithm to make one-step predictions, which can be used to evaluate a Gaussian likelihood function. This algorithm is simple to carry out and avoids the complexity of inverting the covariance matrix of the series. We will use this approach in Chapter 4 when estimating the model parameters for a daily temperature series.

The rest of this dissertation is organized as follows. Chapter 2 reviews some basic concepts, the relationship between PARMA and vector ARMA models, and some inference results for these models. Chapter 3 derives Innovations based algorithms to efficiently calculate PARMA autocovariances and partial autocorrelations, and presents a necessary and sufficient condition for a time series to be a periodic moving-average with order q or a periodic autoregression with order p, respectively. Chapter 4 discusses procedures for fitting a parsimonious PARMA model. A series of daily temperature measurements from Griffin, Georgia is used as an example. Finally, Chapter 5 introduces a new class of models, called seasonal periodic autoregressive moving-average (SPARMA) models. SPARMA models are developed for the purpose of fitting periodically correlated time series linked not only to the seasons in the same period, but also to the same seasons of previous periods. The new models can be viewed as a hybrid of standard PARMA models and Box-Jenkins seasonal models. Some basic properties of SPARMA models are derived.

Chapter 2

Literature Review

2.1 Periodically Stationary Processes

Let $\{X_t\}$ be a mean zero periodic series with period T. The autocovariance function (ACVF) defined by $\gamma_h(\nu) = \text{Cov}(X_{nT+\nu}, X_{nT+\nu-h})$ is arguably the most fundamental feature of the series and provides a measure of the degree of dependence between two values in the series seperated by h time units. The ACVF plays an important role when forecasting future series values in terms of the past and present observations.

Unlike its stationary counterpart, $\gamma_h(\nu)$ is not symmetric in h; however, one sees from (1.0.2) that

$$\gamma_{-h}(\nu) = \gamma_h(\nu + h), \qquad h \geq 0,$$

and

$$\gamma_h(\nu) = \gamma_h(\nu + T), \qquad h \geq 0.$$

Generally, periodicity in the ACVF of a series cannot always be visually detected from the sample ACVF. There are periodicity tests in both time and frequency domains (see Vecchia and Ballerini (1991), Anderson and Vecchia (1993), Hurd and Gerr (1991), Bloomfield *et al.* (1994), and Lund *et al.* (1995)). Among these methods, the average squared coherence method, proposed by Bloomfield *et al.* (1994), and developed further by Lund *et al.* (1995), is straightforward and easy to perform. The procedure of this spectral based test is as follows.

Suppose that $X_1, X_2, \cdots, X_N$ represent d full cycles of observations of a periodic time series $\{X_t\}$ with known period T; that is, $d = N/T$ is assumed to be an integer. The Fourier transformation of the series at the Fourier frequency $\lambda_j = 2\pi j/N$, denoted by I_j, is defined by

$$I_j = (2\pi N)^{-\frac{1}{2}} \sum_{n=0}^{N-1} X_{n+1} e^{-in\lambda_j}.$$

Next calculate the squared coherence

$$\gamma_h^2(j) = \frac{|\sum_{m=0}^{M-1} I_{j+m} \overline{I_{j+h+m}}|^2}{\sum_{m=0}^{M-1} |I_{j+m}|^2 \sum_{m=0}^{M-1} |I_{j+h+m}|^2}, \qquad 0 \le j \le N-1,$$

where M is some fixed integer — the choice of M is not crucial. Here, an overline denotes complex conjugation. Note that $\gamma_h^2(j)$ is the squared sample correlation between $(I_j, I_{j+1}, \cdots, I_{j+M-1})$ and $(I_{j+h}, I_{j+h+1}, \cdots, I_{j+h+M-1})$. Then average the squared coherences over all frequencies for a fixed lag h via

$$\overline{\gamma}_h^2 = N^{-1} \sum_{j=0}^{N-1} \gamma_h^2(j).$$

Finally, plot $\overline{\gamma}_h^2$ against h. If values of $\overline{\gamma}_h^2$ which are larger than a pre-specified threshold appear only at lags which are multiples of $d = N/T$, one can conclude that $\{X_t\}$ is a periodic series with period T. The mentioned threshold can be approximated under the null hypothesis that $\{X_t\}$ is stationary via

$$\mathrm{P}(\overline{\gamma}_h^2 > x) \approx 1 - \Phi[(N/\kappa_M)^{\frac{1}{2}}(x - M^{-1})],$$

where Φ denotes the cumulative distribution function of the standard normal random variable, and κ_M is given by Lund *et al.* (1995) as a function of M.

2.2 Periodic Autoregressive Moving-Average Series

If periodicity is identified in the second moment of a time series, the next step is to find an adequate statistical model that incorporates all relevant information in the observations. Periodic autoregressive moving-average (PARMA) models can be used to model a large class of periodic series. The relationship between periodic series and PARMA models is akin to that between stationary series and ARMA models.

A mean zero series $\{X_t\}$ with finite second moments is said to be a periodic autoregressive moving-average series with period T if it is a solution to the periodic linear difference equation

$$X_{nT+\nu} - \sum_{k=1}^{p(\nu)} \phi_k(\nu) X_{nT+\nu-k} = \epsilon_{nT+\nu} + \sum_{k=1}^{q(\nu)} \theta_k(\nu) \epsilon_{nT+\nu-k}, \tag{2.2.1}$$

where $\{\epsilon_{nT+\nu}\}$ is mean zero periodic white noise: $\{\epsilon_{nT+\nu}\}$ is uncorrelated with $\mathrm{E}[\epsilon_{nT+\nu}] \equiv 0$ and $\mathrm{Var}(\epsilon_{nT+\nu}) = \sigma^2(\nu) > 0$ for all seasons ν. The periodic notation $X_{nT+\nu}$ refers to $\{X_t\}$ during the νth season, $1 \leq \nu \leq T$, of cycle n. The periodic notations $\{X_{nT+\nu}\}$, $\{\epsilon_{nT+\nu}\}$, $\{\phi_k(\nu)\}$, $\{\theta_k(\nu)\}$, etc. are preferred to the non-periodic notations $\{X_t\}$, $\{\epsilon_t\}$, $\{\phi_k(t)\}$, $\{\theta_k(t)\}$, etc. to emphasize periodicity.

Equation (2.2.1) is a generalized version of the classical ARMA difference equation with all model coefficients and orders varying periodically with time. The orders of the autoregressive and moving-average during season ν are respectively $p(\nu)$ and $q(\nu)$, and the model coefficients of the autoregressive and moving-average are respectively $\phi_1(\nu)$,...,$\phi_{p(\nu)}(\nu)$, and $\theta_1(\nu)$,...,$\theta_{q(\nu)}(\nu)$. For mathematical purposes, $p(\nu)$ and $q(\nu)$ can be taken as constant in ν — merely set $p = \max_{1 \leq \nu \leq T} p(\nu)$, $q = \max_{1 \leq \nu \leq T} q(\nu)$, and take $\phi_k(\nu) = 0$ for $k > p(\nu)$ and $\theta_k(\nu) = 0$ for $k > q(\nu)$. Then equation (2.2.1) becomes

$$X_{nT+\nu} - \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k} = \epsilon_{nT+\nu} + \sum_{k=1}^{q} \theta_k(\nu) \epsilon_{nT+\nu-k}. \tag{2.2.2}$$

Such simplification with model orders is inappropriate, however, when specific values of $p(\nu)$ and $q(\nu)$ are under investigation. A PARMA series satisfying (2.2.2) will also be referred to as a PARMA(p, q) series.

Two equivalent forms of (2.2.2) are useful when studying properties of PARMA models. One equivalent form of (2.2.2) is

$$\left(1 - \sum_{k=1}^{p} \phi_k(\nu) B^k\right) X_{nT+\nu} = \left(1 + \sum_{k=1}^{q} \theta_k(\nu) B^k\right) \epsilon_{nT+\nu}, \tag{2.2.3}$$

where B denotes the backward shift operator satisfying $B^k X_t = X_{t-k}$ for $k \geq 0$. Another equivalent form of (2.2.2) is the T-dimensional vector ARMA (VARMA) representation (cf. Vecchia 1985a)

$$\Phi_0 \underline{X}_n - \sum_{k=1}^{p^*} \Phi_k \underline{X}_{n-k} = \Theta_0 \underline{\epsilon}_n + \sum_{k=1}^{q^*} \Theta_k \underline{\epsilon}_{n-k}, \tag{2.2.4}$$

where $\underline{X}_n = (X_{nT+1}, \cdots, X_{nT+T})'$, $\underline{\epsilon}_n = (\epsilon_{nT+1}, \cdots, \epsilon_{nT+T})'$, $p^* = \lceil p/T \rceil$ and $q^* = \lceil q/T \rceil$. Here, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. The (i,j)th entries in the $T \times T$ matrices of $\{\Phi_k, 0 \leq k \leq p^*\}$ and $\{\Theta_k, 0 \leq k \leq q^*\}$ are given by (see Vecchia 1985a)

$$(\Phi_0)_{i,j} = \begin{cases} 1 & i = j \\ 0 & i < j \\ -\phi_{i-j}(i) & i > j, \end{cases} \tag{2.2.5}$$

$$(\Phi_k)_{i,j} = \phi_{kT+i-j}(i) \ , \ 1 \leq k \leq p^*, \tag{2.2.6}$$

$$(\Theta_0)_{i,j} = \begin{cases} 1 & i = j \\ 0 & i < j \\ \theta_{i-j}(i) & i > j, \end{cases} \tag{2.2.7}$$

and

$$(\Theta_k)_{i,j} = \theta_{kT+i-j}(i) \ , \ 1 \leq k \leq q^*. \tag{2.2.8}$$

Many results for PARMA series can be extracted from its VARMA representation. For example, Lund and Basawa (1999) provide a sufficient condition for PARMA causality. Specifically, let $\Phi(z) = \Phi_0 - \sum_{k=1}^{p^*} \Phi_k z^k$ be the T-variate VAR polynomial. When $\Phi(z)$ has no roots within or on the complex unit circle in the sense that $\det(\Phi(z)) \neq 0$ for all complex $|z| \leq 1$, then solutions to (2.2.2) exist and can be uniquely (in mean square) expressed in the infinite order moving-average form

$$X_{nT+\nu} = \sum_{k=0}^{\infty} \psi_k(\nu)\epsilon_{nT+\nu-k}, \tag{2.2.9}$$

where $\max_{1\leq\nu\leq T} \sum_{k=0}^{\infty} |\psi_k(\nu)| < \infty$ (uniqueness only requires that the VAR polynomial have no roots on the unit circle $|z| = 1$). Bentarzi and Hallin (1993) give further discussion and generalities.

The $\psi_k(\nu)$s can be computed from $\psi_0(\nu) \equiv 1$ and the recursion

$$\psi_k(\nu) = \theta_k(\nu)1_{[k\leq q]} + \sum_{j=1}^{\min(k,p)} \phi_j(\nu)\psi_{k-j}(\nu - j), \tag{2.2.10}$$

where $k \geq 1$, and $1 \leq \nu \leq T$ (cf. Lund and Basawa 2000). The notation used in (2.2.10) and elsewhere interprets parameters periodically with period T; for example, $\psi_k(j + T) = \psi_k(j)$ for each $k \geq 0$ and $\sigma^2(j + T) = \sigma^2(j)$.

We mention a generalization of the above: when the VAR polynomial has no roots on the unit circle, solutions to the PARMA difference equation are unique and periodically stationary, but may not be causal.

Let $\Theta(z) = \Theta_0 + \sum_{k=1}^{q^*} \Theta_k z^k$ denote the PARMA model's T-variate MA polynomial. A sufficient condition for PARMA invertibility is that $\Theta(z)$ has no roots within or on the complex unit circle (i.e. $\det(\Theta(z)) \neq 0$) for all complex $|z| \leq 1$. When invertible, solutions to (2.2.2) can be expressed in the infinite order periodic autoregressive form

$$\epsilon_{nT+\nu} = \sum_{k=0}^{\infty} \pi_k(\nu)X_{nT+\nu-k}, \tag{2.2.11}$$

where $\{\pi_k(\nu)\}$ satisfies

$$\max_{1\le\nu\le T}\sum_{k=0}^{\infty}|\pi_k(\nu)| < \infty.$$

PARMA series should not be confused with seasonal ARMA (SARMA) series. Generally speaking, SARMA series are stationary (not periodic) with strong correlations (large in absolute value) at lags which are multiples of the period T. Specifically, a SARMA series is a solution to the linear difference equation

$$X_{nT+\nu} - \sum_{k=1}^{p}\phi_k X_{(n-k)T+\nu} = \epsilon_{nT+\nu} + \sum_{k=1}^{q}\theta_k\epsilon_{(n-k)T+\nu}, \tag{2.2.12}$$

where $\{\epsilon_{nT+\nu}\}$ is white noise (stationary). Equation (2.2.12) is similar to the ARMA difference equation except that all differences appear at multiples of the period T. To see this more clearly, note that under the condition of causality, the solution to equation (2.2.12) can be uniquely (in mean square) expressed as

$$X_{nT+\nu} = \sum_{k=0}^{\infty}\psi_k\epsilon_{(n-k)T+\nu}, \tag{2.2.13}$$

where $\max_{1\le\nu\le T}\sum_{k=0}^{\infty}|\psi_k(\nu)| < \infty$. From equation (2.2.13), it follows that the same model applies to each season ν, and the autocovariance function is actually stationary; that is $\mathrm{Cov}(X_{nT+\nu}, X_{nT+\nu-h})$ depends only on h, but with $\mathrm{Cov}(X_t, X_{t+h})$ large when h is a multiple of T. In fact, SARMA series are characterized by ACVFs which are large at multiples of the period T. Consequently, the ACVF of a SARMA series exhibits 'spikes' at lags which are multiples of T. Although SARMA models are superior to ARMA models for describing some seasonal data, they are appropriate for stationary and are not adequate models for most periodic series. McLeod (1993) illustrates the inadequacies of SARMA models when forecasting periodic streamflow series. In contrast, PARMA series have a truly periodic autocovariance structure in the sense of (1.0.2). By letting the model parameters periodically vary with time, PARMA models have solutions which exhibit periodicity in the second moment.

2.3 Statistical Inference in PARMA Series

The PARMA model in (2.2.2) has $(p+q)T$ model parameters, which we denote by $\underline{\alpha} = (\phi_1(1),\cdots,\phi_p(1),\theta_1(1),\cdots,\ \theta_q(1);\cdots;\phi_1(T),\cdots,\phi_p(T),\theta_1(T),\ \cdots,\theta_q(T))'$, and T additional white noise variance parameters, denoted by $\underline{\sigma}^2 = (\sigma^2(1),\sigma^2(2),\cdots,\sigma^2(T))'$. Many PARMA estimation studies have been conducted (cf. Jones and Brelsford (1967), Pagano (1978), Li and Hui (1988), Vecchia (1983, 1985a and 1985b), Adams and Goodwin (1993), Lund and Basawa (1999, 2000)).

The following results pertain to estimating PARMA model parameters and their asymptotic variances; throughout, we treat $\underline{\sigma}^2$ as a T-dimensional nuisance parameter, and also make the following assumptions:

(C1). $\underline{X}_N = X_1, X_2, \cdots, X_N$ is a sample from a PARMA series containing d full cycles of data; that is, $d = N/T$ is a whole number.

(C2). The PARMA model is causal and invertible.

(C3). The orders of the autoregressive and moving-average in different seasons are constant; that is, $p(\nu) = p$ and $q(\nu) = q$ for all $1 \le \nu \le T$.

(C4). The periodic white noise $\{\epsilon_{nT+\nu}\}$ is independent and has a finite fourth moment: $\mathrm{E}[\epsilon^4_{nT+\nu}] < \infty$ for each season ν.

For PAR series, all parameters can be estimated by periodic versions of the Yule-Walker equations. Suppose that $\{X_{nT+\nu}\}$ is the periodic autoregression

$$X_{nT+\nu} - \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k} = \epsilon_{nT+\nu}. \tag{2.3.1}$$

Multiplying each side of (2.3.1) by $X_{nT+\nu-h}$, taking an expectation, and using causality to evaluate the right hand side, we obtain the so-called periodic Yule-Walker equations

$$\gamma_h(\nu) - \sum_{k=1}^{p} \phi_k(\nu)\gamma_{h-k}(\nu-k) = \sigma^2(\nu) I_{[h=0]}, \qquad h \ge 0. \tag{2.3.2}$$

The Yule-Walker estimates are defined as solutions to (2.3.2), viz.,

$$\hat{\gamma}_h(\nu) - \sum_{k=1}^{p} \hat{\phi}_k(\nu)\hat{\gamma}_{h-k}(\nu - k) = \hat{\sigma}^2(\nu) I_{[h=0]}, \quad \text{where } h = 0, 1, \cdots, p, \tag{2.3.3}$$

where $\hat{\gamma}_h(\nu)$ is the non-negative definite sample estimate

$$\hat{\gamma}_h(\nu) = d^{-1} \sum_{n=0}^{M} X_{nT+\nu} X_{nT+\nu-h}, \tag{2.3.4}$$

where $M = [d - max(\nu, \nu - h)]$ is such that the subscripts $nT + \nu$ and $nT + \nu - h$ in (2.3.4) all lie within $\{1, \cdots, dT\}$. Obviously $\underline{\hat{\alpha}} = (\hat{\phi}_1(1), \cdots, \hat{\phi}_1(T); \cdots; \hat{\phi}_p(1), \cdots, \hat{\phi}_p(T))'$ are moment estimates. The following result (Pagano, 1978) shows that the Yule-Walker estimates are consistent and asymptotically normal.

Theorem 2.1. *If $\{X_{nT+\nu}\}$ is a causal Gaussian PAR(p) series satisfying C1-C3, then the Yule-Walker estimates $\underline{\hat{\alpha}}= (\hat{\phi}_1(1),\cdots,\hat{\phi}_p(1);\cdots; \hat{\phi}_1(T),\cdots,\hat{\phi}_p(T))'$ satisfy*

$$d^{1/2}(\underline{\hat{\alpha}} - \underline{\alpha}) \xrightarrow{D} \mathcal{N}(\underline{0}, F^{-1}(\underline{\alpha})) \quad \text{as } d \to \infty, \tag{2.3.5}$$

where the $(pT) \times (pT)$ block diagonal information matrix $F(\underline{\alpha})$ has the form

$$F(\underline{\alpha}) = \begin{pmatrix} F(1) & 0 & \cdots & 0 & 0 \\ 0 & F(2) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & F(T-1) & 0 \\ 0 & 0 & \cdots & 0 & F(T) \end{pmatrix}, \tag{2.3.6}$$

where $F(\nu)$ $(1 \leq \nu \leq T)$ is a $p \times p$ matrix. The (i, j)th entry of $F(\nu)$ is defined by $F_{i,j}(\nu) = \sigma^{-2}(\nu)\gamma_{i-j}(\nu - i)$ for $1 \leq i, j \leq p$.

From the diagonal structure of $F(\underline{\alpha})$, we have the following result.

Corollary 2.2. *The Yule-Walker estimates, $\underline{\hat{\alpha}}$, are asymptotically independent in ν and asymptotically most efficient.*

Theorem 2.1 can be extended to non-Gaussian settings. The Yule-Walker estimates are not asymptotically most efficient in the non-Gaussian case, however.

For a general PARMA series with Gaussian periodic white noise $\{\epsilon_{nT+\nu}\}$, the model parameters can also be estimated by maximizing the Gaussian likelihood

$$L(\underline{\alpha}) = (2\pi)^{-dT/2} \left(\det(\Gamma_N)\right)^{-1/2} \exp\left(-\frac{1}{2}\underline{X}_N' \Gamma_N^{-1} \underline{X}_N\right), \tag{2.3.7}$$

or equivalently minimizing

$$-2\ln(L(\underline{\alpha})) = dT\ln(2\pi) + \ln(\det(\Gamma_N)) + \underline{X}_N' \Gamma_N^{-1} \underline{X}_N, \tag{2.3.8}$$

where Γ_N is the covariance matrix of $\underline{X}_N$. When the sample size is large, the computation of Γ_N^{-1} can be very intensive. Many authors have studied likelihood evaluation and optimization, including an approximation of the likelihood (Vecchia (1985a)), a Cholesky decomposition of Γ_N (Li and Hui (1988)), and a control approach to likelihood optimization altogether (Adams and Goodwin (1995)).

