2015

Efficient inference for periodic autoregressive coefficients with polynomial spline smoothing approach

Lin Tang
University of Toledo

A Dissertation

entitled

Efficient Inference for Periodic Autoregressive Coefficients with

Polynomial Spline Smoothing Approach

by

Lin Tang

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the

Doctor of Philosophy Degree in Mathematics with emphasis in Statistics

Dr. Qin Shao, Committee Chair

Dr. Donald B. White, Committee Member

Dr. Rong Liu, Committee Member

Dr. Jing Wang, Committee Member

Dr. Patricia R. Komuniecki, Dean
College of Graduate Studies

The University of Toledo
December 2015

An Abstract of

Efficient Inference for Periodic Autoregressive Coefficients with Polynomial Spline Smoothing Approach

by

Lin Tang

Submitted to the Graduate Faculty as partial fulfillment of the requirements for the Doctor of Philosophy Degree in Mathematics with emphasis in Statistics

The University of Toledo
December 2015

First, we propose a two-step estimation method for periodic autoregressive parameters via residuals when the observations contain trend and periodic autoregressive time series. In the first step, the trend is estimated and the residuals are calculated; in the second step, the autoregressive coefficients are estimated from the residuals. To overcome the drawback of a parametric trend estimation, we estimate the trend nonparametrically by polynomial spline smoothing. Polynomial spline smoothing is one of the nonparametric methods commonly used in practice for function estimation. It does not require any assumption about the shape of the unknown function. In addition, it has advantages of computational expediency and mathematical simplicity. The oracle efficiency of the proposed Yule-Walker type estimator is established. The performance is illustrated by simulation studies and real data analysis.

Second, we consider time series that contain a trend, a seasonal component and periodically correlated time series. A semiparametric three-step method is proposed to analyze such time series. The seasonal component and trend are estimated by means of B-splines, and the Yule-Walker estimates of the time series model coefficients are calculated via the residuals after removing the estimated seasonality and trend. The oracle efficiency of the proposed Yule-Walker type estimators is established. Simulation studies suggest that the performance of the estimators coincides with the theoretical results. The proposed method is applied to three data sets.

Third, we will make the inference for the logistic regression models using the nonparametric estimation method. The primary interest of this topic is the estimation of the conditional mean for the logistic regression models. We propose the local likelihood logit method with linear B-spline to estimate the conditional mean. Simulation studies shows that our method works well.

Acknowledgments

Firstly, my sincere appreciation goes to my supervisor, Professor Qin Shao, for her constantly patient guidance and encouragement on my research work, for her precious time on reading my dissertation word by word, for her numerous advice on improving my writing skills, for her highly positive recommendations and suggestions on my job seeking, and for her invaluable help with many other things throughout my study.

Secondly, my great thanks go to Professor Donald White, Professor Biao Zhang, Professor Denis White, Professor Gerard Thompson, Professor Rong Liu for their kind encouragement, valuable advice, and guidance on my study.

Thirdly, I would like to give my gratitude to my fellow graduate students: Dong Zhang, Haifeng Yu, Gang Liu, Jingning Mei, and Kaili Ren for their time on discussing the topics together in our seminars. I benefit much from the discussions with them. And many thanks go to Jie Chen, Qing Zhao, Hui Wang, Yanmei Xie, Wencan He, Suohong Wang, and Xuan Li for their help not only with my study but also many other things during our stay in The University of Toledo.

Deep love and appreciation go to my parents. Without their support and encouragement, this dissertation could not have been completed.

Contents

List of Tables

List of Figures

List of Abbreviations

AR	Autoregressive Models
ARMA	Autoregressive Moving Average Models
ACVF	autocovariance function
ACF	autocorrelation function
B	B-Spline
Ber	Bernoulli
E	Estimation Method
L	Linear Regression
M	Model
MA	Moving Average Models
MAE	mean absolute error
MSE	mean squared error
P	Parameter
PAR	Periodic Autoregressive Models
SD	Standard Deviation

Chapter 1

Literature Review

1.1 Examples of Time Series

A time series data are a set of observations $\{x_t\}_{t=1}^n$, each one of which is recorded at a specified time point t. The observations can be recorded either at a discrete set of time points or continuously over some interval. If observations are obtained at a discrete set of time points, it is discrete-time series. Discrete-time series will be considered in this dissertation. I will first show three examples of time series in the following.

Example 1.1: Population of the U.S.A., 1790-1980.

Figure 1-1 suggests a roughly exponential trend for population of the U.S.A. from 1790 to 1980. The population was measured every ten years.

Example 1.2: Wolfer Sunspot Numbers, 1770-1869.

Figure 1-2 appears to have an apparent periodic component with period of about 11 years. The sunspot number was measured each year from 1770 to 1869.

Example 1.3: Australian Red Wine Sales, January 1980-July 1995.

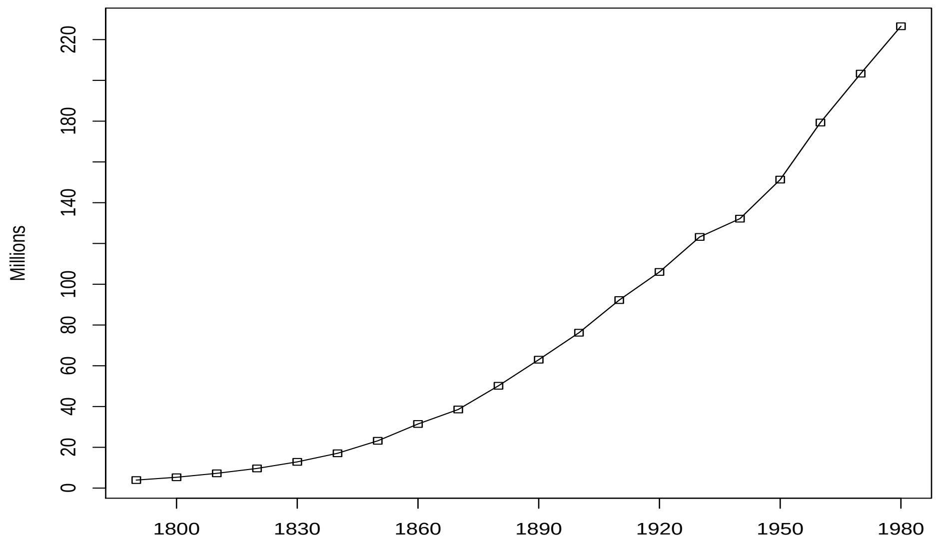

Figure 1-1: Population of the U.S.A., 1790-1980

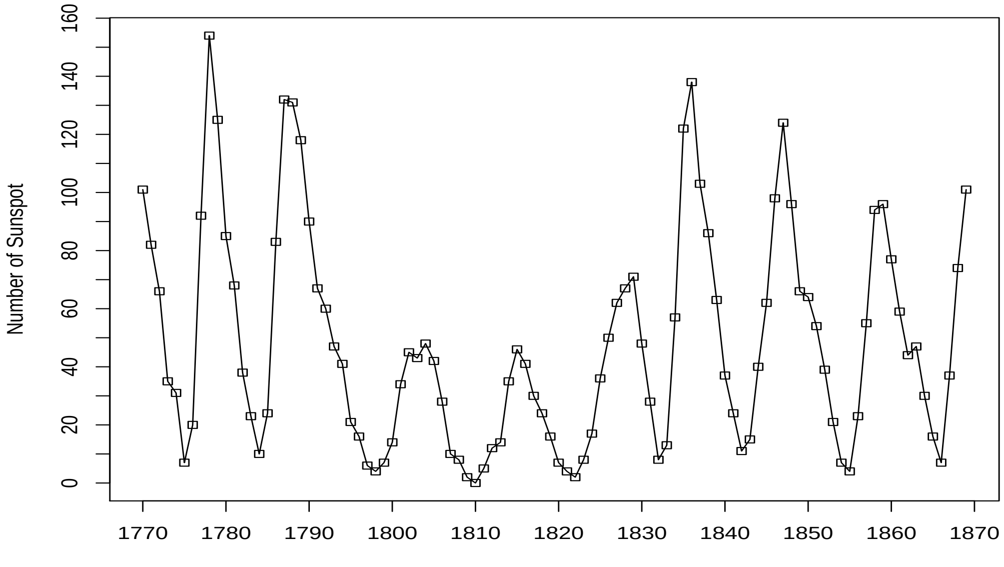

Figure 1-2: Wolfer Sunspot Numbers, 1770-1869

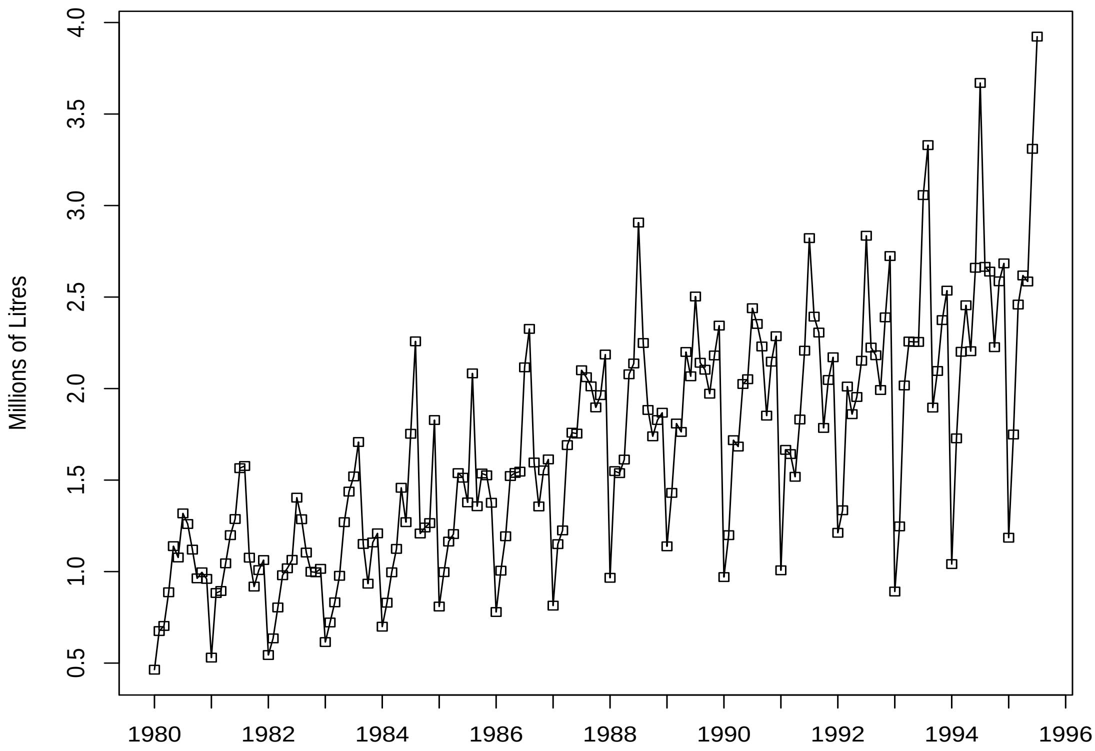

Figure 1-3: The Australian Red Wine Sales, January 1980-July 1995

The data in this example are the monthly Australian red wine sales by wine makers from January 1980 through July 1995. There are a total of 187 observations. In Figure 1-3, the sales show an upward trend and a seasonal pattern with a peak in July and a trough in January for each year.