In Chapter 4, we will use the Innovations Algorithm (cf. Chapter 5 of Brockwell and Davis, 1991) to evaluate Gaussian likelihoods. The Innovations form of the likelihood is

$$L(\underline{\alpha}) = (2\pi)^{-dT/2} \left(\prod_{t=1}^{dT} v_t\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{dT} \frac{(X_t - \hat{X}_t)^2}{v_t}\right), \tag{2.3.9}$$

where

$$\hat{X}_t = \begin{cases} 0 & t = 1 \\ \mathrm{E}[X_t | X_1, \ldots, X_{t-1}] & t > 1 \end{cases} \tag{2.3.10}$$

is the best one-step-ahead linear predictor that minimizes the mean squared prediction error

$$v_t = \mathrm{E}[(X_t - \hat{X}_t)^2]. \tag{2.3.11}$$

$\hat{X}_t$ and v_t can be recursively computed in t for a general PARMA series quite efficiently (cf. Lund and Basawa, 2000). Such an algorithm drastically reduces the calculations and avoids computing the inverse of Γ_N. For PAR(p) series,

$$\hat{X}_{nT+\nu} = \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k}, \quad v_{nT+\nu} = \sigma^2(\nu), \quad nT+\nu > p; \tag{2.3.12}$$

hence, (2.3.8) reduces to

$$\begin{aligned} -2\ln(L(\underline{\alpha})) &= dT\ln(2\pi) + \ln(\det(\Gamma_p)) + \underline{X}_p'\Gamma_p^{-1}\underline{X}_p \\ &\quad + \sum_{t=p+1}^{dT} \ln(\sigma^2(t)) + \sum_{t=p+1}^{dT} \frac{(X_t - \hat{X}_t)^2}{\sigma^2(t)}, \end{aligned} \tag{2.3.13}$$

where $\underline{X}_p = (X_1, \ldots, X_p)'$ denotes the first p series observations and $\Gamma_p = \mathrm{E}[\underline{X}_p\underline{X}_p']$ is its covariance matrix. The covariance matrix Γ_p is easily computed from the seasonal Yule-Walker equations (see Chapter 3 here as well). Likelihood evaluation is still very demanding for general PARMA models with a moving-average component, however. Therefore, for a general PARMA series, Lund and Basawa (2000) present an efficient way to recursively calculate v_t in t. Specifically, define

$$W_{nT+\nu} = \begin{cases} X_{nT+\nu}, & nT+\nu \le \max(p,q), \\ X_{nT+\nu} - \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k}, & nT+\nu > \max(p,q). \end{cases} \tag{2.3.14}$$

It follows that

$$\hat{W}_{nT+\nu} = \hat{X}_{nT+\nu} - \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k},$$

where $\hat{W}_t = \mathrm{E}[W_t | W_1, \cdots, W_{t-1}]$. Therefore,

$$X_t - \hat{X}_t = W_t - \hat{W}_t \quad \text{for all } t,$$

and

$$v_t = \mathrm{E}[(W_t - \hat{W}_t)^2] \quad \text{for all } t.$$

Basawa and Lund (2001) establish the consistency and asymptotic normality of the likelihood estimates in the following theorem. Their methods compute the asymptotic variance of the estimates via general quasi-likelihood techniques. Not surprisingly, the limiting distributions of the likelihood estimates are equivalent to those obtained from weighted least squares techniques.

Theorem 2.3. *If $\{X_{nT+\nu}\}$ is a causal and invertible PARMA(p,q) series satisfying C1-C4, then the maximum likelihood estimates $\underline{\hat{\alpha}} = (\hat{\phi}_1(1),\cdots,\hat{\phi}_p(1),\hat{\theta}_1(1),$ $\cdots,\hat{\theta}_q(1);\cdots;\hat{\phi}_1(T),\cdots,\hat{\phi}_p(T),\hat{\theta}_1(T),\cdots,\ \hat{\theta}_q(T))'$, satisfy*

$$d^{\frac{1}{2}}(\underline{\hat{\alpha}} - \underline{\alpha}) \xrightarrow{D} \mathcal{N}(\underline{0}, F^{-1}(\underline{\alpha})) \quad \textit{as } d \to \infty, \tag{2.3.15}$$

where the information matrix $F(\underline{\alpha})$ is computed from

$$F(\underline{\alpha}) = \sum_{\nu=1}^{T} \sigma^{-2}(\nu)\Gamma(\nu), \tag{2.3.16}$$

and

$$\Gamma(\nu) = E\left[\left(\frac{\partial \epsilon_{nT+\nu}}{\partial \underline{\alpha}}\right)\left(\frac{\partial \epsilon_{nT+\nu}}{\partial \underline{\alpha}}\right)'\right], \tag{2.3.17}$$

where $\partial\epsilon_{nT+\nu}/\partial\underline{\alpha}$, the partial derivative of $\epsilon_{nT+\nu}$, satisfies

$$\frac{\partial \epsilon_{nT+\nu}}{\partial \underline{\alpha}} = -\sum_{k=1}^{p} \frac{\partial \phi_k(\nu)}{\partial \underline{\alpha}} - \sum_{k=1}^{q} \theta_k(\nu) \frac{\partial \epsilon_{nT+\nu-k}(\underline{\alpha})}{\partial\ ul\alpha} - \sum_{k=1}^{q} \frac{\partial \theta_k(\nu)}{\partial \underline{\alpha}} \epsilon_{nT+\nu-k}(\underline{\alpha}).$$

For a PAR(p) model with Gaussian white noise $\{\epsilon_{nT+\nu}\}$, Yule-Walker and likelihood estimates have equivalent asymptotic properties. However, likelihood estimates are superior to Yule-Walker estimates in the case of parametric constraints: likelihood estimates can easily be obtained if $\underline{\alpha}$ is a function of some 'smaller' set of free parameters $\underline{\beta}$. Therefore, likelihood estimators have an irreplaceable role in PARMA parsimony issues. We will revisit this issue in detail in Chapter 4.

Chapter 3

Calculation and Characterization of Parma Autocovariances and Partial Autocorrelations

Throughout this chapter, the PARMA model is assumed causal and the AR and MA model orders constant — say $p(\nu) \equiv p$ and $q(\nu) \equiv q$.

3.1 Introduction

The most fundamental feature of a time series model is its autocovariance structure. Autocovariances can be used to forecast future series values, to evaluate Gaussian-based model likelihoods, and to compute parameter estimates and their standard errors. Efficient computation of PARMA autocovariances, however, remains largely unexplored. Section 2 of this chapter develops a simple algorithm that efficiently computes general PARMA autocovariances.

The partial autocorrelation function also contains much useful diagnostic information. Sample partial autocorrelations can be used to assess whether an autoregressive model is appropriate for the series as well as estimate the autoregressive order (see Cryer 1986, Brockwell and Davis 1991, and Box *et al.* 1994 for stationary series and Vecchia and Ballerini 1991 for periodic series). Section 3 of this chapter considers definition and computation of the partial autocorrelation function in a general periodic series.

Section 4 of this chapter extends two classical stationary autocorrelation and partial autocorrelation characterizations to the periodic setting. Specifically, suppose that $\{X_t\}$ is a mean zero stationary series. Then $\{X_t\}$ is a moving-average of order q if and only if its autocovariances are zero at all lags that exceed q; analogously, $\{X_t\}$ is an autoregression of order p if and only if its partial autocorrelations are zero at all lags that exceed p. Theorems 3.1 and 3.3 establish periodic versions of these results. All plots are provided at the end of this chapter.

3.2 Computation of PARMA Autocovariances

Let

$$\gamma_h(\nu) = \mathrm{Cov}(X_{nT+\nu}, X_{nT+\nu-h}) \tag{3.2.1}$$

be the autocovariance of $\{X_t\}$ at season ν and lag $h \geq 0$. One form for the PARMA autocovariances is easily obtained by multiplying both sides of (2.2.9) by $X_{nT+\nu-h}$, using its causal expression in (2.2.9), and then taking an expectation:

$$\gamma_h(\nu) = \sum_{k=0}^{\infty} \psi_{k+h}(\nu)\psi_k(\nu-h)\sigma^2(\nu-k-h), \tag{3.2.2}$$

where both $\{\psi_k(\nu)\}$ for each $k \geq 0$ and $\{\sigma^2(\nu)\}$ are interpreted periodically in ν with period T. Equation (3.2.2) is computationally impractical since it requires determination and infinite summation of the $\psi_k(\nu)$s (explicit solutions to (2.2.10) are not frequently available). It is worth mentioning that PARMA autocovariances decay geometrically to zero (in lag) uniformly over the seasons whenever $\Phi(z)$ has no roots on the unit circle: $\max_{1\leq\nu\leq T} |\gamma_h(\nu)| \leq \kappa r^{-h}$ for all $h \geq 0$ and some $r > 1$ and finite κ. Hence, truncation approximations to (3.2.2) are feasible.

Instead, we will take a difference equation approach to compute PARMA autocovariances, mimicking Yule-Walker methods for stationary ARMA series. The contribution here lies largely with the bookkeeping. Multiplying both sides of (2.2.2) by

$X_{nT+\nu-h}$, taking expectations, and using causality to evaluate the right hand side gives

$$\gamma_h(\nu) = \sum_{k=1}^{p} \phi_k(\nu)\gamma_{k-h}(\nu - h), \qquad h > \max(p, q). \tag{3.2.3}$$

Equation (3.2.3) expresses $\gamma_h(\nu)$ in terms of autocovariances at the previous p lags when $h > \max(p, q)$ — note that the computational complexity of (3.2.3) does not increase with increasing h. In particular, once $\gamma_h(\nu)$ is identified for all lags $0 \le h \le \max(p, q)$ and seasons ν, the PARMA autocovariances at higher lags can be rapidly computed.

Hence we focus on computation of $\gamma_h(\nu)$ for $0 \le h \le \max(p, q)$ and all ν. Application of (2.2.2) gives

$$\begin{aligned} \mathrm{Cov}(X_{nT+\nu}, X_{nT+\nu-h}) &= \sum_{k=1}^{p} \phi_k(\nu)\mathrm{Cov}(X_{nT+\nu-k}, X_{nT+\nu-h}) \\ &+ \sum_{k=0}^{q} \theta_k(\nu)\mathrm{Cov}(\epsilon_{nT+\nu-k}, X_{nT+\nu-h}), \end{aligned} \tag{3.2.4}$$

where the convention $\theta_0(\nu) = 1$ has been used. The causal representation in (2.2.9) provides

$$\begin{aligned} \mathrm{Cov}(\epsilon_{nT+\nu-k}, X_{nT+\nu-h}) &= \mathrm{Cov}\left(\epsilon_{nT+\nu-k}, \sum_{l=0}^{\infty} \psi_l(\nu - h)\epsilon_{nT+\nu-h-l}\right) \\ &= \sum_{l=0}^{\infty} \psi_l(\nu - h)\sigma^2(\nu - k)1_{[l=k-h]} \\ &= \psi_{k-h}(\nu - h)\sigma^2(\nu - k)1_{[k\ge h]}. \end{aligned} \tag{3.2.5}$$

Combining (3.2.4) and (3.2.5) gives

$$\gamma_h(\nu) - \sum_{k=1}^{p} \phi_k(\nu)\mathrm{Cov}(X_{nT+\nu-k}, X_{nT+\nu-h}) = \sum_{k=h}^{q} \theta_k(\nu)\psi_{k-h}(\nu - h)\sigma^2(\nu - k), \tag{3.2.6}$$

which has the general form

$$\gamma_h(\nu) - \sum_{k=1}^{p} \phi_k(\nu)\gamma_{|h-k|}(s(\max(\nu - k, \nu - h))) = \kappa_h(\nu), \tag{3.2.7}$$

where $s(t) = t - T\lfloor t/T \rfloor$ is the season of time t ($\lfloor x \rfloor$ denotes the greatest integer less than or equal to x), and

$$\kappa_h(\nu) = \sum_{k=h}^{q} \theta_k(\nu)\psi_{k-h}(\nu - h)\sigma^2(\nu - k). \tag{3.2.8}$$

It is a straightforward (but tedious) task to write (3.2.7) into a $T[\max(p, q) + 1]$-dimensional linear system and numerically solve for $\gamma_h(\nu)$ for all $0 \leq h \leq \max(p, q)$ and $1 \leq \nu \leq T$. The matrix associated with this linear system will be invertible whenever the PARMA model is causal. Note that (3.2.7) only requires $\psi_k(\nu)$s for $k \leq q$, which are easily obtained from (2.2.10). The overall computational burden compares favorably to (3.2.2), which requires an infinite number of $\psi_k(\nu)$s.

Example 3.1. Consider the PARMA(2,1) model

$$X_{nT+\nu} - \phi_1(\nu)X_{nT+\nu-1} - \phi_2(\nu)X_{nT+\nu-2} = \epsilon_{nT+\nu} + \theta_1(\nu)\epsilon_{nT+\nu-1} \tag{3.2.9}$$

with period $T = 4$ and the parameters listed below.

PARMA(2,1) coefficients

ν	$\phi_1(\nu)$	$\phi_2(\nu)$	$\theta_1(\nu)$	$\sigma^2(\nu)$
1	0.8	0.1	0.50	1.0
2	0.2	0.7	0.30	9.0
3	-0.2	0.7	-0.30	9.0
4	-0.8	0.1	-0.50	1.0

It can be verified that these parameters induce a causal model. Equation (2.2.10) gives $\psi_0(\nu) = 1$ and $\psi_1(\nu) = \theta_1(\nu) + \phi_1(\nu)$.

Figure 3.1 plots the autocovariances of the above PARMA model for each of the four seasons. The structure of the autocovariance plots differ from season to season, yet series values from adjacent seasons are correlated. Figure 3.2 plots the autocorrelations for this PARMA model and can be regarded as a scaled version of

Figure 3.1. Notice that autocovariances and autocorrelations decay to zero rapidly with increasing lag, an aforementioned property of non-unit root PARMA models. □

3.3 Computation of PARMA Partial Autocorrelations

Define the partial autocorrelation (PACF) $\alpha_h(\nu)$ of a periodic series $\{X_t\}$ at season ν and lag $h \geq 1$ as the correlation between the residuals of $X_{nT+\nu}$ and $X_{nT+\nu-h}$ after linear regression on series values between times $nT+\nu-h$ and $nT+\nu$. Specifically, set $\alpha_1(\nu) = \mathrm{Corr}(X_{nT+\nu}, X_{nT+\nu-1})$ and

$$\alpha_h(\nu) = \mathrm{Corr}(X_{nT+\nu} - \hat{X}^F_{nT+\nu-h,nT+\nu}, X_{nT+\nu-h} - \hat{X}^B_{nT+\nu-h,nT+\nu}), \quad h \geq 2, \tag{3.3.1}$$

where $\hat{X}^F_{nT+\nu-h,nT+\nu} = \mathrm{P}(X_{nT+\nu}|X_{nT+\nu-h+1}, \ldots, X_{nT+\nu-1})$ is the best (minimum mean squared error) linear prediction (forwards) of $X_{nT+\nu}$ from $X_{nT+\nu-h+1}, \ldots,$ $X_{nT+\nu-1}$ and $\hat{X}^B_{nT+\nu-h,nT+\nu} = \mathrm{P}(X_{nT+\nu-h}|X_{nT+\nu-h+1}, \ldots, X_{nT+\nu-1})$ is the best linear prediction (backwards) of $X_{nT+\nu-h}$ from $X_{nT+\nu-h+1}, \ldots, X_{nT+\nu-1}$. Periodic stationarity of $\{X_{nT+\nu}\}$ implies that $\alpha_h(\nu)$ does not depend on n.

Several methods are plausible for numerical evaluation of the PARMA PACF. First, whereas multivariate ARMA techniques could be used to compute and unravel the matrix autocovariances of $\{\underline{X}_n\}$ ($\underline{X}_n = (X_{nT+1}, \cdots, X_{nT+T})'$) into those for the univariate series $\{X_t\}$, an analogous method to extract a PACF is not evident. In particular, the forward and backward predictions appearing in (3.3.1) are not easily obtained from predictions involving $\{\underline{X}_n\}$. For example, when $T \geq 3$, the prediction $\mathrm{P}(X_{nT+3}|X_{nT+2}, X_{nT+1}, X_{nT})$ is a function of X_{nT+2} and X_{nT+1}, which reside in the same cycle as X_{nT+3} — the variable being predicted. It is not possible to obtain $\mathrm{P}(X_{nT+3}|X_{nT+2}, X_{nT+1}, X_{nT})$ from $\mathrm{P}(\underline{X}_n|\underline{X}_{n-1}, \underline{X}_{n-2}, \ldots)$ as X_{nT+2} and X_{nT+1} are not in the closed linear span of the components of $\underline{X}_{n-1}$, $\underline{X}_{n-2}, \ldots$. For these reasons, we will not pursue multivariate PACF evaluation algorithms here. In general, multivariate and periodic time series differ structurally: the components of a multivariate time series may not be time-ordered whereas the components of a periodic time series are necessarily time-ordered.

Second, Sakai (1982, 1983) explores Durbin-Levinson methods for periodic series and derives a set of recursions from which the PARMA PACF can be evaluated. On the whole, some of the Durbin-Levinson recursions are comparatively cumbersome for PARMA PACF computation. Sakai's algorithm and the Innovations based algorithm below induce approximately the same computational burden, which is considerably less than the general recursive procedure of Morrison (1976) for computing partial autocorrelations. The Innovations based approach, however, yields a comparatively simplistic set of recursions cast in terms of backwards and forwards predictors and their mean squared errors.

For simplicity of notation, we consider the general case where $\{X_t\}$ is any mean zero series such that the covariance matrix of $(X_1, \ldots, X_k)'$ is invertible for each $k \geq 1$ (Proposition 4.1 of Lund and Basawa 2000 shows that this invertibility holds for any causal PARMA model). Fix $t \geq 2$ and let

$$\hat{X}^B_{t-i,t} = \mathrm{P}(X_{t-i} | X_{t-i+1}, \ldots, X_{t-1}), \quad i \geq 2, \tag{3.3.2}$$

be the best one-step-ahead backwards linear prediction of X_{t-i} from $X_{t-i+1}, \ldots, X_{t-1}$, with the convention that $\hat{X}^B_{t-1,t} = 0$. Let $v^B_{t-i,t} = \mathrm{E}[(X_{t-i} - \hat{X}^B_{t-i,t})^2]$ denote the backwards mean squared prediction error for $i \geq 1$. Note that $v^B_{t-1,t} = \mathrm{E}[X^2_{t-1}]$.

As the closed linear spans of $\{X_{t-i}, X_{t-i+1}, \ldots, X_{t-1}\}$ and $\{X_{t-i} - \hat{X}^B_{t-i,t}, X_{t-i+1} - \hat{X}^B_{t-i+1,t}, \ldots, X_{t-1} - \hat{X}^B_{t-1,t}\}$ are equal (also note that the latter set of random variables is uncorrelated), one is justified in making the Innovations expansion

$$\hat{X}^B_{t-i,t} = \sum_{j=1}^{i-1} \kappa_{i-1,j} (X_{t-j} - \hat{X}^B_{t-j,t}). \tag{3.3.3}$$

Multiply both sides of (3.3.3) by $X_{t-j} - \hat{X}^B_{t-j,t}$, take expectations, and use that $X_{t-j} - \hat{X}^B_{t-j,t}$ are uncorrelated in j to get

$$\begin{aligned} \kappa_{i-1,j} v^B_{t-j,t} &= \mathrm{E}[\hat{X}^B_{t-i,t}(X_{t-j} - \hat{X}^B_{t-j,t})] \\ &= \mathrm{E}[X_{t-i}(X_{t-j} - \hat{X}^B_{t-j,t})], \quad 1 \le j \le i-1. \end{aligned} \tag{3.3.4}$$

Substituting $\hat{X}^B_{t-j,t}$ with its Innovations expansion in (3.3.3) and applying (3.3.4) for $1 \le k \le j-1$ give

$$\kappa_{i-1,j} = \frac{\mathrm{E}[X_{t-i}X_{t-j}] - \sum_{k=1}^{j-1} \kappa_{j-1,k}\kappa_{i-1,k} v^B_{t-k,t}}{v^B_{t-j,t}}, \quad 1 \le j \le i-1, \tag{3.3.5}$$

which shows how to compute the $\kappa_{i-1,j}$s.

For updating backwards mean squared prediction errors, use the projection relation $v^B_{t-i,t} = \mathrm{E}[X^2_{t-i}] - \mathrm{E}[(\hat{X}^B_{t-i,t})^2]$, (3.3.3), and that $X_{t-j} - \hat{X}^B_{t-j,t}$ are uncorrelated in j to get

$$v^B_{t-i,t} = \mathrm{E}[X^2_{t-i}] - \sum_{j=1}^{i-1} \kappa^2_{i-1,j} v^B_{t-j,t}. \tag{3.3.6}$$

The forwards one-step-ahead linear predictor of X_t from $X_{t-i+1}, \ldots, X_{t-1}$ is

$$\hat{X}^F_{t-i,t} = \mathrm{P}(X_t | X_{t-i+1}, \ldots, X_{t-1}), \qquad i \ge 2, \tag{3.3.7}$$

where the convention $\hat{X}^F_{t-1,t} = 0$ is made. The forwards mean squared prediction errors will be denoted by $v^F_{t-i,t} = \mathrm{E}[(X_t - \hat{X}^F_{t-i,t})^2]$. Note that $v^F_{t-1,t} = \mathrm{E}[X^2_t]$.

An Innovations expansion for $\hat{X}^F_{t-i,t}$ in terms of the backwards prediction errors is

$$\hat{X}^F_{t-i,t} = \sum_{j=1}^{i-1} \eta_j (X_{t-j} - \hat{X}^B_{t-j,t}). \tag{3.3.8}$$

In contrast to the expansion in (3.3.3), the η_js do not depend on i due to the orthogonality of $X_{t-j} - \hat{X}^B_{t-j,t}$ in j and the fact that the variable being predicted (namely X_t) does not change with i.

Now argue as with the backwards predictors to get

$$\eta_j = \frac{\mathrm{E}[X_t X_{t-j}] - \sum_{k=1}^{j-1} \kappa_{j-1,k}\eta_k v_{t-k,t}^B}{v_{t-j,t}^B}, \quad 1 \le j \le i-1, \tag{3.3.9}$$

and

$$\begin{aligned} v_{t-i,t}^F &= \mathrm{E}[X_t^2] - \sum_{j=1}^{i-1} \eta_j^2 v_{t-j,t}^B \\ &= \mathrm{E}[X_t^2] - \sum_{j=1}^{i-2} \eta_j^2 v_{t-j,t}^B - \eta_{i-1}^2 v_{t-(i-1),t}^B \\ &= v_{t-(i-1),t}^F - \eta_{i-1}^2 v_{t-(i-1),t}^B. \end{aligned} \tag{3.3.10}$$

Putting the above together gives

$$\begin{aligned} &\mathrm{Corr}(X_t - \hat{X}_{t-h,t}^F, X_{t-h} - \hat{X}_{t-h,t}^B) \\ = &\ \frac{\mathrm{E}[(X_t - \hat{X}_{t-h,t}^F)X_{t-h}]}{\sqrt{v_{t-h,t}^B v_{t-h,t}^F}} \qquad (3.3.11) \\ = &\ \frac{\mathrm{E}[X_t X_{t-h}] - \sum_{j=1}^{h-1} \kappa_{h-1,j}\eta_j v_{t-j,t}^B}{\sqrt{v_{t-h,t}^B v_{t-h,t}^F}} \qquad (3.3.12) \end{aligned}$$

when the orthogonality of $\hat{X}_{t-h,t}^B$ and $X_t - \hat{X}_{t-h,t}^F$, (3.3.4), and (3.3.8) are applied. Specializing this to the PARMA case yields the following.

PARMA PACF Evaluation Algorithm.

Step 1. Compute the autocovariance and autocorrelation functions of the PARMA model as discussed in Section 2.

Step 2. For $h = 1$, set $\alpha_1(\nu) = \rho_1(\nu)$ for each season ν.

Step 3. To obtain $\alpha_h(\nu)$ for a fixed $h \ge 2$, define $n^* = \min\{n : nT + \nu - h \ge 1\}$ and set $t = n^*T + \nu$. Now solve the PACF recursions in the order $v_{t-1,t}^B$; $\kappa_{1,1}$, $v_{t-2,t}^B$, η_1, $v_{t-2,t}^F$; $\kappa_{2,1}$, $\kappa_{2,2}$, $v_{t-3,t}^B$, η_2, $v_{t-3,t}^F$; $\ldots$; $\kappa_{h-1,1}, \ldots, \kappa_{h-1,h-1}$, $v_{t-h,t}^B$, η_{h-1}, $v_{t-h,t}^F$.

Step 4. Compute $\alpha_h(\nu)$ via (3.3.12).

$\square$

The computational burden for computing the forwards and backwards prediction coefficients and their mean squared prediction errors above could possibly be further reduced by invoking the PARMA difference equation structure (akin to Section 5.3 in Brockwell and Davis for stationary ARMA models). We have not pursued this avenue for two reasons. First, one is typically interested in the PACF for small lags (say $h \leq 100$) and the above recursions are comparatively clean. Second, unlike the autocovariance function which may have to be evaluated many times to optimize a Gaussian based likelihood, the partial autocorrelation function is typically only evaluated once.

Example 3.2. Figure 3.3 plots the PACF of the PARMA(2,1) series in Example 3.1. Again notice that the PACF structure differs in season and that the partial autocorrelations decay to zero rapidly with increasing lag. □

3.4 Characterization of PMA Autocovariances

A stationary series $\{X_t\}$ is a moving average of order q if and only if $\gamma(h) = 0$ for all $h > q$ and $\gamma(q) \neq 0$ (cf. Proposition 3.2.1 in Brockwell and Davis, 1991). The condition $\gamma(q) \neq 0$ is needed to guarantee that the moving-average order is not strictly less than q. We next establish the periodic analogy of this result. Before this, we present several definitions and lemmas.

Definition 3.1. *The norm of a random variable X, denoted by $\| X \|$, is*

$$\| X \| = \sqrt{EX^2}.$$

Definition 3.2. *The orthogonal complement of a closed linear subset $\mathcal{M}$ of the Hilbert Space $\mathcal{H}$ is defined to be the set $\mathcal{M}^\perp$ of all elements of $\mathcal{H}$ which are orthogonal to every element of $\mathcal{M}$. Thus*

$$x \in \mathcal{M}^\perp (\text{ written } x \perp \mathcal{M}) \text{ if and only if } < x, y >= 0 \text{ for all } y \in \mathcal{M},$$

where the inner product is $<X, Y>= E[XY]$. *For all* $x \in \mathcal{H}$,

$$x = P_{\mathcal{M}}(x) + (I - P_{\mathcal{M}})(x), \tag{3.4.1}$$

where $P_{\mathcal{M}}(x)$ *is the projection of* x *onto* $\mathcal{M}$.

Lemma 3.1. *(The Projection Theorem). If* $\mathcal{M}$ *is a closed subspace of the Hilbert space* $\mathcal{H}$ *and* $x \in \mathcal{H}$, *then*
(i) there is a unique element $\hat{x} \in \mathcal{M}$ *such that*

$$\| x - \hat{x} \| = \inf_{y \in \mathcal{M}} \| x - y \|, \ \textit{and}$$

(ii) $\hat{x} \in \mathcal{M}$ *and* $\| x - \hat{x} \| = \inf_{y \in \mathcal{M}} \| x - y \|$ *if and only if* $\hat{x} \in \mathcal{M}$ *and* $(x - \hat{x}) \in \mathcal{M}^{\perp}$.

For proof, see page 51 Brockwell and Davis (1991).

The following theorem gives a necessary and sufficient condition for a series to be a periodic moving-average.

Theorem 3.1. *Assume that* $\{X_{nT+\nu}\}$ *is a mean zero periodically stationary series. Then* $\{X_{nT+\nu}\}$ *is a PMA(q) series if and only if* $\gamma_h(\nu) = 0$ *for* $h > q$ *and* $\gamma_q(\nu) \neq 0$ *for* $1 \leq \nu \leq T$.

Proof of Theorem 3.1. First, assume that $\gamma_h(\nu) = 0$ for $h > q$ and $\gamma_q(\nu) \neq 0$ for $1 \leq \nu \leq T$. Define the subspace $\mathcal{M}_t = \overline{sp}\{X_j, j \leq t\}$, and set

$$\epsilon_{nT+\nu} = X_{nT+\nu} - \mathrm{P}_{\mathcal{M}_{nT+\nu-1}} X_{nT+\nu}. \tag{3.4.2}$$

Obviously, $\epsilon_{nT+\nu} \in \mathcal{M}_{nT+\nu}$. By (3.4.2), for each fixed n and ν, $\epsilon_{nT+\nu} \in \mathcal{M}^{\perp}_{nT+\nu-1}$. Moreover for any $j > 0$,

$$\epsilon_{nT+\nu-j} \in \mathcal{M}_{nT+\nu-j} \subseteq \mathcal{M}_{nT+\nu-1}.$$

Therefore, $\mathrm{E}[\epsilon_{nT+\nu-j}\epsilon_{nT+\nu}] = 0$, and $\{\epsilon_{nT+\nu}\}$ is uncorrelated.

By the conclusion in problem 5.2(c) of Brockwell and Davis (1991), we have

$$\mathrm{P}_{\overline{sp}\{X_j,\ -\infty<j\leq nT+\nu\}} X_{nT+\nu+1} = \lim_{r\to\infty} \mathrm{P}_{\overline{sp}\{X_j,\ n-r<j\leq nT+\nu\}} X_{nT+\nu+1}.$$

Therefore, by continuity of norms,

$$\begin{aligned}
\| \epsilon_{(n+1)T+\nu} \|^2 &= \| X_{(n+1)T+\nu} - P_{\mathcal{M}_{(n+1)T+\nu-1}} X_{(n+1)T+\nu} \|^2 \\
&= \| X_{(n+1)T+\nu} - \lim_{r\to\infty} P_{\overline{sp}\{X_j,(n+1)T+\nu-1-r<j\leq(n+1)T+\nu-1\}} X_{(n+1)T+\nu} \|^2 \\
&= \lim_{r\to\infty} \| X_{(n+1)T+\nu} - P_{\overline{sp}\{X_j,(n+1)T+\nu-1-r<j\leq(n+1)T+\nu-1\}} X_{(n+1)T+\nu} \|^2 \\
&= \lim_{r\to\infty} \| X_{nT+\nu} - P_{\overline{sp}\{X_j,\ nT+\nu-1-r<j\leq nT+\nu-1\}} X_{nT+\nu} \|^2 \\
&= \| X_{nT+\nu} - P_{\mathcal{M}_{nT+\nu-1}} X_{nT+\nu} \|^2 \\
&= \| \epsilon_{nT+\nu} \|^2 .
\end{aligned}$$

Hence, $\| \epsilon_{nT+\nu} \|^2$ does not depend on n and we define the white noise variance $\sigma^2(\nu) = \| \epsilon_{nT+\nu} \|^2$. Since $\mathcal{M}_{nT+\nu-q-1} = \overline{sp}\{X_{nT+\nu-q-j}, j \leq 1\}$, then the causality of $\{X_{nT+\nu}\}$ gives

$$\mathcal{M}_{nT+\nu-q-1} \perp \overline{sp}\{\epsilon_{nT+\nu-1}, \epsilon_{nT+\nu-2}, \cdots, \epsilon_{nT+\nu-q}\}.$$

Since $\gamma_h(\nu) = 0$ for any $h > q$, we have $X_{nT+\nu} \perp \mathcal{M}_{nT+\nu-q-1}$. Therefore,

$$\begin{aligned}
\mathcal{M}_{nT+\nu-1} &= \overline{sp}\{X_{nT+\nu-1}, X_{nT+\nu-j}, j > 1\} \\
&= \overline{sp}\{X_{nT+\nu-j}, j > 1, \epsilon_{nT+\nu-1}\} \\
&= \overline{sp}\{X_{nT+\nu-j}, j > q, \epsilon_{nT+\nu-1}, \epsilon_{nT+\nu-2}, \cdots, \epsilon_{nT+\nu-q}\} \\
&= \mathcal{M}_{nT+\nu-q-1} + \overline{sp}\{\epsilon_{nT+\nu-1}, \epsilon_{nT+\nu-2}, \cdots, \epsilon_{nT+\nu-q}\}.
\end{aligned}$$

By (3.4.1), we have

$$\begin{aligned}
\mathrm{P}_{\mathcal{M}_{nT+\nu-1}} X_{nT+\nu} &= \mathrm{P}_{\mathcal{M}_{nT+\nu-q-1}} X_{nT+\nu} + \mathrm{P}_{\overline{sp}\{\epsilon_{nT+\nu-1},\epsilon_{nT+\nu-2},\cdots,\epsilon_{nT+\nu-q}\}} X_{nT+\nu} \\
&= 0 + \sigma^{-2}(\nu-1)\mathrm{E}[X_{nT+\nu}\epsilon_{nT+\nu-1}]\epsilon_{nT+\nu-1} + \cdots \\
&\quad + \sigma^{-2}(\nu-q)\mathrm{E}[X_{nT+\nu}\epsilon_{nT+\nu-q}]\epsilon_{nT+\nu-q} \\
&= \theta_1(\nu)\epsilon_{nT+\nu-1} + \cdots + \theta_q(\nu)\epsilon_{nT+\nu-q},
\end{aligned}$$

where $\theta_k(\nu) = \sigma^{-2}(\nu - k)\mathrm{E}[X_{nT+\nu}\epsilon_{nT+\nu-k}]$ which does not depend on n. Thus,

$$X_{nT+\nu} - \epsilon_{nT+\nu} = \theta_1(\nu)\epsilon_{nT+\nu-1} + \cdots + \theta_q(\nu)\epsilon_{nT+\nu-q},$$

i.e.,

$$X_{nT+\nu} = \epsilon_{nT+\nu} + \sum_{k=1}^{q} \theta_k(\nu)\epsilon_{nT+\nu-k}$$

which completes the proof of sufficiency.

Now, assume that $\{X_{nT+\nu}\}$ is a mean zero PMA(q) series. The autocovariance function for $h \geq 0$ and $1 \leq \nu \leq T$ is

$$\begin{aligned} \gamma_h(\nu) &= \mathrm{Cov}(X_{nT+\nu}, X_{nT+\nu-h}) \\ &= \mathrm{Cov}(\sum_{k=1}^{q} \theta_k(\nu)\epsilon_{nT+\nu-k}, \sum_{k=1}^{q} \theta_k(\nu-h)\epsilon_{nT+\nu-h-k}). \end{aligned} \tag{3.4.3}$$

Obviously, $\gamma_h(\nu) = 0$ for $h > q$, and $\gamma_h(\nu) \neq 0$ for $h \leq q$. This finishes the proof of necessity. □

3.5 Characterization of PAR Partial Autocorrelations

Our final result in this chapter shows that, essentially, periodic autoregressions are characterized by partial autocorrelations that are zero at lags that exceed a seasonal threshold (see Ramsey (1974) for the stationary autoregressive case). We first prove a theorem in the stationary setting, then extend the result to the periodically stationary case.

The following lemmas are given for later reference.

Lemma 3.2. *Assume that $X, Z_1, \cdots, Z_n$ are mean zero random variables with finite second moments. Let $\underline{\gamma}_n = (E[XZ_1], \cdots, E[XZ_n])'$ and $\Gamma_n = (E[Z_iZ_j])_{i,j=1,\cdots,n}$. If Γ_n is non-singular, then there exists a unique $\underline{\phi}_n = (\phi_1, \cdots, \phi_n)' = \Gamma_n^{-1}\underline{\gamma}_n$ such that*

$$P_{\overline{sp}\{Z_1,\cdots,Z_n\}}X = \sum_{i=1}^{n} \phi_i Z_i.$$

Proof. Let $\mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}X = \sum_{i=1}^{n} \phi_i Z_i$. By the Projection Theorem

$$\mathrm{E}[X - \sum_{i=1}^{n} \phi_i Z_i]Z_j = 0,$$

for all $1 \le j \le n$. Thus, $\underline{\phi}_n = \Gamma_n^{-1}\underline{\gamma}_n$. □

Lemma 3.3. *If $\{X_t\}$ is a mean zero stationary series, then*

(i) there exists $\underline{\phi}_n = (\phi_{n1}, \cdots, \phi_{nn})'$ free of t such that

$$P_{\overline{sp}\{X_n,\cdots,X_1\}}X_{n+1} = \sum_{i=1}^{n} \phi_{ni} X_i, \text{ and}$$

(ii) under the condition that $\gamma(h) \to 0$ as $h \to \infty$, $\underline{\phi}_n$ is unique and determined by

$$\underline{\phi}_n = \Gamma_n^{-1}\underline{\gamma}_n,$$

where $\Gamma_n = [\gamma(i-j)]_{i,j=1,\cdots,n}$ is nonsingular and $\underline{\gamma}_n = (\gamma(1), \cdots, \gamma(n))'$.

Proof. From Lemma 3.2, it suffices merely to verify the non-singularity of Γ_n. For the proof, see Proposition 5.1.1 in Brockwell and Davis (1991) (page 167). □

Lemma 3.4. *For any mean zero stationary series $\{X_t\}$, $E[X_t - P_{\overline{sp}\{X_{t-1},\cdots,X_{t-n}\}}X_t]^2$ is free from t.*

Proof. From stationarity of $\{X_t\}$ and Lemma 3.3 (i), there exits $\underline{\phi}_n = (\phi_{n1}, \cdots, \phi_{nn})'$ such that

$$\mathrm{E}[X_t - \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-n}\}}X_t]^2 = \mathrm{E}[X_t - \sum_{i=1}^{n} \phi_{ni} X_{t-i}]^2,$$

which does not depend on t. □

Lemma 3.5. *If X, Y, $Z_1, \cdots, Z_n$ are mean zero random variables such that*

$$Cov(X - P_{\overline{sp}\{Z_1,\cdots,Z_n\}}X, Y - P_{\overline{sp}\{Z_1,\cdots,Z_n\}}Y) = 0, \tag{3.5.1}$$

then

$$P_{\overline{sp}\{Z_1,\cdots,Z_n\}}X = P_{\overline{sp}\{Z_1,\cdots,Z_n,Y\}}X.$$

Proof. By the Projection Theorem, $\mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}Y \in \overline{sp}\{Z_1,\cdots,Z_n\}$ and $(X - \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}X) \perp \overline{sp}\{Z_1,\cdots,Z_n\}$. Hence $(X - \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}X) \perp \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}Y$. Therefore,

$$\mathrm{E}[(X - \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}X)\mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}Y] = 0. \tag{3.5.2}$$

By (3.5.1), we have

$$\begin{aligned} & \mathrm{E}[(X - \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}X)(Y - \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}Y)] \\ = \ & \mathrm{Cov}(X - \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}X, Y - \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}Y) \\ = \ & 0. \end{aligned} \tag{3.5.3}$$

Hence, by (3.5.2) and (3.5.3)

$$\mathrm{E}[(X - \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}X)Y] = 0.$$

Therefore, $(X - \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}X) \perp \overline{sp}\{Z_1,\cdots,Z_n,Y\}$. Obviously,

$$\mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}X \in \overline{sp}\{Z_1,\cdots,Z_n\} \subseteq \overline{sp}\{Z_1,\cdots,Z_n,Y\}.$$

By the Projection Theorem

$$\mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n\}}X = \mathrm{P}_{\overline{sp}\{Z_1,\cdots,Z_n,Y\}}X,$$

and the proof is complete. □

The following theorem was first proved by Ramsey (1974) and characterizes PACFs for AR(p)s.

Theorem 3.2. *Assume that $\{X_t\}$ is a mean zero causal stationary series with partial autocorrelation function $\alpha(.)$ and autocovariance function $\gamma(.)$ such that $\gamma(h) \to 0$ as $h \to \infty$. Then $\{X_t\}$ is an AR(p) series if and only if $\alpha(h) = 0$ for all $h > p$ and $\alpha(p) \neq 0$.*

Proof of Theorem 3.2. First, we assume that the PACF satisfies $\alpha(h) = 0$ for all $h > p$ and $\alpha(p) \neq 0$.

By stationarity and the Projection Theorem, there exists a unique $\underline{\phi}_p = (\phi_1, \cdots, \phi_p)'$ such that

$$\mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p}\}} X_t = \sum_{k=1}^{p} \phi_k X_{t-k}.$$

Let

$$\epsilon_t = X_t - \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p}\}} X_t = X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p}.$$

By Lemma 3.4, the variance of ϵ_t is constant in t. Set $\sigma^2 = \mathrm{Var}(\epsilon_t)$. From stationarity,

$$\begin{aligned}
& \mathrm{Corr}(X_t - \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p}\}} X_t, X_{t-p-1} - \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p}\}} X_{t-p-1}) \\
= \; & \mathrm{Corr}(X_{p+2} - \mathrm{P}_{\overline{sp}\{X_{p+1},\cdots,X_2\}} X_{p+2}, X_1 - \mathrm{P}_{\overline{sp}\{X_{p+1},\cdots,X_2\}} X_1) \\
= \; & \alpha(p+1) \\
= \; & 0.
\end{aligned}$$

Repeated use of Lemma 3.5 yields that for all $r \geq 1$

$$\mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p}\}} X_t = \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p},\cdots,X_{t-p-r}\}} X_t.$$

Hence we conclude that $\{\epsilon_t\}$ are uncorrelated since for any $i > j$,

$$\begin{aligned}
\mathrm{Cov}(\epsilon_i, \epsilon_j) &= \mathrm{E}[(X_i - \mathrm{P}_{\overline{sp}\{X_{i-1},\cdots,X_{i-p}\}} X_i)(X_j - \mathrm{P}_{\overline{sp}\{X_{j-1},\cdots,X_{j-p}\}} X_j)] \\
&= \mathrm{E}[(X_i - \mathrm{P}_{\overline{sp}\{X_{i-1},\cdots,X_{i-p},\cdots,X_{j-p}\}} X_i)(X_j - \mathrm{P}_{\overline{sp}\{X_{j-1},\cdots,X_{j-p}\}} X_j)] \\
&= 0,
\end{aligned}$$

by the Projection Theorem. Therefore $\{\epsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$, and the proof of sufficiency is complete.