The above sample data sets are only some of numerous time series in real world. There are many other examples of time series data in the fields of engineering, science, finance, economics, business, and sociology, such as monthly streamflows, monthly temperatures, monthly stock returns series of one company, and monthly or quarterly sales data.

Throughout the dissertation, we will use lower case letters to denote both random variables and observations, which is slightly different from the convention of statistics. However, the meaning of the letter is always clear according to the context. Also, we will use bold lower case letters for vectors and bold upper case letters for matrices.

1.2 Concepts and Definitions in Time Series Theory

Definition 1. *Let $\{x_t, t = 0, \pm 1, \pm 2, \ldots\}$ be a time series with $E(x_t^2) < \infty$. For all t, the mean function of $\{x_t\}$ is*

$$\mu(t) = E(x_t).$$

The covariance function of $\{x_t\}$ is

$$\gamma(r, s) = Cov(x_r, x_s) = E[(x_r - \mu(r))(x_s - \mu(s))],$$

for all integers r and s.

Definition 2. *The time series $\{x_t, t \in \mathbb{Z}\}$, with $\mathbb{Z} = 0, \pm 1, \pm 2, \ldots$, is said to be weakly stationary if*

(i) $E(x_t^2) < \infty$ for all $t \in \mathbb{Z}$,

(ii) $\mu(t)$ is independent of t for all integers t,

(iii) $\gamma(t + k, t)$ is independent of t for each integer k.

When the term stationary is used in this dissertation, it means weakly stationarity. Another important notion of stationarity is strict stationarity.

Definition 3. *The time series* $\{x_t, t \in \mathbb{Z}\}$, *with* $\mathbb{Z} = 0, \pm 1, \pm 2, \ldots$ *is said to be strictly stationary if* $(x_1, \ldots, x_n)$ *and* $(x_{1+k}, \ldots, x_{n+k})$ *have the same joint distributions for all integers* k *and all positive integers* n.

If $\{x_t\}$ is strictly stationary and $E(x_t^2) < \infty$ for all integers t, then $\{x_t\}$ is also weakly stationary.

When a time series is stationary, the covariance function can be rewritten as a function of one variable, defined by $\gamma(k) = \gamma(t+k, t)$, where $\gamma(\cdot)$ is referred to as the autocovariance function and $\gamma(k)$ is the value of the autocovariance function at lag k. The autocovariance function is defined as follows:

Definition 4. *Let* $\{x_t, t \in \mathbb{Z}\}$ *be a stationary time series. The autocovariance function (ACVF) of* $\{x_t\}$ *at lag* k *is*

$$\gamma(k) = \gamma(t+k, t) = Cov(x_{t+k}, x_t).$$

The autocorrelation function (ACF) of $\{x_t\}$ *at lag* k *is*

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)} = Corr(x_{t+k}, x_t).$$

The following is one example of stationary time series, which is referred to as white noise.

Definition 5. *If* $\{\epsilon_t\}$ *is a sequence of uncorrelated random variable, each with zero mean and variance* σ^2, *then such a series is referred to as white noise. It is denoted by* $\{\epsilon_t\} \sim WN(0, \sigma^2)$.

Intuitively, white noise series is random with no systematic structures. In addition, white noise $\{\epsilon_t\}$ is stationary and its covariance function is

$$\gamma(k) = \begin{cases} \sigma^2, & \text{if } k = 0, \\ 0, & \text{if } k \neq 0. \end{cases}$$

It is easy to calculate the autocovariance function and the autocorrelation function for simple time series models. However, we usually do not know the model for the given data in real world. Thus we need to select an appropriate model to represent the dependence in the data. One of the useful tools for determining the appropriate model is the sample autocorrelation function of the data. The sample ACF is a consistent estimate of the autocovariance function (see Brockwell and Davis (1991)), and can provide us with information about an appropriate stationary time series model for representing the dependence in the data. The definitions of a sample autocovariance function and sample autocorrelation function are as follows:

Definition 6. *Let $x_1, x_2, \ldots, x_n$ be observations of a stationary time series. The sample mean of $x_1, x_2, \ldots, x_n$ is*

$$\bar{x} = \frac{1}{n} \sum_{t=1}^{n} x_t.$$

The sample autocovariance function (sample ACVF) at lag k is

$$\hat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (x_{t+|k|} - \bar{x})(x_t - \bar{x}), \qquad -n < k < n.$$

The sample autocorrelation function (sample ACF) at lag k is

$$\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)}, \qquad -n < k < n.$$

The sample ACVF is the element of the sample covariance matrix. It is an important tool for time series model diagnostics. Both estimators $\hat{\gamma}(k)$ and $\hat{\rho}(k)$ are biased, but they are nearly unbiased for large sample sizes. The sample covariance matrix is $\hat{\mathbf{\Gamma}}_n = [\hat{\gamma}(i-j)]_{i,j=1}^{n}$.

1.3 Simple Time Series Models

Generally speaking, there are three types of simple time series models that are used for modeling stationary time series: autoregressive models (AR), moving average models (MA), and autoregressive moving average models (ARMA).

1.3.1 Autoregressive Models (AR)

AR models are similar to linear regression models. If we use the past values of a time series as predictors in a classical linear regression model, then we obtain an AR model. The dependent variable is the value at the current time and the independent variables are the lagged values in an AR model. Most of the statistical properties and conclusions for classical linear regression models can be generalized to AR models with slight modifications. Since the classical linear regression models were studied and used very widely, AR models have become one of the most used time series models. An AR(p) model can be written as

$$\phi(B)x_t = \epsilon_t, \tag{1.1}$$

where $\{x_t\}$ is time series, p is the order of the AR model, $\{\epsilon_t\} \sim WN(0, \sigma^2)$, B is the backward shift operator with $\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$ and $B^k x_t = x_{t-k}$, $k = 1, 2, \ldots, p$. Thus we can have

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \epsilon_t. \tag{1.2}$$

A formal definition of an AR(p) model is as follows (e.g. Chan (2010))

Definition 7. $\{x_t\}$ *is said to be an AR(p) process if*

(i) $\{x_t\}$ *is stationary,*

(ii) $\{x_t\}$ *satisfies equation (1.1) for all* t*, where* $Bx_t = x_{t-1}$.

We assume that the mean μ of $\{x_t\}$ is zero in Definition 7. When it is not zero, if $\{x_t - \mu\}$ is an AR(p) process then $\{x_t\}$ is said to be an AR(p) process with mean μ.

Next, let us introduce the Yule-Walker estimation of the parameters $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$ and σ^2 for an AR(p) model from the observations $x_1, \ldots, x_n$. The Yule-Walker method is commonly used for estimation of AR model parameters because of its simplicity and asymptotic efficiency. Causality is needed to ensure the asymptotic distribution of Yule-Walker estimators.

Definition 8. *The time series* $\{x_t\}$ *is causal if for all* t*, there exists a sequence of constants* $\{\psi_j\}$ *such that* $\sum_{j=0}^{\infty} |\psi(j)| < \infty$ *and*

$$x_t = \sum_{j=0}^{\infty} \psi(j)\epsilon_{t-j}.$$

We multiply each side of equation (1.2) by x_{t-j}, $j = 1, 2, \ldots, p$ and take expectations on both sides. We can calculate the expected values of both sides of the equation systems and obtain the Yule-Walker equations:

$$\boldsymbol{\phi} = \boldsymbol{\Gamma}_p^{-1}\boldsymbol{\gamma}_p \tag{1.3}$$

and

$$\sigma^2 = \gamma(0) - \boldsymbol{\phi}'\boldsymbol{\gamma}_p, \tag{1.4}$$

where $\boldsymbol{\Gamma}_p$ is the covariance matrix $[\gamma(i-j)]_{i,j=1}^{p}$ and $\boldsymbol{\gamma}_p = (\gamma(1), \gamma(2), \ldots, \gamma(p))'$. If

$\gamma(0) > 0$ and $\gamma(k) \to 0$ as $k \to \infty$ then the covariance matrix $\mathbf{\Gamma}_p$ is nonsingular for every p (see Brockwell and Davis (1991)).

We need to estimate model parameter $\boldsymbol{\phi}$ from equation (1.3). The difficulty is that, however, $\gamma(k)$ is unknown. We use the corresponding sample covariances $\hat{\gamma}(k)$, $k = 0, 1, \ldots, p$ to replace the covariances $\gamma(k)$. We obtain the Yule-Walker equations as follows:

$$\hat{\boldsymbol{\phi}} = \hat{\mathbf{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p = \hat{\boldsymbol{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p \tag{1.5}$$

and

$$\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\phi}}' \hat{\boldsymbol{\gamma}}_p = \hat{\gamma}(0)[1 - \hat{\boldsymbol{\rho}}_p' \hat{\boldsymbol{R}}_p^{-1} \hat{\boldsymbol{\rho}}_p], \tag{1.6}$$

where $\hat{\mathbf{\Gamma}}_p = [\hat{\gamma}(i-j)]_{i,j=1}^p$, $\hat{\boldsymbol{\gamma}}_p = (\hat{\gamma}(1), \hat{\gamma}(2), \ldots, \hat{\gamma}(p))$, $\hat{\boldsymbol{\rho}}_p = (\hat{\rho}(1), \ldots, \hat{\rho}(p))' = \hat{\boldsymbol{\gamma}}_p / \hat{\gamma}(0)$, and $\hat{\boldsymbol{R}}_p = \hat{\mathbf{\Gamma}}_p / \hat{\gamma}(0)$. If $\hat{\gamma}(0) > 0$, then $\hat{\mathbf{\Gamma}}_p$ is nonsingular for every p (see Brockwell and Davis (1991)).

The asymptotic distribution of the Yule-Walker estimators is normal (see Brockwell and Davis (1991)): If $\{x_t\}$ is the causal AR(p) process (1.2), and $\hat{\boldsymbol{\phi}}$ is the Yule-Walker estimator of $\boldsymbol{\phi}$, then

$$\sqrt{n}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \overset{D}{\Longrightarrow} N(\mathbf{0}, \Sigma \mathbf{\Gamma}_p^{-1}), \tag{1.7}$$

and

$$\hat{\sigma}^2 \overset{P}{\longrightarrow} \sigma^2. \tag{1.8}$$

The order p of an AR time series model is unknown in application and is usually determined empirically. Some general methods have been extensively studied for the order determination in the time series literature. One of the most commonly used methods is information approach (Tsay (2010)). The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are two maximum-likelihood based information criteria to determine the orders of AR(p) models. We select the order p such that the model has the minimum AIC or BIC value.

After an AR model is fitted, the model should be examined carefully to check for adequacy. As indicated by Tsay (2010), if the model is adequate, then the residual series should behave as a white noise. Thus the sample ACF of the residuals is used to detect any discrepancy. If the sample ACF shows that a fitted model is inappropriate, then it should be refined. For example, we may need to remove some insignificant parameters to refine the fitted AR model. In addition, the sample ACF can be used to check for additional serial correlations.

1.3.2 Moving Average Models (MA)

MA models are another class of simple time series models. They can be regarded as a simple extension of white noise series. In particular, a moving average model of order q (MA(q)) is essentially a weighted average of $\{\epsilon_t\}$, and is defined as follows:

$$x_t = \epsilon_t + \theta_1\epsilon_{t-1} + \ldots + \theta_q\epsilon_{t-q}, \quad \epsilon_t \sim WN(0, \sigma^2). \tag{1.9}$$

Moving average models are always weakly stationary since they are finite linear combinations of a white noise series.

Next, we will briefly introduce the innovations algorithm, see Brockwell and Davis (2002). The innovations algorithm is recursive and applicable to all time series with finite moments, regardless of whether they are stationary or not. In particular, it works extremely well to estimate the MA model parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$ as well as the variance of the white noise from observations $x_1, \ldots, x_n$.

When we apply the innovations algorithm to a MA model with parameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q)'$, we assume that in this algorithm equations the parameters are $\boldsymbol{\theta}_n = (\theta_{n1}, \ldots, \theta_{nn})'$ with $n = 1, 2, \ldots$, where $\theta_{nj} = 0$ for $q + 1 \leq j \leq n$, and that the white noise variances are ν_n with $n = 1, 2, \ldots$. Then MA model parameters can be computed recursively by the innovations algorithm from the following equations:

$$\theta_{n,n-k} = \nu_k^{-1} \left(\gamma(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}\nu_j \right), \quad k = 0, 1, \ldots, n-1, \tag{1.10}$$

$$\nu_0 = \gamma(1,1), \tag{1.11}$$

$$\nu_n = \gamma(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j, \tag{1.12}$$

where $\gamma(n+1, k+1) =$E$(x_{n+1}x_{k+1})$ since E$(x_j) = 0$ for all integers j. For the above algorithm, ν_0 should be first solved, then in the following order for θ_{11}, ν_1; θ_{22}, θ_{21}, ν_2; θ_{33}, θ_{32}, θ_{31}, ν_3; $\cdots$. Then we can replace the autocovariance function in the algorithm by the sample ACVF when we compute $\hat{\boldsymbol{\theta}}_n$ and $\hat{\nu}_n$. Thus these coefficients can be estimated from observations $x_1, \ldots, x_n$. For a choice of order q, the sample ACF is useful to estimate MA model parameter.

We can obtain a one-step ahead predictor $\hat{x}_{n+1}$, $n \geq 0$ for MA time series based on the innovations algorithm as follows:

$$\hat{x}_{n+1} = \begin{cases} 0, & \text{if } n = 0, \\ \sum_{j=1}^{n} \hat{\theta}_{nj}(x_{n+1-j} - \hat{x}_{n+1-j}) & \text{if } n \geq 1. \end{cases} \tag{1.13}$$

where $\hat{\boldsymbol{\theta}}_n$ is computed recursively by equations (1.10), (1.11), and (1.12).

1.3.3 Autoregressive Moving Average Models (ARMA)

In some applications, we may need high-order AR(p) or MA(q) models with many parameters to adequately describe the dynamic structure of the data. To overcome this difficulty, a new model class, called autoregressive moving average (ARMA) models, is introduced (Box, Jenkins, and Reinsel (2008)). An ARMA model combines the ideas of AR and MA models into a compact model such that the number of parameters is reduced and parsimony can be achieved.

Definition 9. *$\{x_t\}$ is an ARMA(p, q) process if it is stationary and if for every t,*

$$x_t - \phi_1 x_{t-1} - \ldots - \phi_p x_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q}, \tag{1.14}$$

where $\epsilon_t \sim WN(0, \sigma^2)$ and the polynomials $1 - \phi_1 z - \ldots - \phi_p z^p$ and $1 + \theta_1 z + \ldots + \theta_q z^q$ have no common factors.

The process $\{x_t\}$ is said to be an ARMA(p, q) process with mean μ if $\{x_t - \mu\}$ is an ARMA(p, q) process. It is useful to adopt the backward shift operator to obtain the more concise form of the ARMA model in (1.14)

$$\phi(B) x_t = \theta(B) \epsilon_t, \tag{1.15}$$

where

$$\phi(B) = 1 - \phi_1 B - \ldots - \phi_p B^p,$$

$$\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q,$$

and B is the backward shift operator ($B^j x_t = x_{t-j}, B^j \epsilon_t = \epsilon_{t-j}, |j| \geq 0$).

There are many estimation methods for ARMA model parameters. We briefly introduce the two-step Hannan-Rissanen algorithm. In the first step, we suppose $q =$

0 and obtain the preliminary parameter estimates using the least squares regression method, since an AR model has the form of a linear regression model. In the second step, we replace the unobserved white noise series $\epsilon_{t-1}, \ldots, \epsilon_{t-q}$ in ARMA model (1.14) by the residuals $\hat{\epsilon}_{t-1}, \ldots, \hat{\epsilon}_{t-q}$ from the first step. The parameters $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are then estimated by the least squares method for regression x_t on x_{t-1}, …, x_{t-q}, $\hat{\epsilon}_{t-1}$, …, $\hat{\epsilon}_{t-q}$.

1.4 Periodic Autoregressive Models (PAR)

Although AR(p) are appropriate for many stationary time series, they cannot describe periodically correlated time series data which often occur in climatology, hydrology, economics, biology and environmental toxicology. Unlike a stationary time series, a periodic correlated time series has a time varying autocovariance function. It depends on both the time point and the lag. However, the autocovariance function repeats the same pattern after the period T. The value of the period T depends on the nature of the observations. For example, it is typically 12 if the observations are monthly. A unique feature of a periodic time series is that each season ν which is a time point within the period T has a different autocovariance function.

A time series $\{x_t\}_{t=-\infty}^{\infty}$ is said to be periodic or periodically correlated with period T if the first two moments are bounded and satisfy

$$E(x_{i+T}) = E(x_i), \tag{1.16}$$

and

$$Cov(x_{i+T}, x_{j+T}) = Cov(x_i, x_j), \tag{1.17}$$

for all i and j, where T is the smallest positive integer satisfying the equation (1.16) and (1.17).

Let $\{x_t\}$ be a periodically correlated time series with the period T. Then for the season ν, where $\nu = 1, 2, \ldots, T$, the autocovariance function (ACVF) of $\{x_t\}$ at lag $k \geq 0$ is

$$\gamma_\nu(k) = Cov(x_{iT+\nu}, x_{iT+\nu-k}). \tag{1.18}$$

The autocorrelation function (ACF) at lag k is

$$\rho_\nu(k) = \frac{\gamma_\nu(k)}{\sqrt{\gamma_\nu(0)\gamma_{\nu-k}(0)}}. \tag{1.19}$$

From (1.18), it is obvious that the autocovariance function of a periodically correlated time series does not depend on the cycle i, but does depend on both season ν and lag k. The periodic notation $x_{iT+\nu}$ denotes series at the νth season of the ith cycle.

Generally speaking, periodically correlated time series are not stationary unless $T = 1$. Therefore inference for periodically correlated time series is much more complicated than that for stationary time series. For example, a scatter time series plot is very useful to identify any structural pattern in the first moments, however, it is not so helpful to determine whether there exists periodicity in the second moments. Hurd and Gerr (1991), and Lund and Basawa (1999) proposed some easy and useful methods to detect periodicity in the second moments of a series by graphing the average squared coherence of the series.

Periodic autoregressive (PAR) models are often used for making inference about such periodic time series. If the mean of a periodically correlated time series is zero, it is usually described by the following periodic autoregressive model with order p

(PAR(p)):

$$x_{iT+\nu} - \sum_{k=1}^{p} \phi_k(\nu) x_{iT+\nu-k} = \sigma_\nu \epsilon_{iT+\nu}, \tag{1.20}$$

where $\{\epsilon_{iT+\nu}\}$ is white noise with $\mathrm{E}(\epsilon_{iT+\nu}) = 0$ and $\mathrm{Var}(\epsilon_{iT+\nu}) = 1$. To emphasize the periodicity, the time index t is often written as $iT + \nu$, where i is the number of cycle, and it is an integer, T is the period, and ν ($1 \leq \nu \leq T$) is called the season and is the seasonal time-point in a period. If a periodically correlated time series $\{x_t\}_{t=-\infty}^{\infty}$ satisfies the equation (1.20), then it is called periodic autoregressive time series.