Now, suppose $\{X_t\}$ is an AR(p) series. By the definition of causality and linearity of predictions, we can easily conclude that

$$\begin{aligned} X_t - \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p}\}}X_t &= X_t - \sum_{k=1}^{p} \phi_k X_{t-k}, \\ &= \epsilon_t, \end{aligned} \tag{3.5.4}$$

and for any $h > p$,

$$\begin{aligned} X_t - \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p},\cdots,X_{t-h}\}}X_t &= X_t - \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p}\}}X_t \\ &= \epsilon_t. \end{aligned} \tag{3.5.5}$$

Since $\epsilon_t \perp \overline{sp}\{X_{t-1}, \cdots, X_{t-h}\}$,

$$\begin{aligned} \alpha(h) &= \mathrm{Cov}(X_t - \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-h}\}}X_t, X_{t-h-1} - \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-h}\}}X_{t-h-1}) \\ &= \mathrm{E}[\epsilon_t(X_{t-h-1} - \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-h}\}}X_{t-h-1})] \\ &= 0. \end{aligned} \tag{3.5.6}$$

Suppose that $\alpha(p) = 0$ and that $h_0 < p$ is the largest integer such that $\alpha(h_0) \neq 0$. Then by the sufficiency argument above, $\{X_t\}$ is an AR(h_0) series which contradicts that $\{X_t\}$ is an AR(p) series. Therefore, for a causal AR(p) series, $\alpha(h) = 0$ for all $h > p$ and $\alpha(p) \neq 0$. This completes the necessity. □

The following result generalizes Theorem 3.2 to periodic autoregressions.

Theorem 3.3. *Assume that $\{X_{nT+\nu}\}$ is a mean zero periodically stationary series with autocovariance $\gamma_h(\nu)$ and partial autocorrelation $\alpha_h(\nu)$ at season ν and lag h such that $\Gamma_{m,\nu} = [\gamma_{i-j}(\nu - i)]_{i,j=1,\cdots,m}$ is nonsingular for all ν and $m \geq 1$. Then $\{X_{nT+\nu}\}$ is a PAR(p) series if and only if $\alpha_h(\nu) = 0$ for all $h > p$ and $\alpha_p(\nu) \neq 0$ for $1 \leq \nu \leq T$.*

Proof of Theorem 3.3. First, assume that $\alpha_h(\nu) = 0$ for any $h > p$ and $1 \leq \nu \leq T$, and that $\alpha_p(\nu) \neq 0$. Let

$$\epsilon_{nT+\nu} = X_{nT+\nu} - \mathrm{P}_{\overline{sp}\{X_{nT+\nu-1},\cdots,X_{nT+\nu-p}\}}X_{nT+\nu}.$$

From Lemma 3.2, there exists a unique $\underline{\phi}(\nu) = (\phi_1(\nu), \cdots, \phi_p(\nu))'$ for each ν such that

$$\mathrm{P}_{\overline{sp}\{X_{nT+\nu-1},\cdots,X_{nT+\nu-p}\}} X_{nT+\nu} = \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k}$$

holds for all n and

$$\sigma^2(\nu) := \mathrm{E}[\epsilon_{nT+\nu}]^2 = \mathrm{E}[X_{nT+\nu} - \mathrm{P}_{\overline{sp}\{X_{nT+\nu-1},\cdots,X_{nT+\nu-p}\}} X_{nT+\nu}]^2$$

is free from n by periodic stationarity. Repeated use of Lemma 3.5 yields that

$$\mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p}\}} X_t = \mathrm{P}_{\overline{sp}\{X_{t-1},\cdots,X_{t-p},\cdots,X_{t-p-r}\}} X_t$$

for all $r \geq 1$. Hence, for any $i > j$

$$\begin{aligned} \mathrm{E}[\epsilon_i \epsilon_j] &= \mathrm{E}[(X_i - \mathrm{P}_{\overline{sp}\{X_{i-1},\cdots,X_{i-p}\}} X_i)(X_j - \mathrm{P}_{\overline{sp}\{X_{j-1},\cdots,X_{j-p}\}} X_j)] \\ &= \mathrm{E}[(X_i - \mathrm{P}_{\overline{sp}\{X_{i-1},\cdots,X_{i-p},\cdots,X_{j-p}\}} X_i)(X_j - \mathrm{P}_{\overline{sp}\{X_{j-1},\cdots,X_{j-p}\}} X_j)] \\ &= 0 \end{aligned}$$

by the orthogonality property of projections. Therefore, $\{\epsilon_{nT+\nu}\}$ is periodic white noise and the proof of sufficiency is complete.

Now, suppose that $\{X_{nT+\nu}\}$ is a PAR(p) series. By the definition of causality and linearity of predictions, we have

$$\mathrm{P}_{\overline{sp}\{X_{nT+\nu-1},\cdots,X_{nT+\nu-p}\}} X_{nT+\nu} = \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k},$$

and for any $h > p$,

$$\mathrm{P}_{\overline{sp}\{X_{nT+\nu-1},\cdots,X_{nT+\nu-p},\cdots,X_{nT+\nu-h}\}} X_{nT+\nu} = \mathrm{P}_{\overline{sp}\{X_{nT+\nu-1},\cdots,X_{nT+\nu-p}\}} X_{nT+\nu}.$$

Therefore, for any $h > p$,

$$\hat{X}^F_{t-h,t} X_{t-h} = \hat{X}^F_{t-p-1,t} X_{t-h}, \tag{3.5.7}$$

and hence

$$
\begin{aligned}
\mathrm{E}[(X_t - \hat{X}^F_{t-h,t})X_{t-h}] &= \mathrm{E}[(X_t - \hat{X}^F_{t-p-1,t})X_{t-h}] \\
&= \mathrm{E}[\epsilon_t X_{t-h}] \\
&= 0. \qquad (3.5.8)
\end{aligned}
$$

So by (3.3.11), $\alpha_h(\nu) = 0$ for $h > p$.

Suppose that $\alpha_p(\nu) = 0$ and that $h_0 < p$ is the largest integer such that $\alpha_{h_0}(\nu) \neq 0$. Then by the sufficiency argument above, $\{X_{nT+\nu}\}$ is a PAR(h_0) series which contradicts that $\{X_{nT+\nu}\}$ is a PAR(p) series. Therefore, for a causal PAR(p) series, $\alpha_h(\nu) = 0$ for all $h > p$ and $\alpha_p(\nu) \neq 0$. $\square$

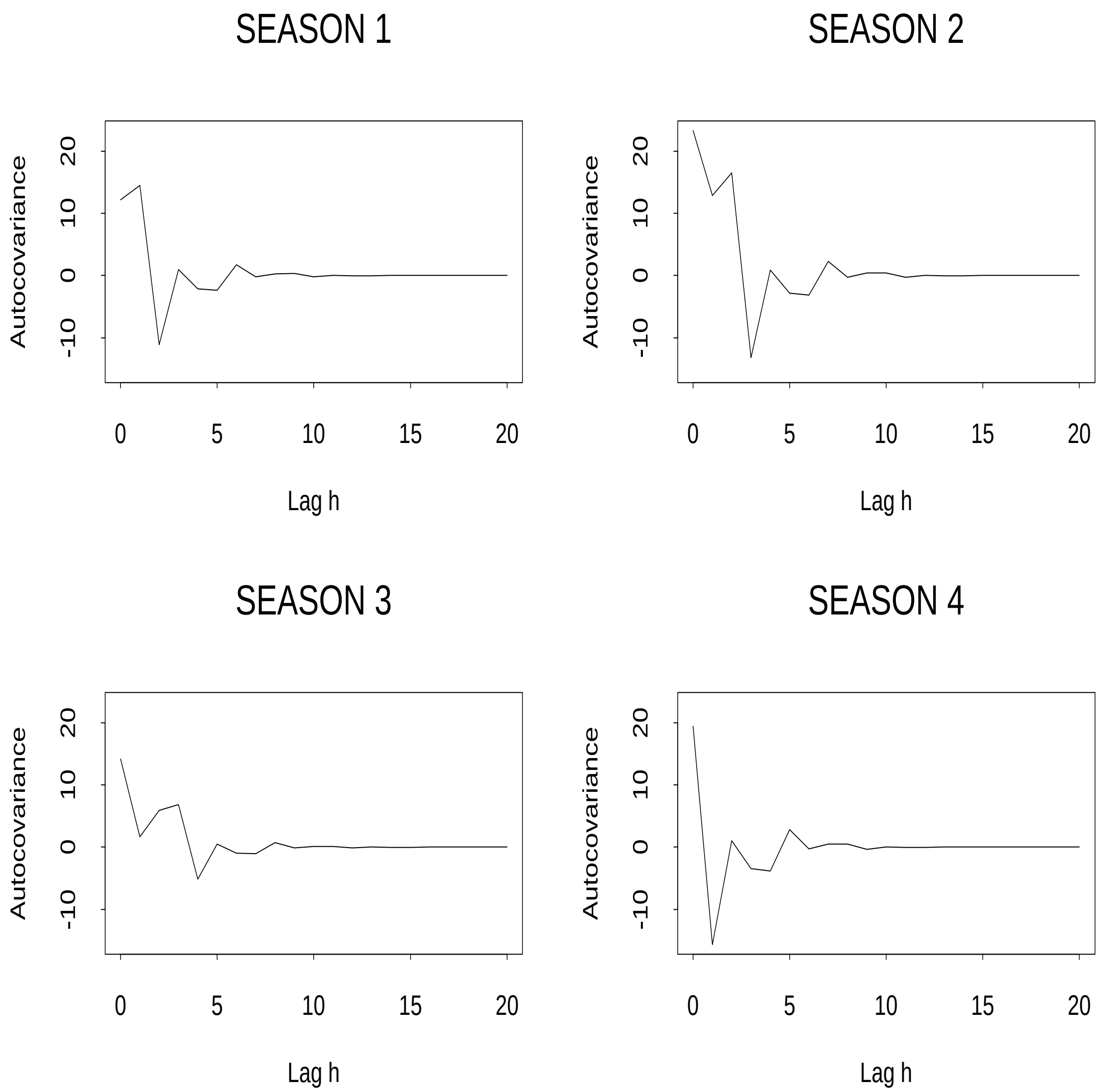
SEASON 1
Autocovariance
-10
0
10
20
0
5
10
15
20
Lag h
SEASON 2
Autocovariance
-10
0
10
20
0
5
10
15
20
Lag h
SEASON 3
Autocovariance
-10
0
10
20
0
5
10
15
20
Lag h
SEASON 4
Autocovariance
-10
0
10
20
0
5
10
15
20
Lag h

Figure 3.1:

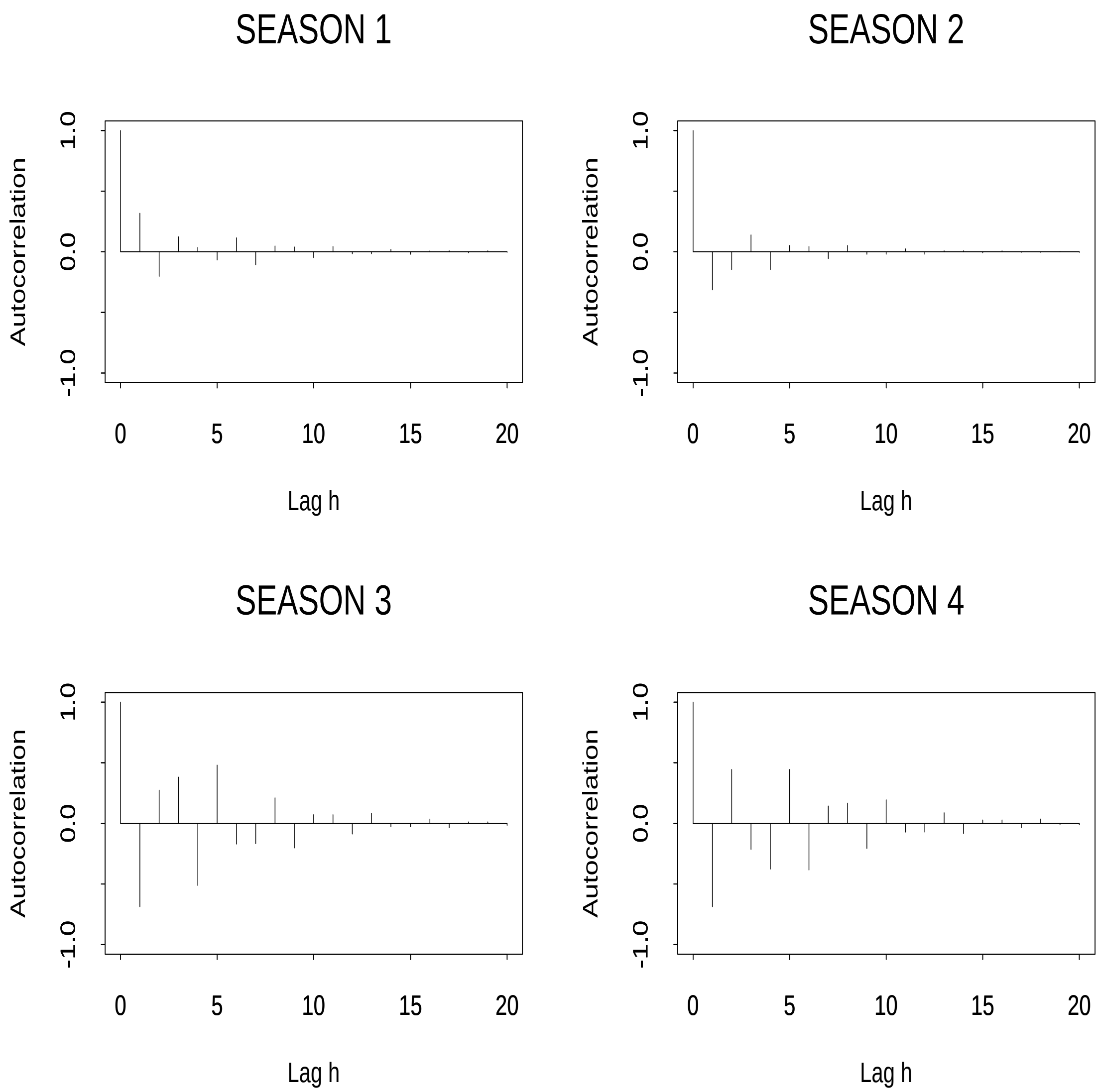
SEASON 1
Autocorrelation
1.0
0.0
-1.0
0
5
10
15
20
Lag h
SEASON 2
Autocorrelation
1.0
0.0
-1.0
0
5
10
15
20
Lag h
SEASON 3
Autocorrelation
1.0
0.0
-1.0
0
5
10
15
20
Lag h
SEASON 4
Autocorrelation
1.0
0.0
-1.0
0
5
10
15
20
Lag h

Figure 3.2:

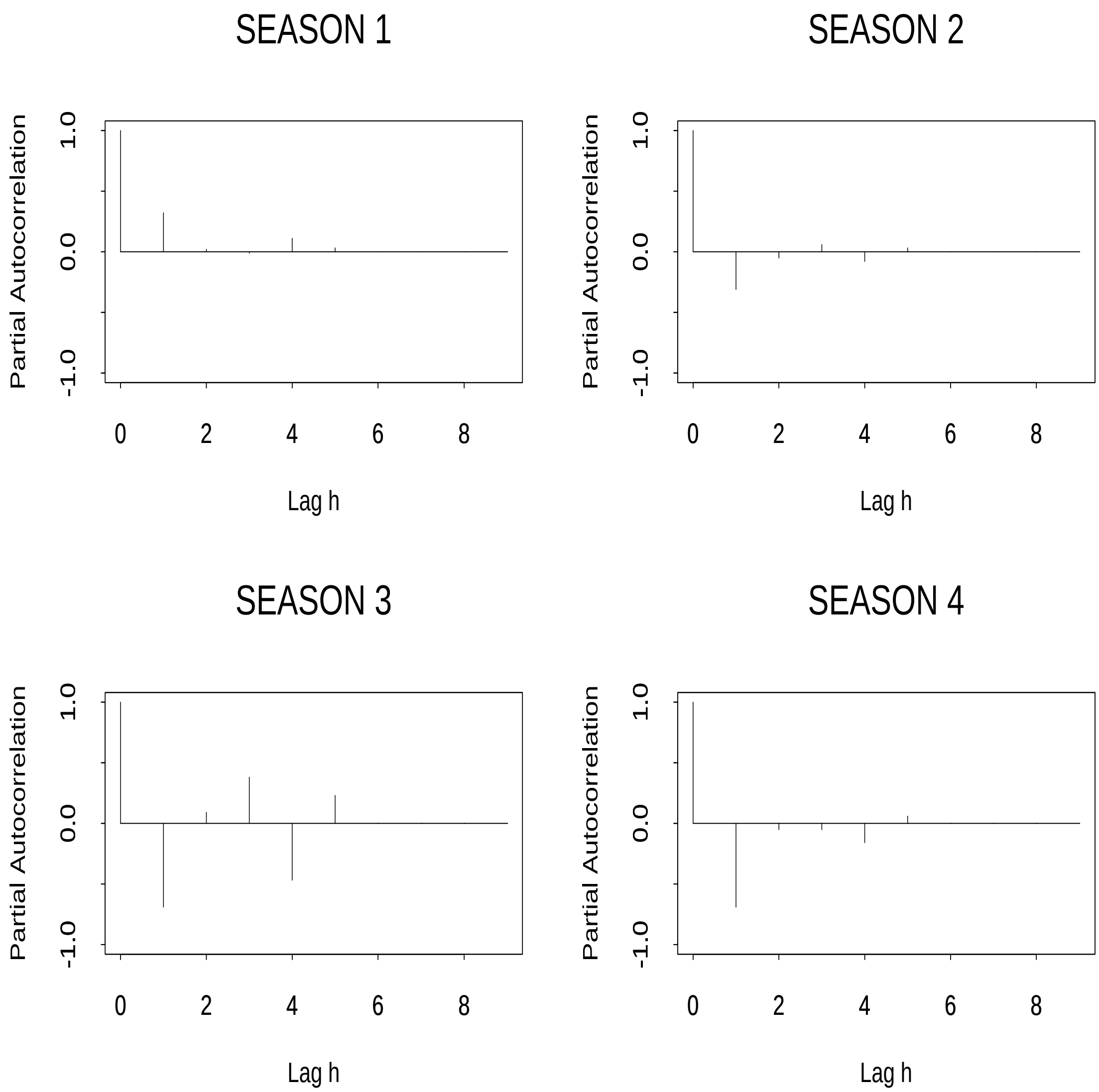
SEASON 1
Partial Autocorrelation
-1.0
0.0
1.0
0
2
4
6
8
Lag h
SEASON 2
Partial Autocorrelation
-1.0
0.0
1.0
0
2
4
6
8
Lag h
SEASON 3
Partial Autocorrelation
-1.0
0.0
1.0
0
2
4
6
8
Lag h
SEASON 4
Partial Autocorrelation
-1.0
0.0
1.0
0
2
4
6
8
Lag h

Figure 3.3:

Chapter 4

Large Sample Parsimonious PARMA Inference

4.1 Introduction

Modeling methods for periodic series focus on the class of periodic autoregressive moving-average (PARMA) models as the PARMA second moment structure is indeed periodic. A ubiquitous problem in fitting a PARMA model to a periodic series, however, lies with parsimony. Even very simple PARMA models can have an inordinately large number of parameters. A first order periodic autoregression for daily data, for example, has 365 autoregressive parameters — typically far more than this number is necessary for an adequate statistical description of the series. This chapter presents results aimed at parsimonious PARMA model development.

The PARMA model in (2.2.2) has $(p+q)T$ autoregressive and moving-average parameters and T additional white noise variance parameters. This parameter total can be large for even moderate T, making some PARMA inference matters unwieldy. For example, it would be a computationally intensive task to numerically optimize a Gaussian-based likelihood in the 365 autoregressive parameters to fit a periodic first order autoregression (PAR(1)) to daily data. Consequently, many authors have investigated parsimonious versions of (2.2.2). Thompstone *et al.* (1985) suggested grouping similar seasons into blocks to reduce parameter totals; alternatively, Hannan (1955), Jones and Brelsford (1967) and Bloomfield *et al.* (1994) consider modeling slow seasonal changes in parameters with Fourier series. For example, in the analysis of a monthly ozone series, Bloomfield *et al.* (1994) found that the

12 parameters in a monthly representation for $\phi_1(\nu)$ in a PAR(1) model could be statistically consolidated into the three parameters R_0, R_1, and τ_1 via

$$\phi_1(\nu) = R_0 + R_1 \cos\left(\frac{2\pi(\nu - \tau_1)}{T}\right). \tag{4.1.1}$$

Wavelet based expansions could be considered if changes between adjacent seasons are more abrupt. A third technique of parsimonious PARMA modeling constrains all off-diagonal matrix parameters in the PARMA model's T-variate ARMA representation (Vecchia 1985a provides the construction) to be zero. See Bartolini *et al.* (1988) and Rasmussen *et al.* (1996) for further discussion of these 'contemporaneous' models.

Parameter estimation and large sample inference for unconstrained PARMA models has now been well explored from several viewpoints. Pagano (1978) and Troutman (1979) first studied moment estimates for periodic autoregressions and established their consistency and asymptotic efficiency for Gaussian series; Lund and Basawa (2000) extend these results to weighted and unweighted least squares and maximum likelihood estimation, and also consider PARMA models with a moving average component. Parzen and Pagano (1979) and Dunsmuir (1981) focus on estimation of the seasonal means and standard deviations of a periodic series.