This model class has been applied to periodic data in climatology (Lund *et al.* (2006)), hydrology (Hipel and McLeod (1994)), economics (Parzen and Pagano (1979)), biology (Seymour (2002)) and environmental toxicology (Shao (2006)). Although model (1.20) is different from seasonal autoregressive time series (Box *et al.* (1994)), as pointed out in Lund and Basawa (1999), it is closely related to an autoregressive model with order p (AR(p)). That is, on one hand, it becomes an AR(p) when the period $T = 1$, and on the other hand, it is an AR(p) for each ν.

Definition 10. *(Causality Condition) The periodically correlated time series $\{x_{iT+\nu}, 1 \leq \nu \leq T\}_{i=0}^{n_T-1}$ is causal; that is, for each fixed ν ($1 \leq \nu \leq T$), there exists a sequence of constants $\{\psi_\nu(j)\}_{j=0}^{\infty}$ such that $\sum_{j=0}^{\infty} |\psi_\nu(j)| < \infty$ and*

$$x_{iT+\nu} = \sum_{j=0}^{\infty} \psi_\nu(j) \epsilon_{iT+\nu-j}.$$

As in an AR model, the order p of a PAR time series model is unknown in application, and has to be determined empirically. We can use information criteria, such as AIC and BIC, to determine the order p of a PAR model; that is, we select the order p such that the model has the minimum AIC or BIC value. After a PAR model is fitted, the model should be examined carefully to check for inadequacy. The sample

ACF of the residuals can be used to check for the model inadequacy. It is useful for us to check for additional serial correlations. If the sample ACF shows that a fitted model is inadequate, then it should be refined.

Denote the PAR(p) model coefficients by $\boldsymbol{\phi} = (\boldsymbol{\phi}'(1), \cdots, \boldsymbol{\phi}'(T))'$ with $\boldsymbol{\phi}(\nu) = (\phi_1(\nu), \cdots, \phi_p(\nu))'$. The inference for $\boldsymbol{\phi}$ based on the observations $\{x_t\}_{t=1}^n$ or $\{x_{iT+\nu}, 1 \leq \nu \leq T\}_{i=0}^{n_T-1}$, where n is the number of observations and n_T is the number of periods, was considered by many researchers. For example, Pagano (1978) discussed the computation and asymptotic properties of Yule-Walker estimators; Basawa and Lund (2001) considered the asymptotic properties of the estimator from estimating equations; Anderson and Meerschaert (2005) derived the large sample distribution of the estimator obtained by the innovations algorithm.

We assume that the mean of a periodic autoregressive time series PAR(p) $\{x_t\}_{t=-\infty}^{\infty}$ is zero. Thus the autocovariance function is $\gamma_\nu(k) = \mathrm{E}(x_{iT+\nu}x_{iT+\nu-k})$. Notice that different from the autocovariance function of a stationary autoregressive, $\gamma_\nu(k)$ is not symmetric for a fixed season ν, that is $\gamma_\nu(k) \neq \gamma_\nu(-k)$. Instead, $\gamma_\nu(k) = \gamma_{\nu-k}(-k)$ and $\gamma_{-\nu}(k) = \gamma_{\nu+T}(k)$. The details can be found, for example, in Shao and Lund (2004). For such a time series with mean zero the sample autocovariance at lag k are computed as follows:

$$\tilde{\gamma}_\nu(k) = \frac{1}{n_T}\sum_{i=i_0}^{n_T-1} x_{iT+\nu}x_{iT+\nu-k}, \quad k \geq 0 \tag{1.21}$$

where i_0 is the smallest integer such that $i_0T + \nu - k \geq 1$.

The Yule-Walker estimators for PAR models are obtained as follows. We multiply each side of equation (1.20) by x_{t-j}, $j = 1, 2, \ldots, p$ and take expectations on both sides. We can calculate the expected values of both sides of the equation systems. Then let $p \times p$ matrix $\tilde{\mathbf{\Gamma}}_\nu$ be symmetric $(1 \le \nu \le T)$, the (i, j)-th entry $(\tilde{\mathbf{\Gamma}}_\nu)_{i,j}$ of which is calculated as follows:

$$
\begin{aligned}
&(\tilde{\mathbf{\Gamma}}_\nu)_{1,j} = \tilde{\gamma}_{\nu-1}(j-1), \quad j = 1, \dots, p, \\
&(\tilde{\mathbf{\Gamma}}_\nu)_{2,j} = \tilde{\gamma}_{\nu-2}(j-2), \quad j = 2, \dots, p, \\
&\vdots \\
&(\tilde{\mathbf{\Gamma}}_\nu)_{\nu-1,j} = \tilde{\gamma}_{\nu-(\nu-1)}(j-(\nu-1)) = \tilde{\gamma}_1(j-(\nu-1)), \quad j = \nu-1, \dots, p, \\
&(\tilde{\mathbf{\Gamma}}_\nu)_{\nu,j} = \tilde{\gamma}_T(j-\nu), \quad j = \nu, \dots, p, \\
&(\tilde{\mathbf{\Gamma}}_\nu)_{\nu+1,j} = \tilde{\gamma}_{T-1}(j-(\nu+1)), \quad j = \nu+1, \dots, p, \\
&\vdots \\
&(\tilde{\mathbf{\Gamma}}_\nu)_{p,j} = \tilde{\gamma}_{T-(p-\nu)}(j-p), \quad j = p.
\end{aligned}
\tag{1.22}
$$

Then let $\tilde{\mathbf{\Gamma}}$ be the diagonal partitioned matrix with $\mathrm{diag}\left(\tilde{\mathbf{\Gamma}}\right) = (\tilde{\mathbf{\Gamma}}_1, \cdots, \tilde{\mathbf{\Gamma}}_T)$, and pT dimensional vector $\tilde{\boldsymbol{\gamma}} = (\tilde{\boldsymbol{\gamma}}_1', \tilde{\boldsymbol{\gamma}}_2', \dots, \tilde{\boldsymbol{\gamma}}_T')'$ with $\tilde{\boldsymbol{\gamma}}_\nu = (\tilde{\gamma}_\nu(1), \tilde{\gamma}_\nu(2), \cdots, \tilde{\gamma}_\nu(p))'$. Then the Yule-Walker estimator of $\boldsymbol{\phi}$ from $\{x_{iT+\nu}, 1 \le \nu \le T\}_{i=0}^{n_T-1}$ is defined by

$$
\tilde{\boldsymbol{\phi}} = \tilde{\mathbf{\Gamma}}^{-1}\tilde{\boldsymbol{\gamma}} \text{ and } \tilde{\sigma}_\nu^2 = \tilde{\gamma}_\nu(0) - \sum_{k=1}^{p} \tilde{\phi}_k(\nu)\, \tilde{\gamma}_{\nu-k}(-k). \tag{1.23}
$$

However, it is not uncommon that the observations are contaminated by trend and thus the periodic autoregressive time series $\{x_t\}_{t=-\infty}^{\infty}$ cannot be observed. Instead the observations are the time series $\{y_t\}_{t=1}^{n}$ that contains a trend function and satisfy the following model:

$$
y_{iT+\nu} = g(u_{iT+\nu}) + x_{iT+\nu}, \tag{1.24}
$$

where $u_{iT+\nu} = (iT+\nu)/n$ and $g(\cdot)$ represents the unknown trend function defined in the interval $[0, 1]$. Model (1.24) can be rewritten in vector format as

$$\boldsymbol{y} = \boldsymbol{g} + \boldsymbol{x}, \tag{1.25}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{g} = (g(u_1), g(u_2), \ldots, g(u_n))'$ with $u_i = i/n$, and $\boldsymbol{x} = (x_1, \ldots, x_n)'$ is a PAR(p) time series defined by (1.20).

Many results have been established when the trend is estimated parametrically. The idea of a parametric method is to assume that the trend follows a known function with an explicit format and some unknown parameters. Such a parametric approach has been applied to model (1.24). For example, Anderson and Meerscheart (1998) took the seasonal-varying constant trend into account; Lund (2006) estimated the trend using a linear function. Although parametric trend estimation works well when the shape of the trend can be approximated by a known analytical function, its major drawback is that it is subjective and it is easily exposed to the risk of model mispecification. For example, as shown in Section 2.4, if the trend function is misspecified, the inference based on the residuals can be very misleading.

To overcome the drawback of a parametric trend estimation, we estimate the trend nonparametrically by polynomial spline smoothing. Polynomial spline smoothing is one of the nonparametric methods commonly used in practice for function estimation. It inherits the advantage shared by all nonparametric methods. That is, it does not require any assumption about the shape of the unknown function. In addition, compared with a local linear method (Fan and Gijbels (1996)), it has advantages of computational expediency and mathematical simplicity. In particular, as shown in Xue and Yang (2006) and Wang and Yang (2007), it needs only a single optimization and thus is much faster than local linear method. Although Truong (1991) and Qiu *et al.* (2013) showed that when trend is estimated by kernels and local linear method, respectively, under model (1.24), the Yule-Walker estimator of $\boldsymbol{\phi}$ from the residuals is as efficient as the one from $\{x_t\}_{t=1}^n$, polynomial spline smoothing is preferred due to the above reason when the data set is very large. Polynomial spline smoothing has been applied to time series settings recently, for example, in Huang and Yang (2004), Wang and Yang (2009), Song and Yang (2010), Wang and Yang (2010), Shao and Yang (2011), Shao and Yang (2012).