Inference for PARMA models with parametric constraints such as those in (4.1.1) have not been directly investigated. For example, an explicit standard error for the estimate (say maximum likelihood) of R_1 in (4.1.1) allows one to test the hypothesis that $\phi_1(\nu)$ is constant in the season ν. It is the purpose of this chapter to consider such problems.

The rest of this chapter proceeds as follows. Section 2 introduces a series of daily temperatures from Griffin, Georgia (GA), whose analysis motivated this study. Section 3 presents a large sample result for parsimonious PARMA model fitting. A likelihood ratio hypothesis test is also investigated there. Section 4 explores the computational mathematics of several simple PARMA models, and Section 5 concludes with a study of the Griffin series. All plots are provided at the end of this chapter.

4.2 Griffin Temperatures

Figure 4.1 plots 67 years of daily temperatures from Griffin, GA observed from Jan 1 1931 through Dec 31 1997 inclusive. Leap year (Feb 29) temperatures were ignored so that $T = 365$. Inclusion of leap year data does not change any of our inferences appreciably. Figure 4.1 reveals that the first two moments of the Griffin data are periodic with summertime temperatures being warmer and less variable than wintertime temperatures. The seasonal cycle in variance can be seen by comparing the 'jaggedness' of peaks and troughs in Figure 4.1. The first two graphs in Figure 4.2 present plots of the series sample means and variances by day and graphically confirm this claimed seasonality.

Let $\gamma_h(\nu) = \mathrm{Cov}(X_{nT+\nu}, X_{nT+\nu-h})$ denote the autocovariance of $\{X_t\}$ at lag $h \geq 0$ and season ν. The bottom plot in Figure 4.2 displays Griffin sample autocovariances by day of year at the lag $h = 1$. These estimates show a similar seasonal structure to that of the sample variances. Autocovariances are estimated from the non-negative definite sample average

$$\hat{\gamma}_h(\nu) = d^{-1} \sum_{n=0}^{d-1} X_{nT+\nu} X_{nT+\nu-h}, \tag{4.2.1}$$

where the employed notation assumes d full cycles of observed data. The number of observed data points is then dT and $\{X_t\}$ is observed at the times $t = 1, \ldots, dT$. The summation limits in (4.2.1) are written more precisely, for a fixed ν and h, as all $n \geq 0$ such that $nT + \nu$ and $nT + \nu - h$ lie within $\{1, \ldots, dT\}$.

Let $\rho_h(\nu) = \mathrm{Corr}(X_{nT+\nu}, X_{nT+\nu-h})$ denote the autocorrelation of $\{X_t\}$ at lag $h \geq 0$ and season ν. Sample versions of the autocorrelations are computed via

$$\hat{\rho}_h(\nu) = \frac{\hat{\gamma}_h(\nu)}{\sqrt{\hat{\gamma}_0(\nu)\hat{\gamma}_0(\nu-h)}}. \tag{4.2.2}$$

The top graph in Figure 4.3 plots the lag one sample autocorrelations by day of year (the ordinate scale on this graph is labelled as 'PAR(1) autoregressive estimate', but the two quantities are equivalent; see (4.5.2) below). From this plot, it is not clear whether the lag one autocorrelations are constant in season, or if there is further structure. Later, we will investigate statistical tests for such hypotheses.

4.3 Large Sample Parsimonious PARMA Inference

Suppose that the $(p+q)T$ PARMA autoregressive and moving average parameters, denoted by $\underline{\alpha}$, are functions of an l dimensional parameter vector $\underline{\beta}$. The dimension l can be considerably smaller than $(p+q)T$. The notations $\phi_k(\nu;\underline{\beta})$ and $\theta_k(\nu;\underline{\beta})$ will explicitly emphasize dependence of the autoregressive and moving-average parameters on $\underline{\beta}$. The T white noise variances, $\sigma^2(\nu), 1 \leq \nu \leq T$, will be treated as nuisance parameters in the exposition below. We now give two examples of PARMA parametric 'consolidations' (constraints) that appear interesting.

Example 4.1. Fourier parameter consolidations can be written in the form

$$\phi_m(\nu) = A_{0,m} + \sum_{k=1}^{r}\left[B_{k,m}\sin\left(\frac{2\pi k\nu}{T}\right) + A_{k,m}\cos\left(\frac{2\pi k\nu}{T}\right)\right], \quad 1 \leq m \leq p; \tag{4.3.1}$$

$$\theta_m(\nu) = C_{0,m} + \sum_{k=1}^{s}\left[D_{k,m}\sin\left(\frac{2\pi k\nu}{T}\right) + C_{k,m}\cos\left(\frac{2\pi k\nu}{T}\right)\right], \quad 1 \leq m \leq q, \tag{4.3.2}$$

where $A_{k,m}, B_{k,m}, C_{k,m}$, and $D_{k,m}$ are Fourier coefficients. There are $(2r+1)p + (2s+1)q$ free parameters in (4.3.1) and (4.3.2) which may be considerably smaller than $(p+q)T$ for small r and s. One could have r and s depend on m, but we do not pursue such generality here. Equations (4.3.1) and (4.3.2) are equivalently reparameterized as

$$\phi_m(\nu) = R_{0,m} + \sum_{k=1}^{r} R_{k,m} \cos\left(\frac{2\pi k(\nu - \tau_{k,m})}{T}\right), \quad 1 \leq m \leq p; \tag{4.3.3}$$

$$\theta_m(\nu) = S_{0,m} + \sum_{k=1}^{s} S_{k,m} \cos\left(\frac{2\pi k(\nu - \eta_{k,m})}{T}\right), \quad 1 \leq m \leq q. \tag{4.3.4}$$

Here, $R_{k,m}$ and $S_{k,m}$ are amplitudes and $\tau_{k,m}$ and $\eta_{k,m}$ are phase shifts. $\square$

Example 4.2. Consider a manufacturing process which operates with different dynamics on weekend days and week days. A simple first order autoregression describing this situation, also considered in Bibi and Francq (2000), is

$$X_{nT+\nu} - \phi_1(\nu) X_{nT+\nu-1} = \epsilon_{nT+\nu}, \tag{4.3.5}$$

where $\phi_1(\nu) = \beta_1$ if $1 \leq \nu \leq 5$ and $\phi_1(\nu) = \beta_2$ if $6 \leq \nu \leq 7$. The period is $T = 7$ and our accounting tracks Mondays as $\nu = 1$ and Sundays as $\nu = 7$. $\square$

Estimates of $\underline{\beta}$ can be computed under several estimation paradigms. A likelihood estimate of $\underline{\beta}$ is calculated as the arguments that maximize the Gaussian likelihood

$$L(\underline{\beta}) = (2\pi)^{-dT/2} \left(\prod_{t=1}^{dT} v_t\right)^{-1/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{dT} \frac{(X_t - \hat{X}_t)^2}{v_t}\right), \tag{4.3.6}$$

where the innovations form of the likelihood (cf. Chapter 8 of Brockwell and Davis, 1991) has been invoked. In particular,

$$\hat{X}_t = E[X_t | X_1, \ldots, X_{t-1}], \quad t \geq 2 \tag{4.3.7}$$

is the best one-step-ahead linear predictor (take $\hat{X}_1 = 0$) that minimizes the mean squared prediction error

$$v_t = E[(X_t - \hat{X}_t)^2]. \tag{4.3.8}$$

Lund and Basawa (2000) present an algorithm that efficiently computes $\hat{X}_t$ and v_t recursively in t for a general PARMA series. In the case of periodic autoregressions of order p (PAR(p)),

$$X_{nT+\nu} - \sum_{k=1}^{p} \phi_k(\nu) X_{nT+\nu-k} = \epsilon_{nT+\nu}, \tag{4.3.9}$$

the one-step-ahead predictions and mean squared errors take the simple form

$$\hat{X}_{nT+\nu} = \sum_{k=1}^{p} \phi_k(\nu;\beta) X_{nT+\nu-k}, \quad v_{nT+\nu} = \sigma^2(\nu), \quad nT+\nu > p; \tag{4.3.10}$$

hence, (4.3.6) reduces to

$$\begin{aligned} -2\ln(L(\underline{\beta})) &= dT\ln(2\pi) + \ln\left(\det(\Gamma_p)\right) + \underline{X}_p' \Gamma_p^{-1} \underline{X}_p \\ &+ \sum_{t=p+1}^{dT} \ln\left(\sigma^2(t)\right) + \sum_{t=p+1}^{dT} \frac{(X_t - \hat{X}_t)^2}{\sigma^2(t)}, \end{aligned} \tag{4.3.11}$$

where $\underline{X}_p = (X_1, \ldots, X_p)'$ denotes the first p series observations and $\Gamma_p = E[\underline{X}_p \underline{X}_p']$ is its covariance matrix. The covariance matrix Γ_p can be computed from the seasonal Yule-Walker equations (cf. Pagano 1978 and Section 4 here). The exact likelihood above compares quite favorably, in terms of computational demands, to the methods of Li and Hui (1988) or the approximate likelihood in Vecchia (1985b).

For PARMA models with a moving average component, $\underline{\beta}$ can be estimated as a solution to the l-dimensional optimal estimating equation $S_d(\underline{\beta}) = \underline{0}_{(l\times 1)}$ where

$$S_d(\underline{\beta}) = \sum_{n=0}^{d-1} \sum_{\nu=1}^{T} \frac{\epsilon^*_{nT+\nu}(\underline{\beta})}{\sigma^2(\nu)} \frac{\partial \epsilon^*_{nT+\nu}(\underline{\beta})}{\partial \underline{\beta}}. \tag{4.3.12}$$

Here, $\epsilon_t^*(\underline{\beta})$, an estimate of ϵ_t in (2.2.2), is computed recursively in t via

$$\epsilon^*_{nT+\nu}(\underline{\beta}) = X_{nT+\nu} - \sum_{k=1}^{p} \phi_k(\nu;\underline{\beta}) X_{nT+\nu-k} - \sum_{k=1}^{q} \theta_k(\nu;\underline{\beta}) \epsilon^*_{nT+\nu}(\nu;\underline{\beta}) \tag{4.3.13}$$

with the convention that $X_t = \epsilon_t^* = 0$ for $t \leq 0$. The derivatives $\partial \epsilon_t^*(\underline{\beta})/\partial \underline{\beta}$ in (4.3.12) are easily computed (recursively in t) by differentiating (4.3.13). We tacitly assume that the first order partial derivatives of $\phi_k(\nu;\underline{\beta})$ and $\theta_k(\nu;\underline{\beta})$ exist.

Using general quasilikelihood techniques, Basawa and Lund (2001) establish the consistency and asymptotic normality of the above maximum likelihood and optimal estimating equation estimates of $\underline{\beta}$ and identify the asymptotic covariance matrix of these estimates for general causal and invertible PARMA models. The limiting distribution for the estimating equation and likelihood estimates are equivalent. Noting that

$$S_d(\underline{\beta}) = \sum_{n=0}^{d-1} \sum_{\nu=1}^{T} \frac{\epsilon^*_{nT+\nu}(\underline{\beta})}{\sigma^2(\nu)} \frac{\partial \epsilon^*_{nT+\nu}(\underline{\beta})}{\partial \underline{\alpha}} \frac{\partial \underline{\alpha}}{\partial \underline{\beta}} \tag{4.3.14}$$

and arguing as in Basawa and Lund (2001) will produce the following result.

Theorem 4.1. If $\{X_t\}$ is a causal and invertible PARMA series where $\{\epsilon_t\}$ is independent periodic white noise with $E[\epsilon_t] \equiv 0$ and $E[\epsilon^4_{nT+\nu}] \in (0, \infty)$ for each season ν, then

$$d^{\frac{1}{2}}(\hat{\underline{\beta}} - \underline{\beta}) \xrightarrow{D} N(0, F^{-1}(\underline{\beta})) \tag{4.3.15}$$

as $d \to \infty$ where the information matrix $F(\underline{\beta})$ is computed from

$$F(\underline{\beta}) = \sum_{\nu=1}^{T} \sigma^{-2}(\nu) D' \Gamma(\nu) D, \tag{4.3.16}$$

$D = \partial \underline{\alpha} / \partial \underline{\beta}$ is a $(q+p)T \times l$ dimensional matrix of partial derivatives translating constrained estimation to unconstrained estimation (the $(i,j)th$ element of D is $D_{i,j} = \partial \alpha_i / \partial \beta_j$ for $1 \le i \le (p+q)T$ and $1 \le j \le l$), and

$$\Gamma(\nu) = E\left[\left(\frac{\partial \epsilon_{nT+\nu}}{\partial \underline{\alpha}} \right) \left(\frac{\partial \epsilon_{nT+\nu}}{\partial \underline{\alpha}} \right)' \right]. \tag{4.3.17}$$

□

The next section gives several examples illustrating the practical uses of Theorem 4.1.

Likelihood ratio statistics for hypothesis tests involving the PARMA parameters are easily developed. Specifically suppose that a parsimonious PARMA model is specified via $\underline{\alpha} = g(\underline{\beta})$ where g is a known $(p+q)T$ dimensional function of the l dimensional parameter vector $\underline{\beta}$. Of course, one wants l much less than $(p+q)T$. To test the null hypothesis $H : \underline{\alpha} = g(\underline{\beta})$ nested within the unrestricted PARMA model (but subject to periodic stationarity and invertibility conditions), one can examine the likelihood ratio statistic

$$\Lambda = -2\ln\left(\frac{L(g(\hat{\underline{\beta}}_{ML}))}{L(\hat{\underline{\alpha}}_{ML})}\right), \tag{4.3.18}$$

where $\hat{\underline{\beta}}_{ML}$ and $\hat{\underline{\alpha}}_{ML}$ are the maximum likelihood estimates of $\underline{\beta}$ and $\underline{\alpha}$.

Under the conditions of Theorem 4.1, and regarding the white noise variances $\sigma^2(\nu), 1 \leq \nu \leq T$ as known nuisance parameters, one can establish the usual null hypothesis chi-squared limit law $\Lambda \xrightarrow{D} \chi^2_{(p+q)T-l}$ as $d \to \infty$, where χ^2_k denotes a chi-squared random variable with k degrees of freedom. If the white noise variances are unknown, consistent estimates of these quantities can be substituted without altering the limit distribution. Whereas we will focus on likelihoods and likelihood ratios in the ensuing computations, Wald Tests, Rao Tests, and AICC discrepancies are also easily developed.

Assuming H is true, one may wish to further test a hypothesis of the form $H^* : \underline{\beta} = \underline{\beta}_0$ against the alternative that $\underline{\beta} \neq \underline{\beta}_0$. The likelihood ratio statistic for this hypothesis test is

$$\lambda^* = -2\ln\left(\frac{L(g(\underline{\beta}_0))}{L(g(\hat{\underline{\beta}}_{ML}))}\right); \tag{4.3.19}$$

in this case, $\lambda^* \xrightarrow{D} \chi^2_l$ as $d \to \infty$. Applications of these methods are given in Section 5.

4.4 Computational Examples

Example 4.3. Consider the causal PAR(p) model in (4.3.9) with AR parameters as in (4.3.1). Our derivation will enter the PARMA parameters in the order

$$\underline{\alpha} = (\phi_1(1), \ldots, \phi_p(1); \phi_1(2), \ldots, \phi_p(2); \ldots; \phi_1(T), \ldots, \phi_p(T))', \tag{4.4.1}$$

and

$$\underline{\beta} = (A_{0,1}, B_{1,1}, A_{1,1}, \ldots, B_{r,1}, A_{r,1}; \ldots; A_{0,p}, B_{1,p}, A_{1,p}, \ldots, B_{r,p}, A_{r,p})'. \tag{4.4.2}$$

Here, $l = p(2r+1)$ and

$$D_{mT+\nu,j} = \frac{\partial \phi_{m+1}(\nu)}{\partial \beta_j}, \quad 0 \leq m \leq p-1, \quad 1 \leq \nu \leq T, \quad 1 \leq j \leq l. \tag{4.4.3}$$

Taking derivatives in (4.3.1) and performing some tedious simplifications gives, for a fixed i, j satisfying $1 \leq i \leq pT$ and $1 \leq j \leq l$,

$$D_{i,j} = \begin{cases} \cos\left(\frac{2\pi k\nu}{T}\right), \text{ if } i = mT+\nu \text{ and } j = m(2r+1)+2k+1 \\ \sin\left(\frac{2\pi k\nu}{T}\right), \text{ if } i = mT+\nu \text{ and } j = m(2r+1)+2k \\ 0, \text{ otherwise} \end{cases} \tag{4.4.4}$$

for some values satisfying $0 \leq m \leq p-1$, $1 \leq \nu \leq T$, and $0 \leq k \leq r$. □

Example 4.4. Consider a causal PAR(1) series as a specific case of Example 4.3. The PAR(1) causality condition is well known as $|\prod_{\nu=1}^{T} \phi_1(\nu)| < 1$ (cf. Vecchia 1985a and Hurd *et al.* 1999). Here, $\underline{\alpha} = (\phi_1(1), \ldots, \phi_1(T))'$. By (4.3.5),

$$\frac{\partial \epsilon_{nT+\nu}}{\partial \underline{\alpha}} = -X_{nT+\nu-1} \frac{\partial \phi_1(\nu)}{\partial \underline{\alpha}} = -X_{nT+\nu-1} e_{\nu}^{T} \tag{4.4.5}$$

where e_i^T denotes a $T \times 1$ unit vector whose entries are all zero except for a one in the ith row. Combining (4.3.17) and (4.4.5) gives

$$\Gamma(\nu) = \text{Var}(X_{nT+\nu-1}) E_{\nu,\nu}^{T}, \tag{4.4.6}$$

where $E^T_{i,j} = \underline{e}^T_i \underline{e}^T_j{}'$ is a $T \times T$ matrix whose entries are all zero except for a unit entry in the jth column of row i.

When the PAR(1) parameters are constrained as in Example 4.1, the constrained parameters will be tracked in the order $\underline{\beta} = (A_{0,1}, B_{1,1}, A_{1,1}, \ldots, B_{r,1}, A_{r,1})'$. Differentiating (4.3.1) and simplifying shows that the odd columns of D have the form

$$D_{\nu,2k+1} = \cos\left(\frac{2\pi k\nu}{T}\right), \quad 1 \le \nu \le T, \quad 0 \le k \le r, \tag{4.4.7}$$

and that even columns of D have entries

$$D_{\nu,2k} = \sin\left(\frac{2\pi n\nu}{T}\right), \quad 1 \le \nu \le T, \quad 1 \le k \le r. \tag{4.4.8}$$

Using (4.4.6) in (4.3.16) gives

$$F(\underline{\beta}) = \sum_{\nu=1}^{T} \frac{\mathrm{Var}(X_{nT+\nu-1}) D' E^T_{\nu,\nu} D}{\sigma^2(\nu)}. \tag{4.4.9}$$

All quantities in (4.4.9) are now explicitly identified except for the periodic variance of $\{X_t\}$, which Bloomfield *et al.* (1994) compute as

$$\gamma_0(\nu) = \mathrm{Var}(X_{nT+\nu}) = r_\nu^2 \left[\frac{r_T^2}{1 - r_T^2} \sum_{k=1}^{T} \frac{\sigma^2(k)}{r_k^2} + \sum_{k=1}^{\nu} \frac{\sigma^2(k)}{r_k^2} \right], \tag{4.4.10}$$

where $r_\nu = \prod_{j=1}^{\nu} \phi_1(j)$ for $1 \le \nu \le T$.

For the PAR(1) parameterization in Example 4.2, $\underline{\alpha} = (\phi_1(1), \ldots, \phi_1(T))'$, $\underline{\beta} = (\beta_1, \beta_2)'$, $T = 7$, and the causality condition is $|\beta_1^5 \beta_2^2| < 1$. For white noise variances, we take $\sigma^2(\nu) = \sigma_1^2$ if $1 \le \nu \le 5$ and $\sigma^2(\nu) = \sigma_2^2$ if $6 \le \nu \le 7$. Here, D is a 7×2 matrix whose entries are

$$D_{i,j} = \begin{cases} 1, & \text{if } 1 \le i \le 5 \text{ and } j = 1 \\ 1, & \text{if } 6 \le i \le 7 \text{ and } j = 2 \\ 0, & \text{otherwise} \end{cases} . \tag{4.4.11}$$

Hence,

$$D' E^T_{\nu,\nu} D = \begin{cases} E^2_{1,1}, & \text{if } 1 \le \nu \le 5 \\ E^2_{2,2}, & \text{if } 6 \le \nu \le 7 \end{cases} \tag{4.4.12}$$

and substitution into (4.4.9) gives

$$F(\beta_1, \beta_2) = \text{Diag}\left(\sigma_1^{-2} \sum_{\nu=1}^{5} \text{Var}(X_{nT+\nu-1}), \sigma_2^{-2} \sum_{\nu=6}^{7} \text{Var}(X_{nT+\nu-1})\right), \tag{4.4.13}$$

where $\text{Var}(X_{nT+\nu-1})$ is computed from (4.4.10) as an explicit function of β_1, β_2, σ_1^2, and σ_2^2. Some tedious manipulations with (4.4.10) produce

$$\text{Var}(X_{nT+\nu}) = \beta_1^{2\nu} \left[\left(\frac{\beta_2^4}{1 - \beta_1^{10}\beta_2^4} \right) \left(\frac{\sigma_1^2(\beta_1^{10} - 1)}{\beta_1^2 - 1} + \frac{\sigma_2^2(\beta_2^2 + 1)}{\beta_2^4} \right) + \sigma_1^2 \left(\frac{\beta_1^{2\nu} - 1}{\beta_1^{2\nu}(\beta_1^2 - 1)} \right) \right]$$

for $1 \le \nu \le 5$ and

$$\text{Var}(X_{nT+\nu}) = \beta_2^{2(\nu-5)} \left[\text{Var}(X_{nT+5}) + \sigma_2^2 \left(\frac{(\beta_2^2)^{\nu-5} - 1}{(\beta_2^2)^{\nu-5}(\beta_2^2 - 1)} \right) \right]$$

for $6 \le \nu \le 7$. The diagonal structure of $F(\underline{\beta})$ implies that the estimates of β_1 and β_2 are asymptotically independent — a structure inherited from general unconstrained periodic autoregressions (cf. Pagano 1978). □

4.5 Model Development for Griffin Temperatures

A PARMA model will now be developed for the Griffin data in Figure 4.1. To focus initially on autocovariances, a daily mean $\mu_\nu = E[X_{nT+\nu}]$ was estimated via

$$\hat{\mu}_\nu = d^{-1} \sum_{n=0}^{d-1} X_{nT+\nu} \tag{4.5.1}$$

and subtracted from the series. Next, an unconstrained PAR(1) model was fitted to this mean adjusted data via the method of moments:

$$\hat{\phi}_1(\nu) = \hat{\rho}_1(\nu) \quad \text{and} \quad \hat{\sigma}^2(\nu) = d^{-1} \sum_{n=0}^{d-1} (X_{nT+\nu} - \hat{X}_{nT+\nu})^2. \tag{4.5.2}$$

Figure 4.3 plots the PAR(1) moment estimates of autoregressive and white noise variance parameters. The estimated white noise variances have a distinctly sinusoidal shape; however, it is not clear how the autoregressive parameters vary in season if at all. After some exploratory Fourier fits, the parameterizations

$$\phi_1(\nu) = R_0 + R_1 \cos\left(\frac{2\pi(\nu - \tau_1)}{T}\right) + R_2 \cos\left(\frac{4\pi(\nu - \tau_2)}{T}\right) \tag{4.5.3}$$

and

$$\sigma^2(\nu) = S_0 + S_1 \cos\left(\frac{2\pi(\nu - \kappa_1)}{T}\right) \tag{4.5.4}$$

were chosen for further study. Seasonal variation of $\phi_1(\nu)$ in ν is controlled through R_1 and R_2 (we are interested in testing if R_1 and/or R_2 is zero); seasonal variation of $\sigma^2(\nu)$ in ν is controlled through S_1. The PAR(1) likelihood was evaluated via (4.3.11) and optimized while imposing the constraints in (4.5.3) and (4.5.4). The likelihood estimates, to three decimal places, are $\hat{R}_0 = 0.705 \pm 0.00423$, $\hat{R}_1 = 0.0263 \pm 0.00541$, $\hat{R}_2 = 0.00574 \pm 0.00541$, $\hat{\tau}_1 = 3.169 \pm 14.265$, $\hat{\tau}_2 = 3.427 \pm 32.625$, $\hat{S}_0 = 8.618$, $\hat{S}_1 = 7.050$, and $\hat{\kappa}_1 = 29.533$. The error margins were computed from the information matrix derived in Theorem 4.1 and Example 4.4 (standard errors for nuisance parameters are omitted). A likelihood of $-2\ln(L) = 116128.814$ was achieved.

The estimate and standard error of R_2 produce a standard normal z-statistic of 1.061 for the null hypothesis test that $R_2 = 0$; hence, the PAR(1) model in (4.5.3) and (4.5.4) was refitted with R_2 constrained as zero. The likelihood estimates for this 'reduced' model are $\hat{R}_0 = 0.708 \pm 0.00405$, $\hat{R}_1 = 0.0281 \pm 0.00490$, $\hat{\tau}_1 = 3.169 \pm 13.288$, $\hat{S}_0 = 8.620$, $\hat{S}_1 = 7.054$, and $\hat{\kappa}_1 = 29.520$; a likelihood of $-2\ln(L) = 116129.598$ was achieved.

The likelihood ratio test in (4.3.19) also favors $R_2 = 0$. Here, a test statistic of 0.784 is obtained, which, under the null hypothesis, has one degree of freedom (the restriction $R_2 = 0$ imposes one parametric constraint) and a P-value of 0.377.

Figure 4.3 compares the constrained likelihood estimates of $\phi_1(\nu), 1 \le \nu \le T$ and $\sigma^2(\nu), 1 \le \nu \le T$ with their moment counterparts in (4.5.2) — notice that the constrained parameters retain the general structure of the moment estimates with fewer expended parameters.

A set of residuals was computed for this model fit. In general, the residual at time t, denoted by L_t, is defined as

$$L_{nT+\nu} = \frac{X_{nT+\nu} - \hat{X}_{nT+\nu}}{\sqrt{v_{nT+\nu}}} \tag{4.5.5}$$

and has the usual interpretation of observation minus prediction. If the fitted model is adequate, the residuals should be approximately uncorrelated; otherwise, the model can be improved. The denominator in (4.5.5) scales $\{L_t\}$ to a unit variance series.

Figure 4.4 plots estimated versions of $\{L_t\}$ along with their sample autocorrelations and partial autocorrelations. Ninety five percent confidence bounds for white noise are included for comparison's sake. The autocorrelations and partial autocorrelations exceed the white noise confidence bounds over the first four lags and indicate a departure from white noise. The empirical tests of Bloomfield *et al.* (1994) were applied to $\{L_t\}$ and did not reveal any periodicities. Hence, the residuals appear to be short-memory and stationary, but decisively non-white. An ARMA model was next fitted to these residuals. The optimal fitting ARMA model was judged to be an AR(3) model via the usual ARMA fitting criterion (cf. Chapter 9 of Brockwell and Davis 1991).

Mathematically combining the PAR(1) difference equation in (4.3.5) with errors that are a solution to the AR(3) difference equation

$$\epsilon_t - \eta_1\epsilon_{t-1} - \eta_2\epsilon_{t-2} - \eta_3\epsilon_{t-3} = W_t, \tag{4.5.6}$$

where $\{W_t\}$ is mean zero white noise with $\mathrm{Var}(W_t) \equiv 1$, results in a PAR(4) difference equation with the coefficients

$$\begin{aligned}
\phi_{PAR(4),1}(\nu) &= \phi_{PAR(1),1}(\nu) + \eta_1 \frac{\sigma(\nu)}{\sigma(\nu-1)}, \\
\phi_{PAR(4),2}(\nu) &= \eta_2 \frac{\sigma(\nu)}{\sigma(\nu-2)} - \eta_1 \phi_{PAR(1),1}(\nu-1)\frac{\sigma(\nu)}{\sigma(\nu-1)}, \\
\phi_{PAR(4),3}(\nu) &= \eta_3 \frac{\sigma(\nu)}{\sigma(\nu-3)} - \eta_2 \phi_{PAR(1),1}(\nu-2)\frac{\sigma(\nu)}{\sigma(\nu-2)}, \\
\phi_{PAR(4),4}(\nu) &= -\eta_3 \phi_{PAR(1),1}(\nu-3)\frac{\sigma(\nu)}{\sigma(\nu-3)} \qquad (4.5.7)
\end{aligned}$$

(PAR(1) and PAR(4) coefficients are distinguished with the obvious notation). We now have a good PARMA model candidate and proceed to 'tune' the model parameters via likelihood.

Much of the motivation for the analysis of the Griffin series lies with forecasting. Because of this, we also take care to model the seasonal mean cycle of Griffin temperatures. Using the Fourier parameterization

$$\mu_\nu = E[X_{nT+\nu}] = M_0 + \sum_{k=1}^{u} M_k \cos\left(\frac{2\pi k(\nu - \xi_k)}{T}\right) \qquad (4.5.8)$$

to describe the periodic mean, we will estimate the value of u, $M_0, M_1, \ldots, M_u$ and $\xi_1, \ldots, \xi_u$ jointly with the above PAR(4) model parameters. Imposing (4.5.7), (4.5.3) with $R_2 = 0$, and (4.5.4), the following optimal model likelihoods were obtained for varying u:

u	$-2\ln(L)$
1	116074.063
2	116044.695
3	116035.086
4	116029.320
5	116029.258.

Comparing the nested likelihoods above with a chi-squared test with one degree of freedom suggests that the optimal value of u is four. The periodic mean estimate with $u = 2$ is substantially better than that with $u = 1$, reducing $-2\ln(L)$ by about 30 points, but the three- and four-term fits also improve daily mean description. The above parameterization saves $T - (2u + 1) = 356$ parameters over general unconstrained seasonal means. The top graph in Figure 4.2 compares the Fourier estimates with $u = 4$ against the general periodic means in (4.5.1) and reveals a close agreement. The Fourier fit to the mean also graphically confirms a climatological principle common in the temperate zone: the descent from summer into winter is quicker than the ascent from winter into summer.

The likelihood estimates for the PAR(4) autoregressive and white noise variance parameters with $u = 4$ are $\hat{R}_0 = 0.708 \pm 0.0106$, $\hat{R}_1 = 0.0331 \pm 0.00629$, $\hat{\tau}_1 = 0.982 \pm 11.180$, $\hat{S}_0 = 8.176$, $\hat{S}_1 = 6.469$, $\hat{\kappa}_1 = 25.454$, $\hat{\eta}_1 = 0.103 \pm 0.0122$, $\hat{\eta}_2 = -0.092 \pm 0.00858$, and $\hat{\eta}_3 = -0.0430 \pm 0.00842$. The error margins are again based on Theorem 3.1. The fitted model saves $5T - 9 = 1816$ parameters over a general mean zero PAR(4) model for daily data with general white noise variances. Estimates of the mean parameters in (4.5.8) are $\hat{M}_0 = 13.110$, $\hat{M}_1 = -10.062$, $\hat{\xi}_1 = 22.596$, $\hat{M}_2 = 0.194$, $\hat{\xi}_2 = 74.990$, $\hat{M}_3 = 0.0893$, $\hat{\xi}_3 = -20.014$, $\hat{M}_4 = 0.0834$, and $\hat{\xi}_4 = -20.014$; a likelihood of $-2\ln(L) = 116029.320$ was achieved with this model fit.

As a final diagnostic check, a set of residuals for the above PAR(4) model was computed via (4.5.5) after subtracting daily means estimated by (4.5.8). The sample autocorrelations and partial autocorrelations of these residuals, along with 95% confidence bounds for white noise, are displayed in Figure 4.5. As all autocorrelations lie inside or very close to the 95% white noise bounds, the PAR(4) residuals are concluded to be white noise. Hence, the Griffin series is parsimoniously described with a PAR(4) model as constrained in (4.5.7), (4.5.3) with $R_2 = 0$, (4.5.4), and (4.5.8) with $u = 4$.

The above PAR(4) model expends 18 total parameters, 9 of which describe the seasonal mean structure of the series. This is substantially smaller than the $6 \times T = 2190$ parameters in a general mean PAR(4) model with general white noise variances.

A layering strategy for modeling general short memory periodic series now emerges. First, find a simple PARMA model that leaves short memory and stationary (not periodic) residuals. In the case of the Griffin series, a PAR(1) model fit served this purpose. If the PAR(1) fit had non-periodic or long memory residuals, then a PAR(2), PARMA(1,1), etc. model could alternatively be investigated as a first layer. Next, consolidate the parameters of the first layer with the standard errors derived in Theorem 4.1 or a likelihood ratio test. Reduce the parameters in the first layer as much as possible. For the second layer, fit a stationary ARMA model to the residuals of the first layer. Then mathematically combine the difference equations governing the two layers. The result will be a PARMA model which is parsimonious. As a final step, retune the PARMA parameters jointly in both layers by maximizing a combined two-layer model likelihood or estimating equation. For the Griffin series, more complex first layers (e.g. PAR(2), PARMA(1,1)) were explored, but all non-PAR(1) first layers ultimately led to worse fits with increased parameter counts.

The above methods have worked very well in describing the Griffin series. In general, the results and techniques in this article now make it possible to quickly fit parsimonious PARMA models to periodic short-memory series.

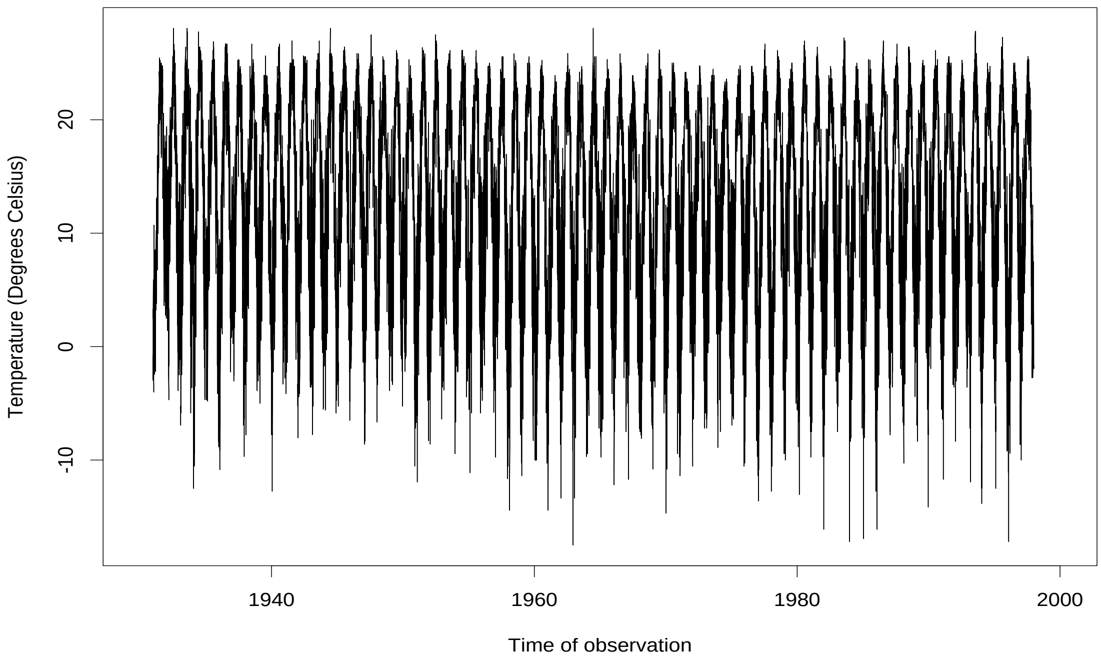

Figure 4.1: Griffin, GA Temperatures.

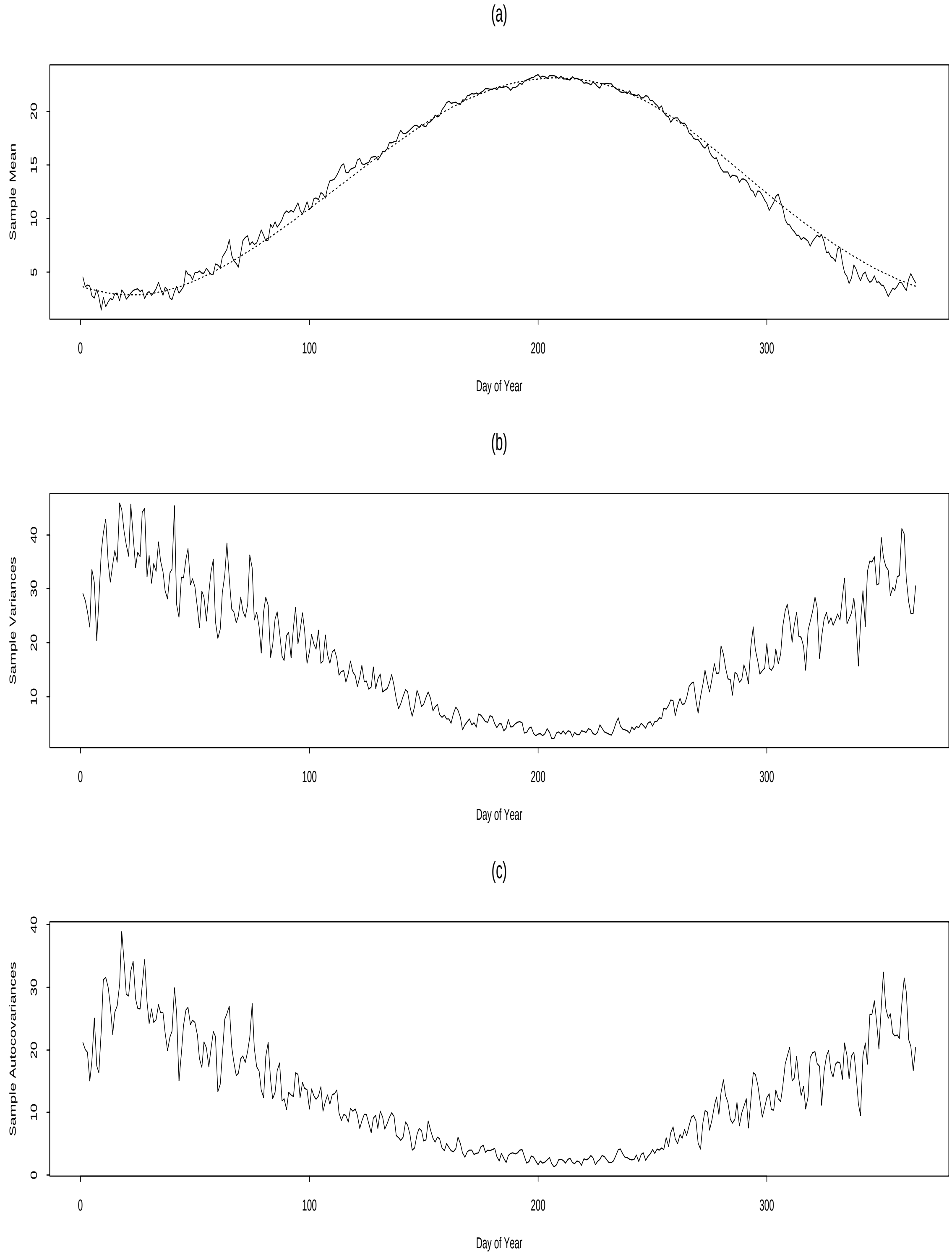

Figure 4.2: (a) Griffin Daily Sample Means; (b) Griffin Daily Sample Variances; (c) Griffin Lag One Autocovariances.

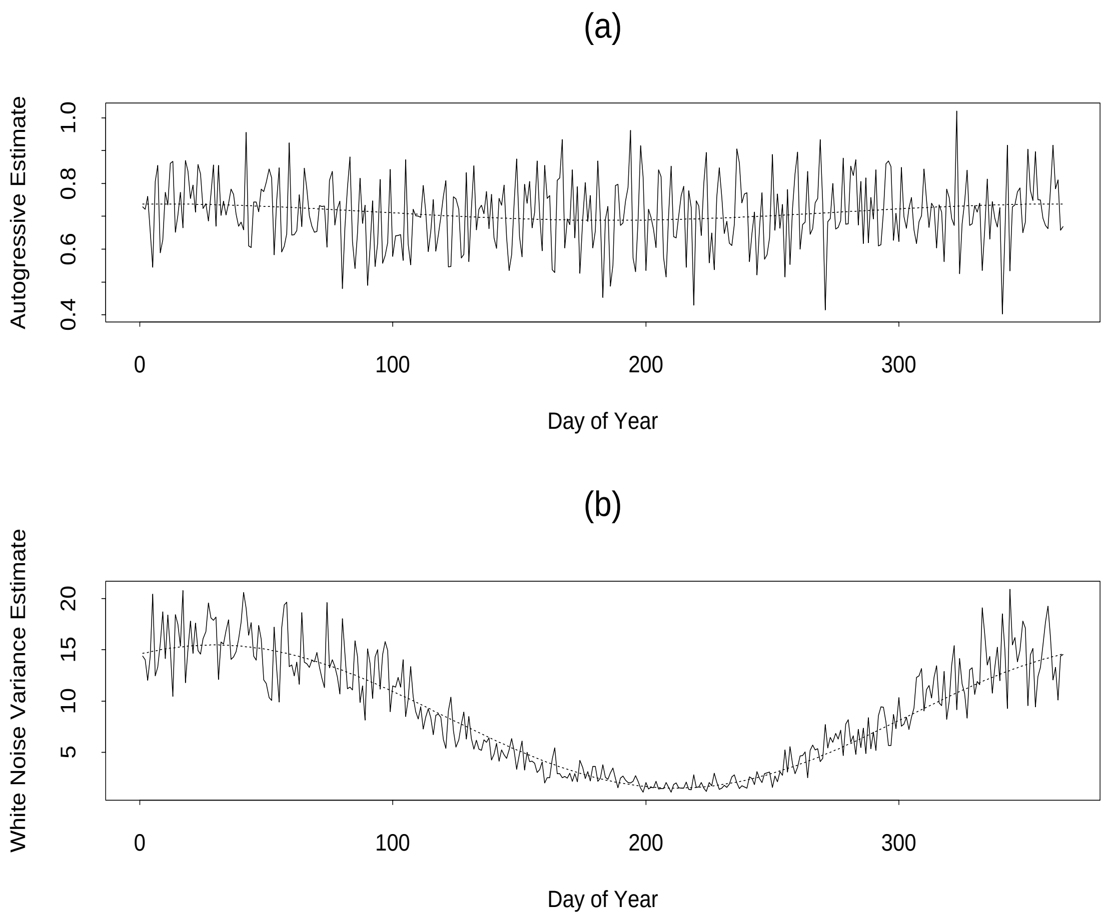

Figure 4.3: (a) PAR(1) Autoregressive Estimates; (b) PAR(1) White Noise Variance Estimates.