Next, let us introduce the basis of B-splines, which is used in the polynomial spline smoothing. Consider a sequence of equally spaced points $\{(-m+1)h,\ (-m+2)h,\ \ldots\ ,-h,\ 0,\ h,\ 2h,\ \ldots\ ,Nh,\ 1\}$, where m is a positive integer. Notice that there are $N+m$ subintervals $J_j = [jh, (j+1)h), j = -m+1, -m+2, \ldots, N-1$ and $J_N = [Nh, 1]$, of length $h = (N+1)^{-1}$. The interval $[0,1]$ can be divided into these positive subintervals. Let $G_N^{(m-2)} = G_N^{(m-2)}[0,1]$ denote the space of polynomial functions of degree $m-1$ on each J_j and have continuous $(m-2)$-th derivatives. Then the B-spline basis of $G_N^{(m-2)}$ is $\{b_{j,m}(u), j = -m+1, \ldots, N\}$. We are going to discuss two cases: $G_N^{(-1)}$ is constant on each J_j, where $m = 1$; $G_N^{(0)}$ is linear on each J_j and continuous on $[0,1]$, where $m = 2$. For case $G_N^{(-1)}$, the B-spline basis is $\{b_{j,1}(u)\}_{j=0}^N$, where $b_{j,1}(u)$ is defined as follows:

$$b_{j,1}(u) = \begin{cases} 1, & u \in J_j, \\ 0, & \text{otherwise.} \end{cases} \tag{1.26}$$

For case $G_N^{(0)}$, the B-spline basis is $\{b_{j,2}(u)\}_{j=-1}^N$, where $b_{j,2}(u)$ is defined as follows:

$$b_{j,2}(u) = K\left(\frac{u-(j+1)h}{h}\right), \tag{1.27}$$

where $K(x) = (1-|x|)_+$ with $(x)_+ = \max(x,0)$. Thus we can write down $b_{j,2}(u)$

specifically as follows:

$$b_{j,2}(u) = \begin{cases} \frac{u}{h} - j, & u \in J_j, \\ j + 2 - \frac{u}{h}, & u \in J_{j+1}, \\ 0, & \text{otherwise.} \end{cases} \tag{1.28}$$

In some fields, time series often contain trend, seasonality, and periodically correlated random components due to the seasonal and periodic nature of dynamical systems. The inference for this type of time series can be considered in partially linear models. Partially linear models have received considerable attention due to their flexibility, relative computation simplicity, and wide applications. Research on this model class has focused on the inference of the nonparametric function and parametric regression component. For example, see the works of Heckman (1986), Speckman (1988), Gao (1995), Ferreira *et al.* (2000), and You and Chen (2007). In particular, Speckman (1988) provided a detailed discussion about inference for partially linear models with independent and identically distributed error terms. They focused on the inference of the seasonal component by estimating the nonparametric trend function using splines. However, they did not take the time series structure of error terms into account. In their papers they assumed the error terms to be independent and identically distributed. Shao (2009) took correlation in the error terms into consideration and applied the partially linear models to analyzing time series using local linear smoothing for trends.

We will elaborate our proposed estimation method for PAR(p) model coefficient $\boldsymbol{\phi}$ when the time series contains a trend or trend as well as seasonal components. In particular, we will use B-splines to estimate the trend and calculate $\hat{\boldsymbol{\phi}}$ from the detrended sequence. Our method is computationally simple, but theoretically efficient under very general conditions.

Chapter 2

Nonparametric Efficient Estimation for Periodic Autoregressive Coefficients via Residuals

2.1 Introduction

The primary interest of this topic is inference for periodically correlated time series in the presence of a nonlinear trend. As in Section 1.4, If the mean of such a time series $\{x_t\}_{t=-\infty}^{\infty}$ is zero, it is usually described by the following periodic autoregressive model with order p (PAR(p))

$$x_{iT+\nu} - \sum_{k=1}^{p} \phi_k(\nu) x_{iT+\nu-k} = \sigma_\nu \epsilon_{iT+\nu}, \tag{2.1}$$

where $\{\epsilon_{iT+\nu}\}$ is white noise with $\mathrm{E}(\epsilon_{iT+\nu}) = 0$ and $\mathrm{Var}(\epsilon_{iT+\nu}) = 1$. To emphasize the periodicity, the time index t is often written as $iT + \nu$, where i is the number of cycle, T is the period, and ν ($1 \leq \nu \leq T$) is called the season and is the seasonal time-point in a period. The PAR(p) model coefficients $\boldsymbol{\phi}$ are given in Section 1.4. The unique feature of this model is that there are T autocorrelation functions that vary with season.

As we mentioned in Section 1.4, many results about inference for autoregressive coefficients have been established when the trend is estimated parametrically. In this chapter, we propose a two-step procedure to make inference for periodic autoregressive coefficients $\boldsymbol{\phi}$ from $\{y_t\}_{t=1}^n$: in the first step, the trend is estimated and the residuals are calculated; in the second step, the periodic autoregressive coefficients $\boldsymbol{\phi}$ are estimated from the residuals.

The essence of the proposed two-step method is to replace the unobservable time series $\{x_t\}_{t=1}^n$ by the residuals of the polynomial spline smoothing trend estimates. This procedure is appealing, as it is easy to implement. Practitioners can utilize any software package that has a build-in function for PAR. For example, in the section of simulation studies and data analysis of this paper, we will use "`pear` " (Mcleod and Balcilar (2011)), a package of R which is an open access environment for statistical computing and graphics developed by the R Core Team (2013). This package utilizes the Yule-Walker estimation method for periodic autoregressive coefficients of PAR(p) models for any given. When calculating the Yule-Walker estimates according to the method proposed in this paper by "`pear`", one simply needs to substitute the residuals for the time series in the package. However, an interesting and critical question is whether the estimator is asymptotically equivalent to the one based on $\{x_t\}_{t=1}^n$. The oracle efficiency of the proposed Yule-Walker estimator ensures that this replacement is appropriate. Shao and Yang (2011) was one of the first to consider such a two-step approach. They considered that observations contain trend plus autoregressive error terms, and the trend was estimated by polynomial spline smoothing. They concluded that under certain conditions, the difference between these two estimators is negligible. We will generalize their results to a periodic stationary time series setting and establish the oracle efficiency of the proposed estimator.

The chapter is organized as follows. In Section 2.2, we will consider how to estimate the trend by B-splines and how to obtain the residuals. In Section 2.3, we will focus on the Yule-Walker estimator of $\boldsymbol{\phi}$ from the residuals and present its asymptotic properties. In Section 2.4, we will provide the results of simulation studies for several PAR(1) models, and in Section 2.5 we will apply the procedure to the quarterly streamflows of the Fraser River at Hope, BC from 1913 to 2011. In Section 2.6, we will end the chapter with a brief summary. The proof of the theoretical result in Section 2.3 is given in Section 2.7.

2.2 Residuals of B-Spline Trend Estimation

As in Section 1.4, we consider a sequence of equally spaced points $\{(-m+1)h, (-m+2)h, \ldots, -h, 0, h, 2h, \ldots, Nh, 1\}$, where m is a positive integer. The interval $[0,1]$ is then divided into $N+1$ subintervals of equal length $h = (N+1)^{-1}$ as $J_j = [jh, (j+1)h), j = 0, 1, 2, \ldots, N-1$ and $J_N = [Nh, 1]$. Let $G_N^{(m-2)} = G_N^{(m-2)}[0,1]$ denote the space of polynomial functions of degree $m-1$ on each J_j and have continuous $(m-2)$-th derivatives. Then the B-spline basis of $G_N^{(m-2)}$ is $\{b_{j,m}(u), j = -m+1, \ldots, N\}$. We are going to discuss the same two cases as in Section 1.4: when $m = 1$, we have the space of functions $G_N^{(-1)}$, which is constant on each J_j; when $m = 2$, we have the space of functions $G_N^{(0)}$, which is linear on each J_j and continuous on $[0,1]$. For case $G_N^{(-1)}$, the B-spline basis is $\{b_{j,1}(u)\}_{j=0}^{N}$, where $b_{j,1}(u)$ is defined in equation (1.26). For case $G_N^{(0)}$, the B-spline basis is $\{b_{j,2}(u)\}_{j=-1}^{N}$, where $b_{j,2}(u)$ is defined in equations (1.27) and (1.28).

For observations $\{y_t\}_{t=1}^{n}$, we define that a vector $\boldsymbol{b}_j = (b_{j,m}(u_1), \ldots, b_{j,m}(u_n))'$ with $u_i = i/n$ and $n \times (N+m)$ matrix

$$\boldsymbol{B} = (\boldsymbol{b}_{-m+1}, \ldots, \boldsymbol{b}_N),$$

where $b_{j,m}(u_1), \ldots, b_{j,m}(u_n)$ are defined in equation (1.26) or (1.28), which depends on the value of m.

n_j is the number of observations in J_j, and $j_i = \sum_{k=0}^{j-1} n_k + i$. Thus x_{j_i} is the i-th observation in the j-th interval. Therefore, $\sum_{j=0}^{N} n_j = n$. The polynomial spline estimator of $g(u)$ for every $u \in [0,1]$ is

$$\hat{g}(u) = \sum_{j=1-m}^{N} b_{j,m}(u)\hat{a}_j, \tag{2.2}$$

where $b_{j,m}(u)$ is defined in equation (1.26) or (1.28). $\hat{\boldsymbol{a}} = (\hat{a}_{1-m}, \ldots, \hat{a}_N)'$ is calculated by

$$\hat{\boldsymbol{a}} = \underset{\boldsymbol{a}}{\operatorname{argmin}} (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{a}). \tag{2.3}$$

According to linear model theory, it is straightforward to obtain the trend estimate $\hat{\boldsymbol{g}} = (\hat{g}(u_1), \ldots, \hat{g}(u_n))'$ as follows from (2.2) and (2.3)

$$\hat{\boldsymbol{g}} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{y}. \tag{2.4}$$

The residual sequence $\hat{\boldsymbol{x}} = (\hat{x}_1, \ldots, \hat{x}_n)'$ is calculated by subtracting the estimated trend in (2.4) from the observation $\{y_t\}_{t=1}^{n}$.

$$\hat{\boldsymbol{x}} = \boldsymbol{y} - \hat{\boldsymbol{g}} = \{\boldsymbol{I} - \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\}\boldsymbol{y}.$$

We will replace $\{x_t\}_{t=1}^n$ by $\{\hat{x}_t\}_{t=1}^n$ in calculation of the Yule-Walker estimate of $\boldsymbol{\phi}$.

2.3 Periodic Autoregressive Coefficient Estimation

When the time series contains trend, we propose to estimate the periodic autoregressive coefficients by residuals of B-splines described in Section 2.2. We use $\hat{\mathbf{\Gamma}}$ and $\hat{\boldsymbol{\gamma}}$ to denote $\tilde{\mathbf{\Gamma}}$ and $\tilde{\boldsymbol{\gamma}}$ defined in equation (1.23) when $\{x_t\}_{t=1}^n$ is replaced by the residuals $\{\hat{x}_t\}_{t=1}^n$. The proposed Yule-Walker estimator of $\boldsymbol{\phi}$ is

$$\hat{\boldsymbol{\phi}} = \hat{\mathbf{\Gamma}}^{-1}\hat{\boldsymbol{\gamma}} \text{ and } \hat{\sigma}_\nu^2 = \hat{\gamma}_\nu(0) - \sum_{k=1}^{p} \hat{\phi}_k(\nu)\,\hat{\gamma}_{\nu-k}(-k). \tag{2.5}$$

The formulas of (1.23) and (2.5) are very similar except that the time series $\{x_t\}_{t=1}^n$ is replaced by the B-spline residuals $\{\hat{x}_t\}_{t=1}^n$. Hereafter, we will use hat and tilde to represent the formulas based on $\{\hat{x}_t\}$ and $\{x_t\}$, respectively, unless otherwise indicated. As we pointed out, $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ are different. However, according to the Theorem below $\hat{\boldsymbol{\phi}}$ is oracle or asymptotically equivalent to $\tilde{\boldsymbol{\phi}}$ under the following assumptions:

1. The trend function $g(\cdot) \in C^{(m)}[0,1]$, $m = 1, 2$; that is, the trend function has m continuous derivatives.

2. The subinterval length $h \sim n^{-1/(2m+1)}$; that is, the number of interior knots $N \sim n^{1/(2m+1)}$.

3. The time series $\{x_{iT+\nu}, 1 \le \nu \le T\}_{i=0}^{n_T-1}$ is causal; that is, for each fixed ν ($1 \le \nu \le T$), there exists a sequence of constants $\{\psi_\nu(j)\}_{j=0}^\infty$ such that $\sum_{j=0}^\infty |\psi_\nu(j)| < \infty$ and
$$x_{iT+\nu} = \sum_{j=0}^{\infty} \psi_\nu(j)\epsilon_{iT+\nu-j}.$$

4. $\mathrm{E}(\epsilon^4) < \infty$.

Remark 2.3.1. *Under Assumption 3,* $\sum_{|k|=0}^{\infty} |\gamma_\nu(k)| < \infty$ *for each season* ν*.*

Remark 2.3.2. *Pagano (1978) showed that under Assumptions 3 and 4,* $\sqrt{n_T}(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}) \overset{D}{\Longrightarrow} N(\mathbf{0}, \boldsymbol{\Sigma}^{-1})$ *and* $\tilde{\sigma}_\nu^2 \xrightarrow{P} \sigma_\nu^2$ *as* $n_T \to \infty$*, where* $\boldsymbol{\Sigma}$ *is the diagonal partitioned matrix with diag*$(\boldsymbol{\Sigma}) = (\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_T)$ *and the* (i,j)*-th entry of* $\boldsymbol{\Sigma}_\nu$ *is* $(\boldsymbol{\Sigma}_\nu)_{i,j} = \gamma_{\nu-i}(i-j)/\sigma_\nu^2$*.*

The primary conclusion of this chapter is that under the above assumptions, the difference between $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ are negligible, which is summarized in the following theorem:

Theorem 2.3.1. *Under Assumptions 1-4, the Yule-Walker estimator* $\hat{\boldsymbol{\phi}}$ *defined in (2.5) and* $\tilde{\boldsymbol{\phi}}$ *defined in (1.23) satisfy*

$$\hat{\boldsymbol{\phi}} - \tilde{\boldsymbol{\phi}} = o_p(n^{-1/2}),$$

and

$$\hat{\sigma}_\nu^2 - \tilde{\sigma}_\nu^2 = o_p(1).$$

We will prove Theorem 2.3.1 in the Section 2.7. Theorem 2.3.1 implies that $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ are equivalent in terms of efficiency; that is, both are $\sqrt{n}$-consistent estimators, which is given in the corollary below:

Corollary 2.3.1. *Under Assumptions 1-4,* $\sqrt{n_T}(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}) \overset{D}{\Longrightarrow} N(\mathbf{0}, \boldsymbol{\Sigma}^{-1})$ *and* $\hat{\sigma}_\nu^2 \xrightarrow{P} \sigma_\nu^2$ *as* $n_T \to \infty$*, where* $\boldsymbol{\Sigma}$ *are defined in Remark 2.3.2.*

We omit the proof of Corollary 2.3.1, as it is obvious according to Slusky's Theorem.

2.4 Simulation Studies

All the calculation is carried out in the R environment for statistical computing and graphics, using the R package "`pear` "for periodic autoregressive time series modeling. We simulate 100 sample paths of the time series in (1.25) with the error terms from eight different PAR(1) models with period $T = 4$ and the variances of periodic white noise $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.7$, $\sigma_3^2 = 0.85$, $\sigma_4^2 = 1$. The model coefficients $\boldsymbol{\phi} = (\phi(1), \phi(2), \phi(3), \phi(4))'$ of these eight different PAR(1) models are given in Table 2.1. We omit the subscripts of $\boldsymbol{\phi}$, as there is only one coefficient for each season of PAR(1). The coefficients in Table 2.1 are positive, or negative, or mixed, and their absolute values are from 0 to 2, so that these models represent a wide range of PAR(1) time series. It is known that the periodic stationarity condition for PAR(1) is

$$\left| \prod_{\nu=1}^{T} \phi(\nu) \right| < 1.$$

Unlike causal stationary AR(1) models, the coefficients of which are in the interval of $(-1, 1)$, PAR(1) can have coefficients larger than 1. Some models in the simulation studies are particularly chosen to illustrate the performance of $\hat{\boldsymbol{\phi}}$ when the absolute values of the coefficients are close to or even larger than 1, for example, Model 4 in Table 2.1. We use the same trend function as in Shao and Yang (2011),

$$g(u) = \sin(2\pi u), u \in [0, 1].$$

The simulations are conducted for sample sizes 200, 400, 800, 1600. We will provide the results for $n = 200, 400, 800, 1600$ or $n_T = 50, 100, 200, 400$ with $T = 4$ to illustrate the performance of the proposed estimators from relatively small to large sample sizes.

We estimate the trend by the linear B-spline where $m = 2$. The number of knots here as well as in the real data analysis in the next section is $N = [n^{1/5}]$ which is adopted from Shao and Yang (2011). To illustrate the effect of model misspecification, we also estimate the trend by linear regression. We obtain the estimates of PAR(1) model coefficients based on the residuals of B-spline and linear regression, respectively. In comparison with the estimates of different models, the sample means and sample standard deviations of the estimates of $\boldsymbol{\phi}$ are summarized in Tables 2.2-2.5. It is worth pointing out that when the absolute values of the true coefficients are close to or larger than one, the estimates tend to be smaller and unstable with relatively large standard deviations. The bias and variability decrease when the sample size becomes larger. In addition, as shown in these tables, in general the estimates from the B-spline residuals are not only closer to the true values, but they are more robust as well. When the trend model is misspecified, the results could be very misleading.

To demonstrate the performance of $\hat{\boldsymbol{\phi}}$, we also calculate $\tilde{\boldsymbol{\phi}}$ based on time series $\{x_t\}_{t=1}^{n}$ without trend and obtain the boxplots of the ratios $\{\tilde{\phi}(\nu)/\hat{\phi}(\nu)\}_{\nu=1}^{4}$ for the eight models. Figures 2-1–2-8 are the boxplots of the ratios of four seasons for the eight models. When the sample sizes are large, the shapes of the boxes look similar, which implies that $\hat{\boldsymbol{\phi}}$'s of the eight models perform not very differently. For each season, the boxes of the plots become narrower and narrower as well as closer and closer to 1, as sample sizes increases from 200 to 1600, which is in accordance with Theorem 2.3.1. However, $\hat{\boldsymbol{\phi}}$ tends to underestimate the coefficients.

2.5 Application

The time series we will use to demonstrate our proposed method is the quarterly streamflows of the Fraser River at Hope, B.C. from 1913 to 2011. The monthly streamflows of the Fraser River is one of the classic periodically stationary time series.

Table 2.1: PAR(1) Models for Simulation Studies

Model	True Value of $(\phi(1), \phi(2), \phi(3), \phi(4))$
1	(0.3, 0.6, 0.4, 0.2)
2	(0.2, 0.4, 0.6, 0.9)
3	(0.2, 0.9, 0.6, 0.9)
4	(0.2, 2, 1.5, 0.9)
5	$(-0.1, 0.2, -0.6, 0.4)$
6	$(0.6, -0.4, 0.2, -0.9)$
7	$(0.2, -2, -1.5, 0.9)$
8	$(-0.1, -0.2, -0.4, -0.6)$

Table 2.2: $n = 200$

M	E	Mean Estimate of $(\phi(1), \phi(2), \phi(3), \phi(4))\pm$ SD
1	B	$(0.284 \pm 0.106,\ 0.561 \pm 0.165,\ 0.378 \pm 0.158,\ 0.162 \pm 0.168)$
	L	$(0.411 \pm 0.147,\ 0.694 \pm 0.167,\ 0.496 \pm 0.165,\ 0.314 \pm 0.189)$
2	B	$(0.172 \pm 0.073,\ 0.366 \pm 0.148,\ 0.545 \pm 0.159,\ 0.862 \pm 0.135)$
	L	$(0.259 \pm 0.098,\ 0.554 \pm 0.228,\ 0.669 \pm 0.153,\ 0.904 \pm 0.135)$
3	B	$(0.175 \pm 0.070,\ 0.855 \pm 0.152,\ 0.568 \pm 0.115,\ 0.871 \pm 0.133)$
	L	$(0.266 \pm 0.100,\ 0.889 \pm 0.151,\ 0.671 \pm 0.146,\ 0.928 \pm 0.137)$
4	B	$(0.119 \pm 0.052,\ 1.303 \pm 0.287,\ 1.432 \pm 0.070,\ 0.881 \pm 0.053)$
	L	$(0.179 \pm 0.051,\ 1.603 \pm 0.471,\ 1.465 \pm 0.081,\ 0.898 \pm 0.045)$
5	B	$(-0.122 \pm 0.093,\ 0.179 \pm 0.175,\ -0.565 \pm 0.148,\ 0.373 \pm 0.129)$
	L	$(0.042 \pm 0.174,\ 0.424 \pm 0.263,\ -0.234 \pm 0.391,\ 0.512 \pm 0.187)$
6	B	$(0.576 \pm 0.081,\ -0.429 \pm 0.116,\ 0.208 \pm 0.120,\ -0.879 \pm 0.147)$
	L	$(0.616 \pm 0.072,\ -0.190 \pm 0.239,\ 0.326 \pm 0.192,\ -0.556 \pm 0.379)$
7	B	$(0.128 \pm 0.048,\ -1.393 \pm 0.421,\ -1.119 \pm 0.157,\ 0.882 \pm 0.047)$
	L	$(0.183 \pm 0.058,\ -1.276 \pm 0.784,\ -1.270 \pm 0.290,\ 0.887 \pm 0.050)$
8	B	$(-0.120 \pm 0.095,\ -0.226 \pm 0.166,\ -0.376 \pm 0.168,\ -0.616 \pm 0.152)$
	L	$(0.042 \pm 0.170,\ 0.121 \pm 0.361,\ -0.078 \pm 0.368,\ -0.318 \pm 0.316)$