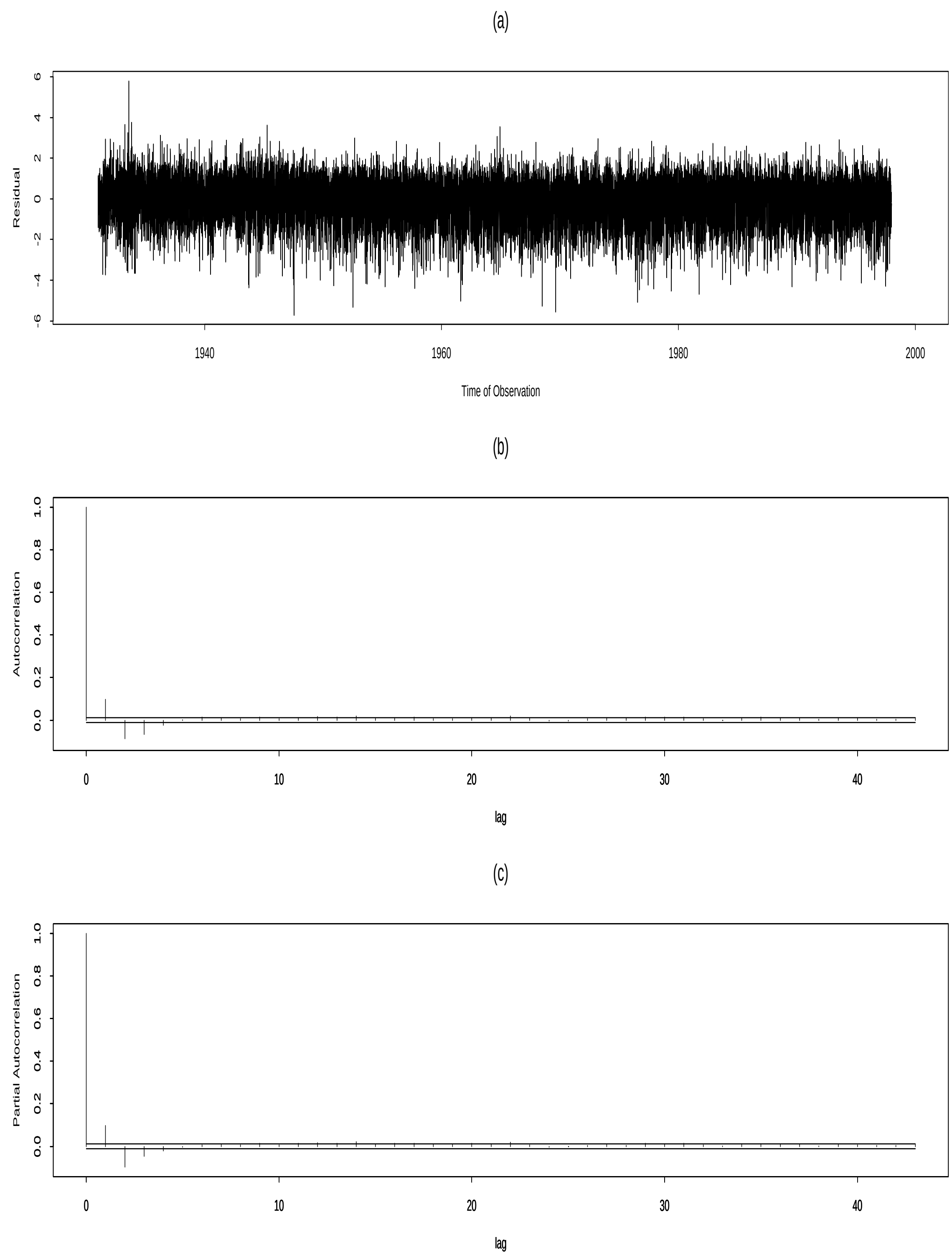

Figure 4.4: (a) PAR(1) Residuals; (b) Autocorrelation; (c) Partial Autocorrelation.

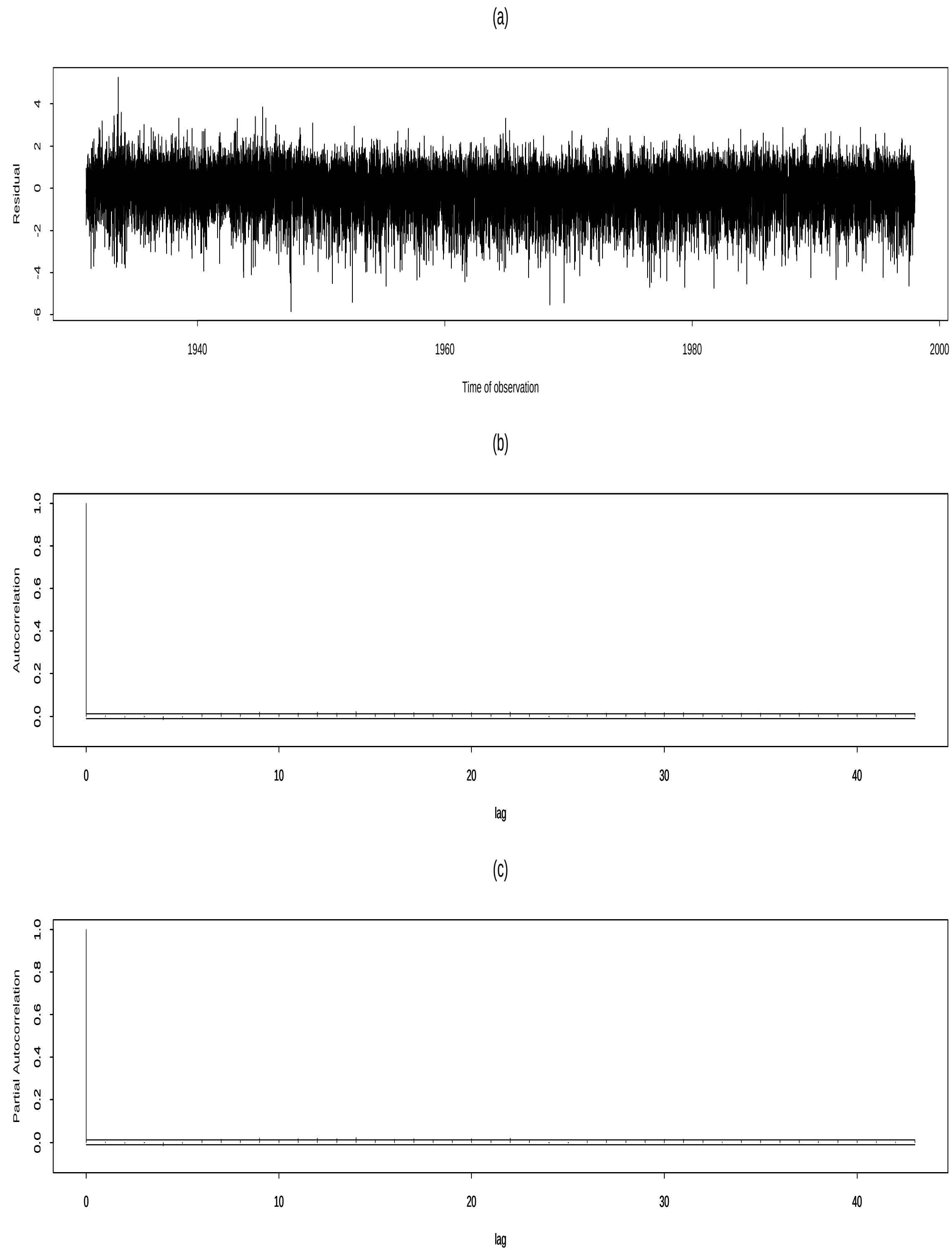

Figure 4.5: (a) PAR(4) Residuals; (b) Autocorrelation; (c) Partial Autocorrelation.

Chapter 5

Seasonal Periodic Autoregressive Moving-Average Models: Future Research

5.1 Motivation

In this section, we introduce a new class of time series models, called seasonal periodic autoregressive moving-average models (SPARMA), by generalizing Box-Jenkins seasonal ARMA (SARMA) models.

The Box-Jenkins seasonal ARMA model (SARMA$((p_1 \times p_2),(q_1 \times q_2))$) is defined by the difference equation:

$$\begin{aligned} & X_{nT+\nu} - \sum_{i=1}^{p_1} \phi_i X_{nT+\nu-i} - \sum_{j=1}^{p_2} \zeta_j X_{(n-j)T+\nu} + \sum_{i=1}^{p_1}\sum_{j=1}^{p_2} \phi_i \zeta_j X_{(n-j)T+\nu-i} \\ = \quad & \epsilon_{nT+\nu} + \sum_{i=1}^{q_1} \theta_i \epsilon_{nT+\nu-i} + \sum_{j=1}^{q_2} \xi_j \epsilon_{(n-j)T+\nu} + \sum_{i=1}^{q_1}\sum_{j=1}^{q_2} \theta_i \xi_j \epsilon_{(n-j)T+\nu-i}, \end{aligned} \tag{5.1.1}$$

where $\{\epsilon_t\} \sim \mathrm{WN}(0, \sigma^2)$. This model incorporates correlations within each period and also correlations between periods. For instance, for monthly data, one may be interested in the following questions:

(a) How is each successive monthly observations related to the previous monthly observations?

(b) How is each month of a given year related to the same month in previous years?

The difference equations in (5.1.1) allow one to model two types of correlations in a convenient form. The product terms involving $\phi\zeta$ and $\theta\xi$ represent interaction effects.

In the above model, all the autoregressive parameters, moving-average parameters, and the white noise variances are constant over time.

By allowing the parameters in (5.1.1) to periodically vary with time, we obtain the corresponding model which we will refer to as a SPARMA model.

A time series $\{X_t\}$ with finite second moments is said to be a seasonal periodic autoregressive moving-average (SPARMA) series of order $((p_1 \times p_2), (q_1 \times q_2))$, or SPARMA $((p_1 \times p_2), (q_1 \times q_2))$, with period $T \geq 1$, if it satisfies the difference equation

$$\begin{aligned} & X_{nT+\nu} - \sum_{i=1}^{p_1} \phi_i(\nu) X_{nT+\nu-i} - \sum_{j=1}^{p_2} \zeta_j(\nu) X_{(n-j)T+\nu} \\ & + \sum_{i=1}^{p_1} \sum_{j=1}^{p_2} \phi_i(\nu) \zeta_j(\nu) X_{(n-j)T+\nu-i} \\ = \; & \epsilon_{nT+\nu} + \sum_{i=1}^{q_1} \theta_i(\nu) \epsilon_{nT+\nu-i} + \sum_{j=1}^{q_2} \xi_j(\nu) \epsilon_{(n-j)T+\nu} \\ & + \sum_{i=1}^{q_1} \sum_{j=1}^{q_2} \theta_i(\nu) \xi_j(\nu) \epsilon_{(n-j)T+\nu-i}, \end{aligned} \tag{5.1.2}$$

where $\{\epsilon_{nT+\nu}\} \sim \mathrm{PWN}(0, \sigma^2(\nu))$.

Equation (5.1.2) becomes a seasonal ARMA (SARMA) model (Box *et al.* (1994)) when the model parameters $\phi_i(\nu), \zeta_j(\nu), \theta_i(\nu), \xi_j(\nu), \sigma^2(\nu)$ do not depend on ν. Model (5.1.2) reduces to a PARMA(p_1, q_1) model if all $\{\zeta_j(\nu)\}$ and $\{\xi_j(\nu)\}$ are set equal to zero. Thus, a SPARMA model includes both SARMA and PARMA as special cases.

We will denote SPARMA$((p_1 \times p_2), (0 \times 0))$ and SPARMA$((0 \times 0), (p_1 \times p_2))$ models respectively as SPAR$(p_1 \times p_2)$ and SPMA$(q_1 \times q_2)$.

Example 5.1. The SPAR(1×1) model is defined by

$$\begin{aligned} & X_{nT+\nu} - \phi(\nu) X_{nT+\nu-1} - \zeta(\nu) X_{(n-1)T+\nu} + \phi(\nu)\zeta(\nu) X_{(n-1)T+\nu-1} \\ = \; & \epsilon_{nT+\nu}. \end{aligned} \tag{5.1.3}$$

From (5.1.3), we obtain

$$\begin{aligned} \text{SAR}(1 \times 1): \qquad & X_{nT+\nu} - \phi X_{nT+\nu-1} - \zeta X_{(n-1)T+\nu} + \phi\zeta X_{(n-1)T+\nu-1} \qquad (5.1.4) \\ = \quad & \epsilon_{nT+\nu} \end{aligned}$$

and

$$\text{PAR}(1): \qquad X_{nT+\nu} - \phi(\nu) X_{nT+\nu-1} = \epsilon_{nT+\nu} \qquad (5.1.5)$$

by setting $\phi(\nu) = \phi$ and $\zeta(\nu) = \zeta$ in (5.1.3), and by setting $\zeta(\nu) = 0$ in (5.1.3), respectively.

Example 5.2. Consider the models PAR(1), SAR(1×1), and SPAR(1×1) with period $T = 4$ given in Table 5.1. All model parameters and white noise variances periodically vary for both PAR(1) and SPAR(1×1), but remain constant for SAR(1×1).

Both the PAR(1) and SPAR(1×1) models have different autocorrelation and partial autocorrelation structures in different seasons. The autocorrelations in the PAR(1) model die out more rapidly than those of SPAR(1×1) model (see Figure 5.1 and Figure 5.3). For example, in season 1, the absolute values of autocorrelations for PAR(1) is reduced from 0.3149 at *lag* 1 to 0.0003 at *lag* 10, while the absolute values of autocorrelation for SPAR(1×1) is reduced from 0.3188 at *lag* 1 to 0.0462 at *lag* 10. The difference between autocorrelations at *lag* 1 and *lag* 10 for PAR(1) is 0.3146, while it is only 0.2726 for SPAR(1×1). The SAR(1×1) model is stationary, and hence, it has the same autocorrelation structure in different seasons (see Figure 5.2). According to Theorem 3.2 and Theorem 3.3, their PACFs are zero after lags 1, 5, and 5. The PACFs are given in Figures 5.4-5.6. □

Model (5.1.2) can be written in the PARMA$(p_1 + p_2 T, q_1 + q_2 T)$ form

$$X_{nT+\nu} - \sum_{k=1}^{p_1+p_2T} \phi_k^*(\nu) X_{nT+\nu-k} = \epsilon_{nT+\nu} + \sum_{k=1}^{q_1+q_2T} \theta_k^*(\nu) \epsilon_{nT+\nu-k} \qquad (5.1.6)$$

with

$$\phi_k^*(\nu) = \begin{cases} \phi_i(\nu) & \text{for} \quad 0 \le k \le p_1 \\ \zeta_j(\nu) & \text{for} \quad k = jT \ (0 \le j \le p_2) \\ \phi_i(\nu)\zeta_j(\nu) & \text{for} \quad k = i + jT \ (1 \le i \le p_1 \text{ and } 1 \le j \le p_2) \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\theta_k^*(\nu) = \begin{cases} \theta_i(\nu) & \text{for} \quad 0 \le k \le q_1 \\ \xi_j(\nu) & \text{for} \quad k = jT \ (0 \le j \le q_2) \\ \theta_i(\nu)\xi_j(\nu) & \text{for} \quad k = i + jT \ (1 \le i \le q_1 \text{ and } 1 \le j \le q_2) \\ 0 & \text{otherwise.} \end{cases}$$

Example 5.3. The SPAR(1×1) model can be reparameterized as a PAR$(1 + T)$ model with

$$\phi_k^*(\nu) = \begin{cases} \phi(\nu) & \text{for} \quad k = 1 \\ \zeta(\nu) & \text{for} \quad k = T \\ \phi(\nu)\zeta(\nu) & \text{for} \quad k = 1 + T \\ 0 & \text{otherwise,} \end{cases}$$

□

It follows that some fundamental features of SPARMA models are easily obtained from PARMA properties.

5.2 Stationarity and Computation of Autocovariances

Based on the foregoing results on PARMA series, we will briefly discuss some properties, such as stationarity and autocovariance computation, via an example.

Example 5.4. Consider the SPARMA(1×1) model with period T governed by the difference equation

$$\begin{aligned} & X_{nT+\nu} - \phi(\nu)X_{nT+\nu-1} - \zeta(\nu)X_{(n-1)T+\nu} + \phi(\nu)\zeta(\nu)X_{(n-1)T+\nu-1} \\ = \; & \epsilon_{nT+\nu}. \end{aligned} \tag{5.2.1}$$

The T-variate ARMA version for (5.2.1) is

$$\Phi_0 \underline{X}_n - \Phi_1 \underline{X}_{n-1} - \Phi_2 \underline{X}_{n-2} = \underline{\epsilon}_n, \tag{5.2.2}$$

where $\{\Phi_0, \Phi_1, \Phi_2, \Theta_0\}$ are respectively determined by (2.2.5) and (2.2.6). Therefore, the causality condition for this model is

$$\det(\Phi_0 - \Phi_1 z - \Phi_2 z^2) = \prod_{\nu=1}^{T} (1 - \zeta(\nu)z) \left(1 - \prod_{\nu=1}^{T} \phi(\nu) z^T\right) \neq 0$$

for all complex z satisfying $|z| \leq 1$. Hence a unique causal solution to (5.2.1) exists when

$$\prod_{1 \leq \nu \leq T} |\phi(\nu)| < 1, \text{ and } |\zeta(\nu)| < 1 \text{ for all } 1 \leq \nu \leq T. \tag{5.2.3}$$

Under (5.2.3), the causal solution to (5.2.1) has the PMA(∞) representation

$$X_{nT+\nu} = \sum_{k=0}^{\infty} \psi_k(\nu)\epsilon_{nT+\nu-k}, \tag{5.2.4}$$

where $\{\psi_k(\nu)\}$ satisfies $\max_{1 \leq \nu \leq T} \sum_{k=0}^{\infty} |\psi_k(\nu)| < \infty$. Furthermore, we can determine $\{\psi_k(\nu)\}$ via (2.2.10), i.e.,

$$\begin{aligned} \psi_k(\nu) = \; & \phi(\nu)\psi_{k-1}(\nu-1)I_{[k \geq 1]} + \zeta(\nu)\psi_{k-T}(\nu)I_{[k \geq T]} \\ & + \phi(\nu)\zeta(\nu)\psi_{k-(T+1)}(\nu-1)I_{[k \geq T+1]}, \end{aligned}$$

where $k \geq 1$, and $1 \leq \nu \leq T$. Similar to PARMA models, autocovariances are found by multiplying each side of (5.2.1) by $X_{nT+\nu-h}$, and then taking expectations:

$$\begin{aligned} & \gamma_h(\nu) - \phi(\nu)\gamma_{h-1}(\nu-1) - \zeta(\nu)\gamma_h(\nu-T) + \phi(\nu)\zeta(\nu)\gamma_{h-1}(\nu-(T+1)) \\ = \; & \sigma^2(\nu)1_{[h=0]}. \end{aligned} \tag{5.2.5}$$

In particular, the recursion

$$\gamma_h(\nu) = \phi(\nu)\gamma_{h-1}(\nu-1) + \zeta(\nu)\gamma_h(\nu-T) - \phi(\nu)\zeta(\nu)\gamma_{h-1}(\nu-(T+1)), \quad (5.2.6)$$

for $h > T+1$ emerges. □

5.3 SPARMA Parameter Estimation

We will illustrate the problem via two examples.