Table 2.3: $n = 400$

M	E	Mean Estimate of $(\phi(1), \phi(2), \phi(3), \phi(4))\pm$ SD
1	B	$(0.287 \pm 0.078,\ 0.587 \pm 0.099,\ 0.387 \pm 0.098,\ 0.190 \pm 0.111)$
	L	$(0.399 \pm 0.122,\ 0.721 \pm 0.158,\ 0.508 \pm 0.147,\ 0.335 \pm 0.162)$
2	B	$(0.180 \pm 0.052,\ 0.376 \pm 0.097,\ 0.566 \pm 0.092,\ 0.877 \pm 0.084)$
	L	$(0.274 \pm 0.095,\ 0.528 \pm 0.169,\ 0.680 \pm 0.125,\ 0.896 \pm 0.092)$
3	B	$(0.178 \pm 0.051,\ 0.864 \pm 0.110,\ 0.578 \pm 0.095,\ 0.880 \pm 0.088)$
	L	$(0.266 \pm 0.090,\ 0.932 \pm 0.106,\ 0.668 \pm 0.115,\ 0.909 \pm 0.089)$
4	B	$(0.153 \pm 0.031,\ 1.470 \pm 0.221,\ 1.471 \pm 0.057,\ 0.888 \pm 0.035)$
	L	$(0.203 \pm 0.033,\ 1.718 \pm 0.319,\ 1.468 \pm 0.062,\ 0.898 \pm 0.035)$
5	B	$(-0.114 \pm 0.070,\ 0.215 \pm 0.105,\ -0.614 \pm 0.107,\ 0.386 \pm 0.109)$
	L	$(0.048 \pm 0.162,\ 0.458 \pm 0.278,\ -0.255 \pm 0.359,\ 0.488 \pm 0.126)$
6	B	$(0.587 \pm 0.056,\ -0.414 \pm 0.079,\ 0.205 \pm 0.092,\ -0.881 \pm 0.109)$
	L	$(0.630 \pm 0.063,\ -0.189 \pm 0.226,\ 0.342 \pm 0.176,\ -0.552 \pm 0.361)$
7	B	$(0.152 \pm 0.031,\ -1.660 \pm 0.215,\ -1.212 \pm 0.118,\ 0.888 \pm 0.034)$
	L	$(0.198 \pm 0.041,\ -1.363 \pm 0.673,\ -1.306 \pm 0.223,\ 0.896 \pm 0.032)$
8	B	$(-0.111 \pm 0.060,\ -0.215 \pm 0.121,\ -0.422 \pm 0.113,\ -0.580 \pm 0.098)$
	L	$(0.034 \pm 0.148,\ 0.141 \pm 0.355,\ -0.091 \pm 0.325,\ -0.328 \pm 0.293)$

Table 2.4: $n = 800$

M	E	Mean Estimate of $(\phi(1), \phi(2), \phi(3), \phi(4)) \pm$ SD
1	B	$(0.295 \pm 0.054,\ 0.593 \pm 0.075,\ 0.391 \pm 0.067,\ 0.197 \pm 0.069)$
	L	$(0.406 \pm 0.116,\ 0.699 \pm 0.121,\ 0.517 \pm 0.127,\ 0.321 \pm 0.139)$
2	B	$(0.192 \pm 0.041,\ 0.387 \pm 0.078,\ 0.587 \pm 0.076,\ 0.885 \pm 0.066)$
	L	$(0.270 \pm 0.079,\ 0.558 \pm 0.176,\ 0.669 \pm 0.102,\ 0.911 \pm 0.073)$
3	B	$(0.185 \pm 0.038,\ 0.882 \pm 0.075,\ 0.588 \pm 0.060,\ 0.890 \pm 0.065)$
	L	$(0.270 \pm 0.080,\ 0.923 \pm 0.067,\ 0.653 \pm 0.078,\ 0.911 \pm 0.065)$
4	B	$(0.179 \pm 0.021,\ 1.713 \pm 0.170,\ 1.488 \pm 0.032,\ 0.894 \pm 0.023)$
	L	$(0.208 \pm 0.026,\ 1.755 \pm 0.265,\ 1.467 \pm 0.051,\ 0.899 \pm 0.023)$
5	B	$(-0.092 \pm 0.038,\ 0.206 \pm 0.084,\ -0.596 \pm 0.083,\ 0.405 \pm 0.068)$
	L	$(0.058 \pm 0.166,\ 0.425 \pm 0.237,\ -0.261 \pm 0.348,\ 0.490 \pm 0.113)$
6	B	$(0.597 \pm 0.035,\ -0.405 \pm 0.061,\ 0.197 \pm 0.076,\ -0.908 \pm 0.063)$
	L	$(0.638 \pm 0.054,\ -0.181 \pm 0.228,\ 0.339 \pm 0.160,\ -0.559 \pm 0.350)$
7	B	$(0.178 \pm 0.020,\ -1.824 \pm 0.129,\ -1.347 \pm 0.088,\ 0.893 \pm 0.023)$
	L	$(0.207 \pm 0.023,\ -1.406 \pm 0.612,\ -1.338 \pm 0.180,\ 0.902 \pm 0.021)$
8	B	$(-0.105 \pm 0.040,\ -0.212 \pm 0.082,\ -0.395 \pm 0.065,\ -0.587 \pm 0.066)$
	L	$(0.046 \pm 0.153,\ 0.131 \pm 0.339,\ -0.113 \pm 0.296,\ -0.323 \pm 0.285)$

Table 2.5: $n = 1600$

M	E	Mean Estimate of $(\phi(1), \phi(2), \phi(3), \phi(4))\pm$ SD
1	B	$(0.296 \pm 0.036,\ 0.603 \pm 0.052,\ 0.399 \pm 0.049,\ 0.201 \pm 0.046)$
	L	$(0.411 \pm 0.115,\ 0.693 \pm 0.105,\ 0.511 \pm 0.119,\ 0.328 \pm 0.136)$
2	B	$(0.195 \pm 0.026,\ 0.397 \pm 0.050,\ 0.598 \pm 0.049,\ 0.895 \pm 0.046)$
	L	$(0.276 \pm 0.081,\ 0.547 \pm 0.153,\ 0.673 \pm 0.089,\ 0.915 \pm 0.052)$
3	B	$(0.193 \pm 0.023,\ 0.890 \pm 0.058,\ 0.597 \pm 0.043,\ 0.895 \pm 0.041)$
	L	$(0.271 \pm 0.076,\ 0.926 \pm 0.052,\ 0.661 \pm 0.073,\ 0.918 \pm 0.045)$
4	B	$(0.188 \pm 0.013,\ 1.817 \pm 0.100,\ 1.493 \pm 0.022,\ 0.896 \pm 0.018)$
	L	$(0.211 \pm 0.021,\ 1.781 \pm 0.226,\ 1.471 \pm 0.036,\ 0.901 \pm 0.016)$
5	B	$(-0.100 \pm 0.032,\ 0.198 \pm 0.051,\ -0.596 \pm 0.052,\ 0.396 \pm 0.040)$
	L	$(0.066 \pm 0.169,\ 0.420 \pm 0.225,\ -0.256 \pm 0.349,\ 0.492 \pm 0.101)$
6	B	$(0.601 \pm 0.026,\ -0.402 \pm 0.041,\ 0.198 \pm 0.050,\ -0.900 \pm 0.052)$
	L	$(0.643 \pm 0.052,\ -0.193 \pm 0.212,\ 0.344 \pm 0.152,\ -0.545 \pm 0.360)$
7	B	$(0.186 \pm 0.012,\ -1.914 \pm 0.069,\ -1.400 \pm 0.047,\ 0.899 \pm 0.016)$
	L	$(0.216 \pm 0.021,\ -1.414 \pm 0.593,\ -1.378 \pm 0.136,\ 0.899 \pm 0.016)$
8	B	$(-0.100 \pm 0.030,\ -0.201 \pm 0.058,\ -0.400 \pm 0.049,\ -0.608 \pm 0.047)$
	L	$(0.042 \pm 0.146,\ 0.139 \pm 0.345,\ -0.104 \pm 0.302,\ -0.326 \pm 0.279)$