Example 5.5. Consider a causal and invertible SPAR((1×1)) model,

$$\begin{aligned} & X_{nT+\nu} - \phi(\nu)X_{nT+\nu-1} - \zeta(\nu)X_{(n-1)T+\nu} + \phi(\nu)\zeta(\nu)X_{(n-1)T+\nu-1} \\ = \; & \epsilon_{nT+\nu}. \end{aligned} \quad (5.3.1)$$

Let

$$\underline{\alpha} = (\phi(1), \zeta(1); \phi(2), \zeta(2); \cdots; \phi(T), \zeta(T))'.$$

By Theorem 2.3 in Chapter 2, we have $d^{\frac{1}{2}}(\underline{\hat{\alpha}}_{MLE} - \underline{\alpha}) \xrightarrow{D} \mathcal{N}(0, F^{-1}(\underline{\alpha}))$ as $d \to \infty$, where $F(\alpha)$ is determined as follows. Taking partial derivatives in (5.3.1) gives

$$\frac{\partial \epsilon_{nT+\nu}(\underline{\alpha})}{\partial \underline{\alpha}} = -\underline{1}_{2\nu-1}X_{nT+\nu-1} - \underline{1}_{2\nu}X_{nT+\nu-T} + \underline{e}_{2\nu-1}X_{nT+\nu-(T+1)}, \quad (5.3.2)$$

where $\underline{1}_i$ is a $2T \times 1$ vector whose entries are all zeros except for a one in the ith entry; $\underline{e}_i$ is a $2T \times 1$ vector whose entries are all zeros except for $\zeta(i)$ on the ith entry and $\phi(i)$ on the $(i+1)$th entry. By (4.3.17)

$$\begin{aligned} & F_\nu(\underline{\alpha}, \underline{\sigma}^2) \\ = \; & E[(\frac{\partial \epsilon_{nT+\nu}(\underline{\alpha})}{\partial \underline{\alpha}})(\frac{\partial \epsilon_{nT+\nu}(\underline{\alpha})}{\partial \underline{\alpha}})'] \\ = \; & E[(-\underline{1}_{2\nu-1}X_{nT+\nu-1} - \underline{1}_{2\nu}X_{nT+\nu-T} + \underline{e}_{2\nu-1}X_{nT+\nu-(T+1)}) \\ & (-\underline{1}_{2\nu-1}X_{nT+\nu-1} - \underline{1}_{2\nu}X_{nT+\nu-T} + \underline{e}_{2\nu-1}X_{nT+\nu-(T+1)})'] \\ = \; & \underline{1}_{2\nu-1}\underline{1}'_{2\nu-1}\gamma_0(\nu-1) + \underline{1}_{2\nu}\underline{1}'_{2\nu-1}\gamma_{T-1}(\nu-1) - \underline{e}_{2\nu-1}\underline{1}'_{2\nu-1}\gamma_T(\nu-1) \\ & + \underline{1}_{2\nu-1}\underline{1}'_{2\nu}\gamma_{T-1}(\nu-1) + \underline{1}_{2\nu}\underline{1}'_{2\nu}\gamma_0(\nu) - \underline{e}_{2\nu-1}\underline{1}'_{2\nu}\gamma_1(\nu) \\ & - \underline{1}_{2\nu-1}\underline{e}'_{2\nu-1}\gamma_T(\nu-1) - \underline{1}_{2\nu}\underline{e}'_{2\nu-1}\gamma_1(\nu) + \underline{e}_{2\nu-1}\underline{e}'_{2\nu-1}\gamma_0(\nu-1), \end{aligned}$$

which is a $(2T \times 2T)$ matrix whose entries are all zeros except for the (i, i), $(i+1, i)$, $(i, i+1)$, and $(i+1, i+1)$ elements. Specifically,

$$F_\nu(\underline{\alpha}, \underline{\sigma}^2) = \begin{pmatrix} 0 & \cdots & 0 & 0 & 0 & 0 \cdots & 0 & \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 \cdots & 0 & \\ 0 & \cdots & 0 & f_{2\nu-1,2\nu-1} & f_{2\nu-1,2\nu} & 0 \cdots & 0 & \\ 0 & \cdots & 0 & f_{2\nu,2\nu-1} & f_{2\nu,2\nu} & 0 \cdots & 0 & \\ 0 & \cdots & 0 & 0 & 0 & 0 \cdots & 0 & \\ \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & 0 & 0 \cdots & 0 & \end{pmatrix}, \tag{5.3.3}$$

where

$$\begin{aligned} f_{2\nu-1,2\nu-1} &= \gamma_0(\nu-1) - 2\zeta(\nu)\gamma_0(\nu-1) + \zeta^2(\nu)\gamma_0(\nu-1), \\ f_{2\nu-1,2\nu} &= f_{2\nu,2\nu-1} \\ &= \gamma_1(\nu) - \zeta(\nu)\gamma_1(\nu) - \phi(\nu)\gamma_0(\nu-1) + \phi(\nu)\zeta(\nu)\gamma_0(\nu-1), \\ f_{2\nu,2\nu} &= \gamma_0(\nu) - 2\phi(\nu)\gamma_1(\nu) + \phi^2(\nu)\gamma_0(\nu-1), \end{aligned}$$

where $\gamma_0(\nu)$, $\gamma_1(\nu)$, $\gamma_T(\nu)$ and $\gamma_{T+1}(\nu)$ can be evaluated by (5.2.5) and (5.2.6). Hence the $2T \times 2T$ matrix $F(\underline{\alpha}) = \sum_{\nu=1}^{T} \sigma^{-2}(\nu) F_\nu(\underline{\alpha}, \underline{\sigma}^2)$ is a block diagonal matrix. This indicates that the MLEs for different seasons are asymptotically independent. $\square$

5.4 Concluding Remarks

As we mentioned before, one needs to pay special attention to the parsimony issue when fitting a PARMA model. SPARMA models provide methods that can help achieve this goal. For a general PARMA $(p_1 + p_2T, q_1 + q_2T)$ model, there are $p_1 + q_1 + (p_2 + q_2)T$ parameters (the white noise variances are treated as nuisance parameters). If one can identify the model as a SPARMA$((p_1 \times p_2), (q_1 \times q_2))$, then at least $(p_2+q_2)(T-1)$ parameters can be saved. When the period is large, the number of parameters can be drastically reduced. On the other hand, a SPARMA$((p_1 \times p_2), (q_1 \times q_2))$ model has more parameters than a PARMA (p_1, q_1) model. It would be of interest to test whether the extra $(p_2 + q_2)$ parameters (representing between season correlations) are significant. Due to the relationship between PARMA and SPARMA models, many of the results developed in Chapter 3 and 4 for PARMA models can be extended to SPARMA models. However, further work is needed in dealing with the problems of identification, prediction, and inference for SPARMA models.

Table 5.1: Comparison of PAR(1), SAR(1×1), and SPAR(1×1)

	PAR(1)	SAR(1×1)	SPAR(1×1)
	$X_{nT+\nu} - \phi(\nu)X_{nT+\nu-1}$ $= \epsilon_{nT+\nu}$	$X_{nT+\nu} - \phi X_{nT+\nu-1}$ $-\zeta X_{(n-1)T+\nu}$ $+\phi\zeta X_{(n-1)T+\nu-1}$ $= \epsilon_{nT+\nu}$	$X_{nT+\nu} - \phi(\nu)X_{nT+\nu-1}$ $-\zeta(\nu)X_{(n-1)T+\nu}$ $+\phi(\nu)\zeta(\nu)X_{(n-1)T+\nu-1}$ $= \epsilon_{nT+\nu}$
Model	$\phi(1) = 0.3$ $\sigma^2(1) = 1$	$\phi = 0.3$ $\zeta = 0.1$ $\sigma^2 = 1$	$\phi(1) = 0.3$ $\zeta(1) = 0.1$ $\sigma^2(1) = 1$
Parameters	$\phi(2) = -0.3$ $\sigma^2(2) = 1$		$\phi(2) = -0.3$ $\zeta(2) = -0.1$ $\sigma^2(2) = 1$
	$\phi(3) = -0.9$ $\sigma^2(3) = 0.8$		$\phi(3) = -0.9$ $\zeta(3) = -0.3$ $\sigma^2(3) = 0.8$
	$\phi(4) = -0.5$ $\sigma^2(4) = 0.8$		$\phi(4) = -0.5$ $\zeta(4) = -0.1$ $\sigma^2(4) = 0.8$

Table 5.2: Autocorrelations of PAR(1), SAR(1×1), and SPAR(1×1)

season	lag	PAR(1)	SAR(1×1)	SPAR(1×1)
$\nu = 1$	1	0.3149	.3105	0.3188
	2	-0.1851	.1057	-0.2019
	3	0.1344	.0659	0.1214
	4	-0.0405	.1291	0.0355
	5	-0.0128	.0998	-0.0659
	6	0.0075	.0498	0.1133
	7	-0.0054	.0247	-0.1078
	8	0.0016	.0223	0.0470
	9	0.0005	.0205	0.0384
	10	-0.0003	.0141	-0.0462
$\nu = 2$	1	-0.3014		-0.3118
	2	-0.0949		-0.1469
	3	0.0558		0.1375
	4	-0.0405		-0.1459
	5	0.0122		0.0506
	6	0.0039		0.0441
	7	-0.0023		-0.0539
	8	0.0016		0.0506
	9	-0.0005		-0.0181
	10	-0.0002		-0.0179
$\nu = 3$	1	-0.7259		-0.6853
	2	0.2188		0.2754
	3	0.0689		0.3794
	4	-0.0405		-0.5117
	5	0.0294		0.4807
	6	-0.0089		-0.1715
	7	-0.0028		-0.1682
	8	0.0016		0.2122
	9	-0.0012		-0.2025
	10	0.0004		0.0720
$\nu = 4$	1	-0.5880		-0.6870
	2	0.4268		0.4463
	3	-0.1286		-0.2144
	4	-0.0405		-0.3767
	5	0.0238		0.4464
	6	-0.0173		-0.3838
	7	0.0052		0.1444
	8	0.0016		0.1650
	9	-0.0010		-0.2073
	10	0.0007		0.1931

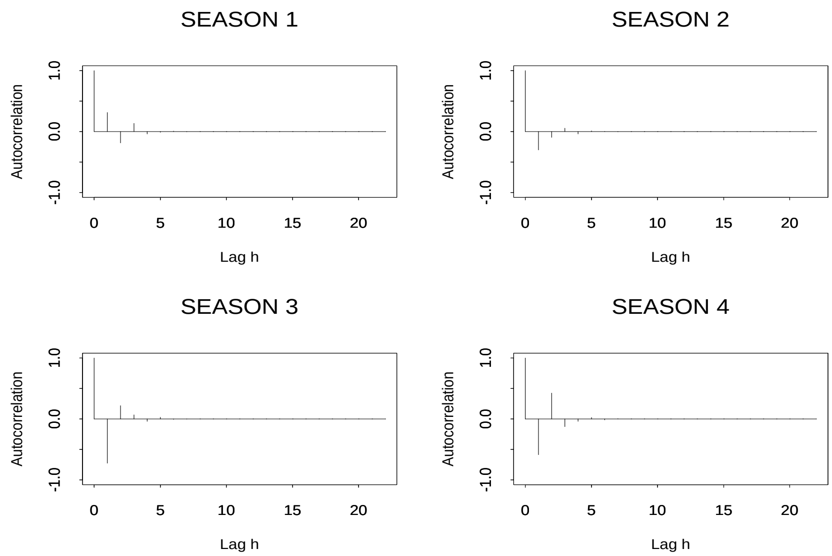

Figure 5.1: Autocorrelations of PAR(1).

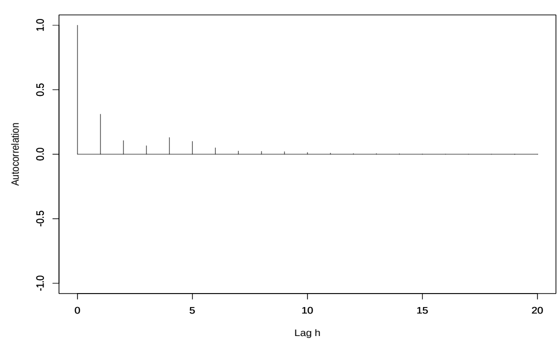

Figure 5.2: Autocorrelations of SAR(1×1).

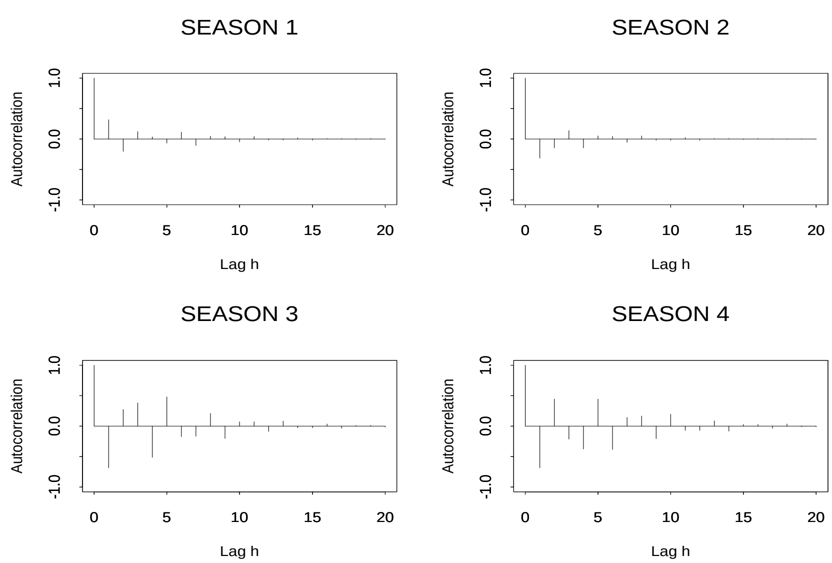

Figure 5.3: Autocorrelations of SPAR(1×1).

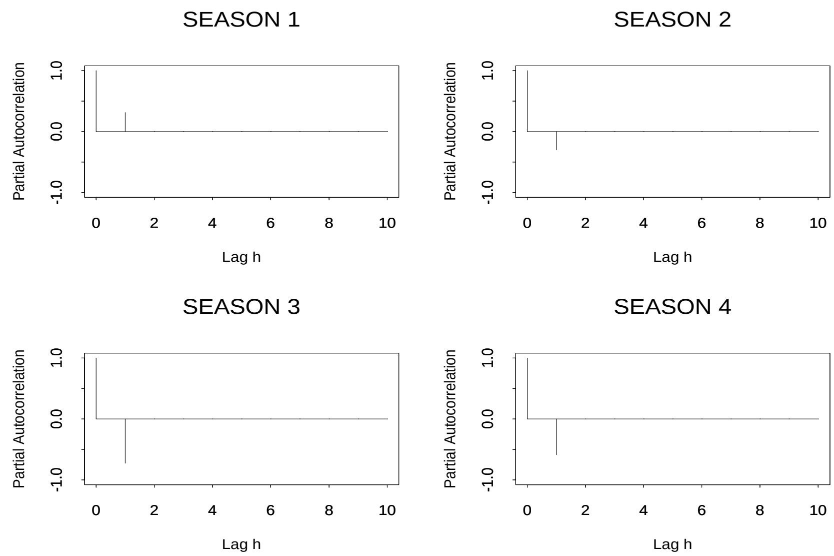

Figure 5.4: Partial Autocorrelations of PAR(1).

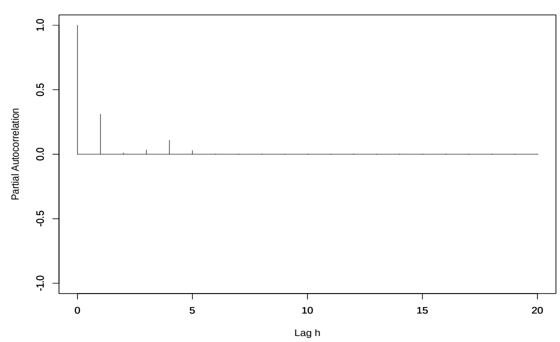

Figure 5.5: Partial Autocorrelations of SAR(1×1).

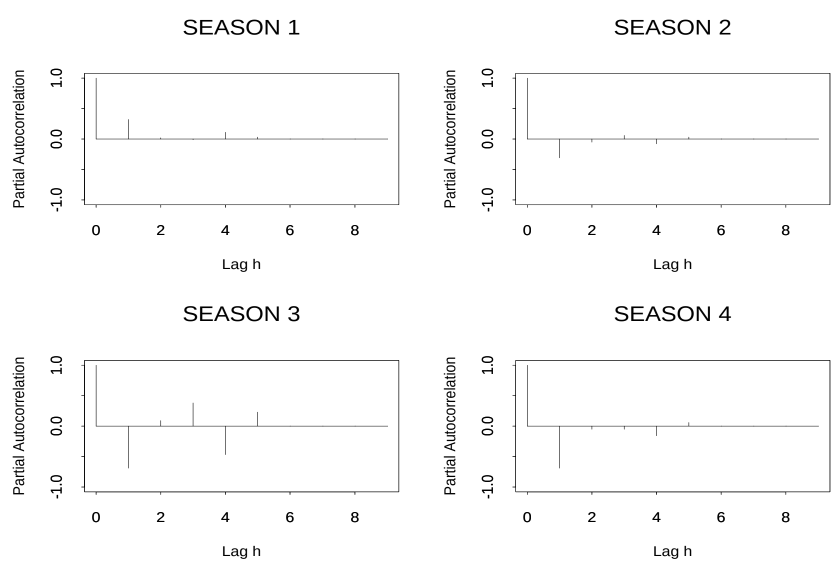

Figure 5.6: Partial Autocorrelations of SPAR(1×1).

Bibliography

Adams, G. J. and Goodwin, G. C. (1995). Parameter estimation for periodic ARMA models. *Journal of Time Series Analysis*, **16**, 127-145.

Anderson, P. L. and Vecchia, A. V. (1993). Asymptotic results for periodic autoregressive moving-average processes. *Journal of Time Series Analysis*, **14**, 1-18.

Bartolini, P., Salas, J. D., and T. B. Obeysekera (1988). Multivariate periodic ARMA(1, 1) processes, *Water Resources Research*, **24**, 1237-1246.

Basawa, I. V. and R. B. Lund (2001). Large sample properties of parameter estimates for periodic ARMA models, *Journal of Time Series Analysis*, **22(4)**, 1-13.

Bentarzi, M. and Hallin, M. (1993). On the invertibility of periodic moving-average models, *Journal of Time Series Analysis*, **15**, 263-268.

Bibi, A. and C. Francq (2001). Consistent and asymptotically normal estimators for time-dependent linear models, To Appear, *The Institute of Statistical Mathematics.*

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

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994). *Time Series Analysis. Forecasting and Control*, Third Edition, Prentice-Hall, Englewood Cliffs, NJ.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, Second Edition, Springer-Verlag, New York.

Cryer, J. D. (1986). *Time Series Analysis*, PWS-Kent, Boston.

Dunsmuir, W. (1981). Estimation of periodically varying means and standard deviations in time series data, *Journal of Time Series Analysis*, **2**, 129-153.

Gardner, W. and Franks, L. (1975). Characterization of cyclostationary random signal processes, *IEEE Transactions on Information Theory*, **21**, 4-41.

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

Hannan, E. J. (1955). A test for singularities in Sydney rainfall, *Australian Journal of Physics*, **8**, 289-297.

Hurd, H. L. and Gerr, N. L. (1991). Graphical methods for determining the presence of periodic correlation. *Journal of Time Series Analysis*, **12**, 337-350.

Hurd, H. L., Makagon, A., and A. G. Miamee (1999). On AR(1) models with periodic and almost periodic coefficients, Preprint, Department of Statistics, The University of North Carolina.

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

Li, W. K. and Hui, Y. V. (1988). An algorithm for the exact likelihood of periodic autoregressive moving average models, *Communications in Statistics, Simulation and Computation*, **17**, 1483-1494.

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

Lund, R. and Basawa, I. V. (1999). Modeling and inference for periodically correlated time series. *Asymptotics, Nonparametrics, and Time Series(ed. S. Ghosh)*, 37-62. Marcel Dekker, New York.

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

Monin, A. (1963). Stationary and periodic time series in the general circulation of the atmosphere, In: *Proceedings Symposium on Time Series Analysis*, Edited by M. Rosenblatt, 144-151.

Morrison, D. F. (1976). *Multivariate Statistical Methods*, Second Edition, McGraw Hill, New York.

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

Parzen, E. and Pagano, M. (1979). An approach to modeling seasonally stationary time series. *Journal of Econometrics*, **9**, 137-153.

Ramsey, F. (1974). Characterization of the partial autocorrelation function, *Annals of Statistics*, **2**, 1296-1301.

Rasmussen, P. F., Salas, J. D., Fagherazzi, L., Rasam, J., and B. Bernard (1996). Estimation and validation of contemporaneous PARMA models for streamflow simulation, *Water Resources Research*, **32**, 3151-3160.

Sakai, H. (1982). Circular lattice filtering using Pagano's method, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **ASSP-30**, 279-287.

Sakai, H. (1983). Covariance matrices characterization by a set of scalar partial autocorrelation coefficients, *Annals of Statistics*, **11**, 337-340.

Seymour (2002). Modeling Plant Growth Using Regression with Periodically Correlated Errors. *Journal of Agricultural, Biological, and Environmental Statistics*, **7**, 1-17.

Thompstone, R. M., Hipel, K. W., and A. I. McLeod (1985). Grouping of periodic autoregressive models, In: *Time Series Analysis: Theory and Practice. Volume 6*, Edited by O. D. Anderson, J. K. Ord, and E. A. Robinson, 35-49.

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

Vecchia, A. V., Obeysekera, J. T., Salas, J. D., and D. C. Boes (1983). Aggregation and estimation for low-order periodic ARMA models, *Water Resources Research*, **5**, 1297-1306.

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

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

Vecchia, A. V. and Ballerini, R. (1991). Testing for periodic autocorrelations in seasonal time series data, *Biometrika*, **78**, 53-63.