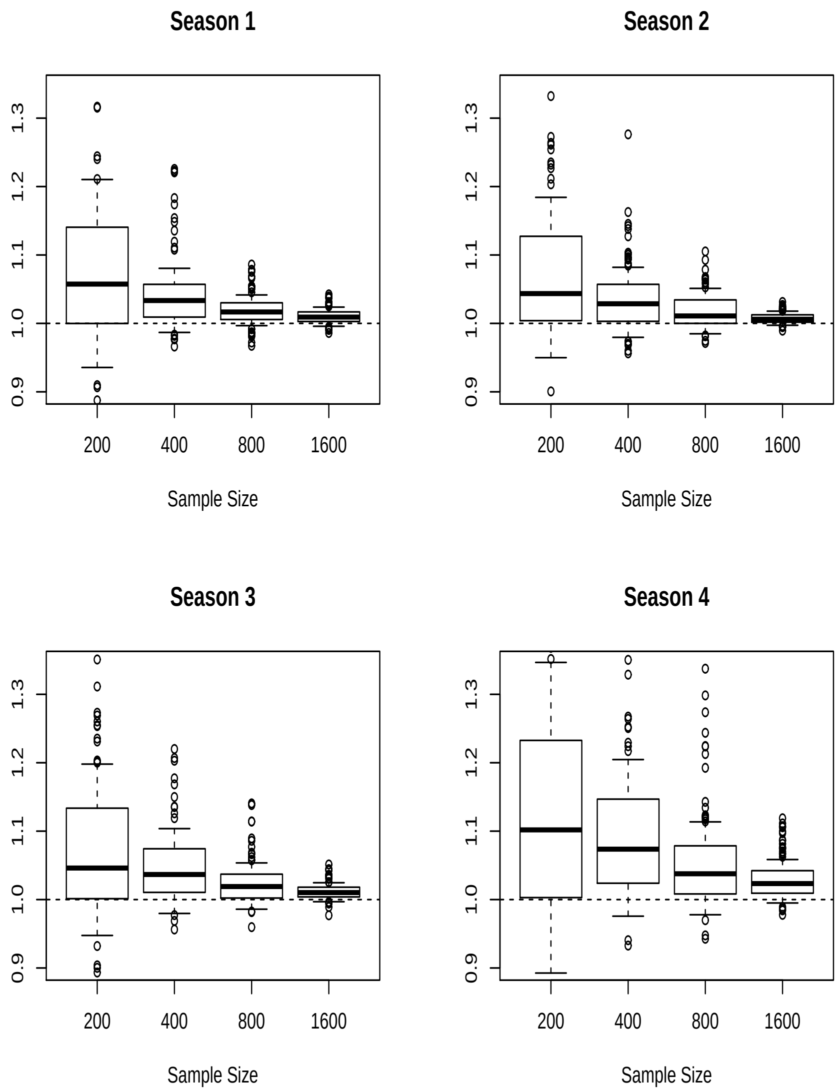

Figure 2-1: Model 1

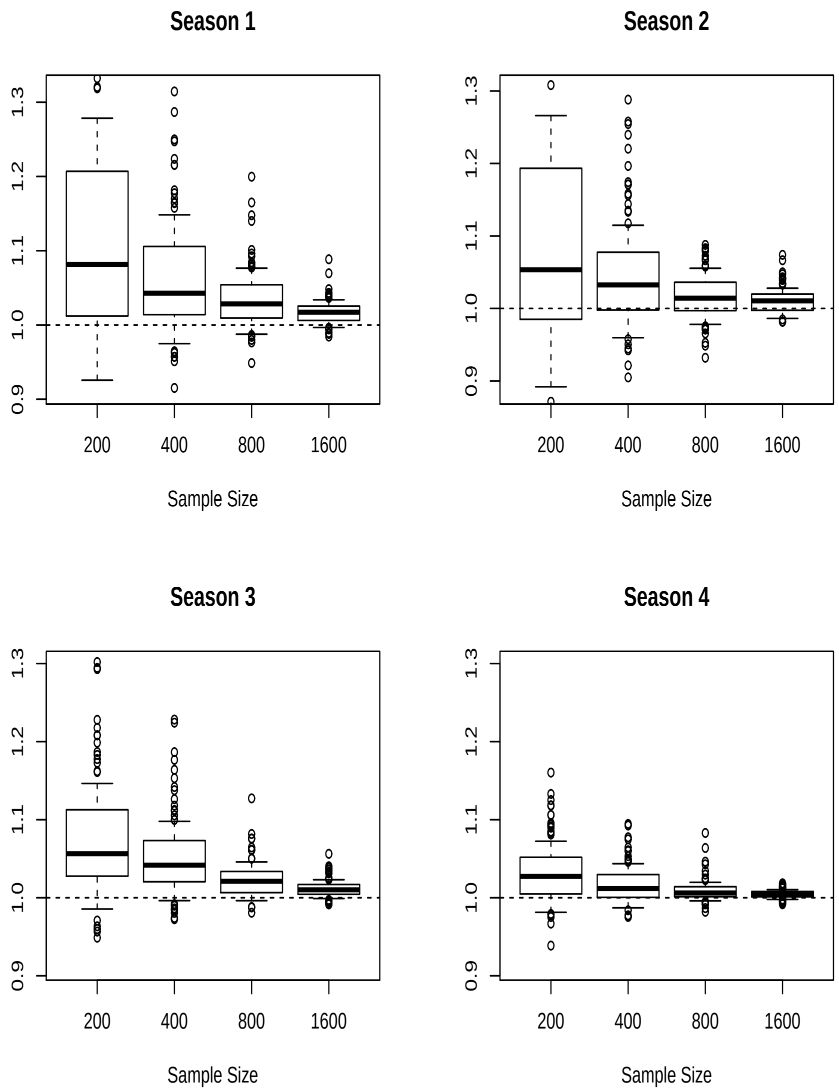

Figure 2-2: Model 2

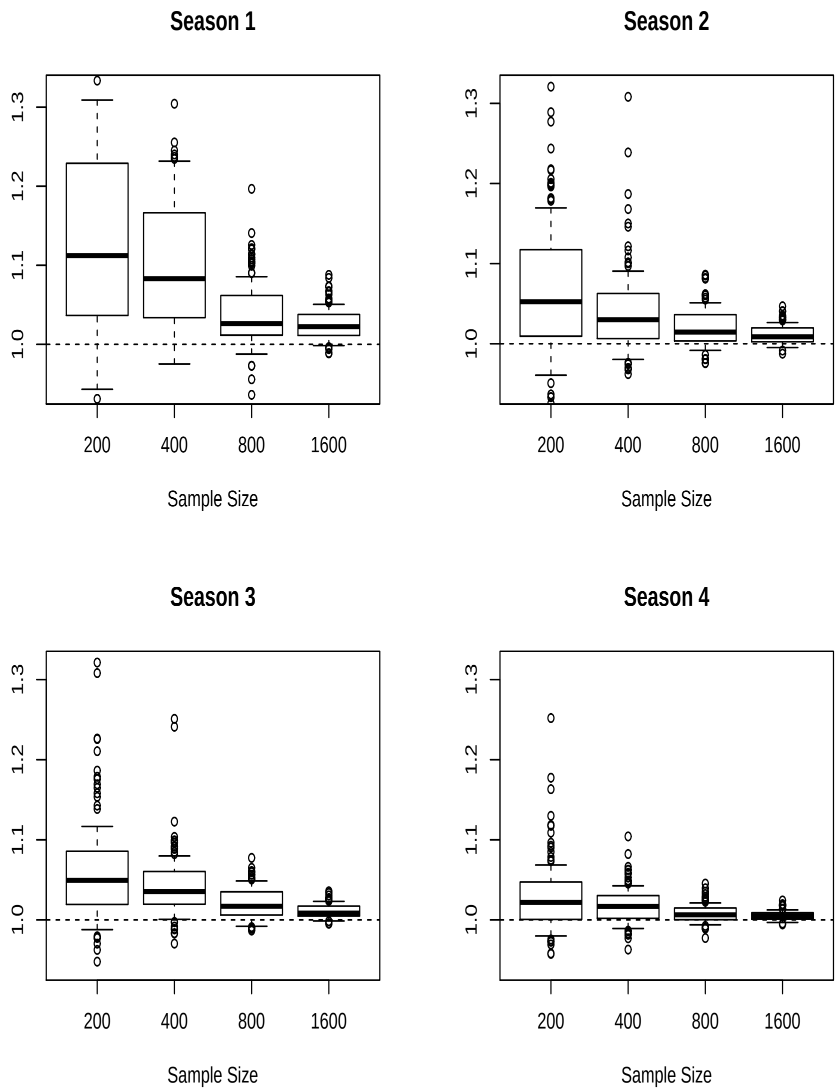

Figure 2-3: Model 3

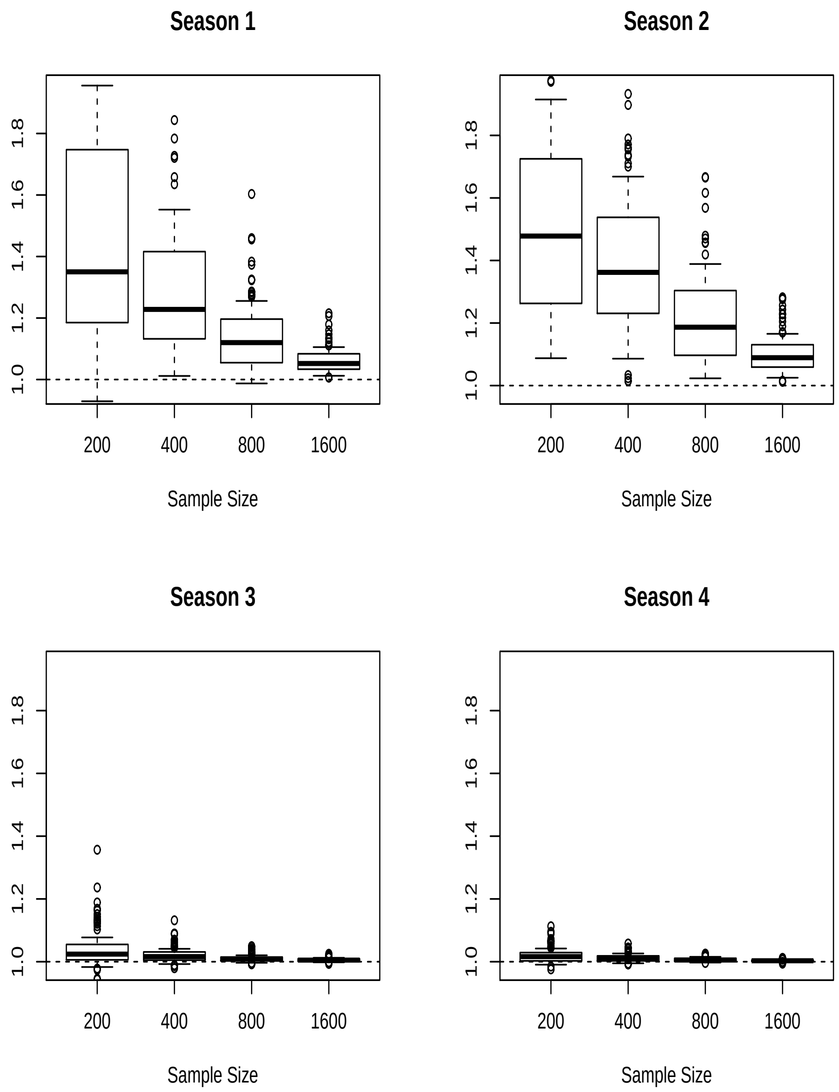

Figure 2-4: Model 4

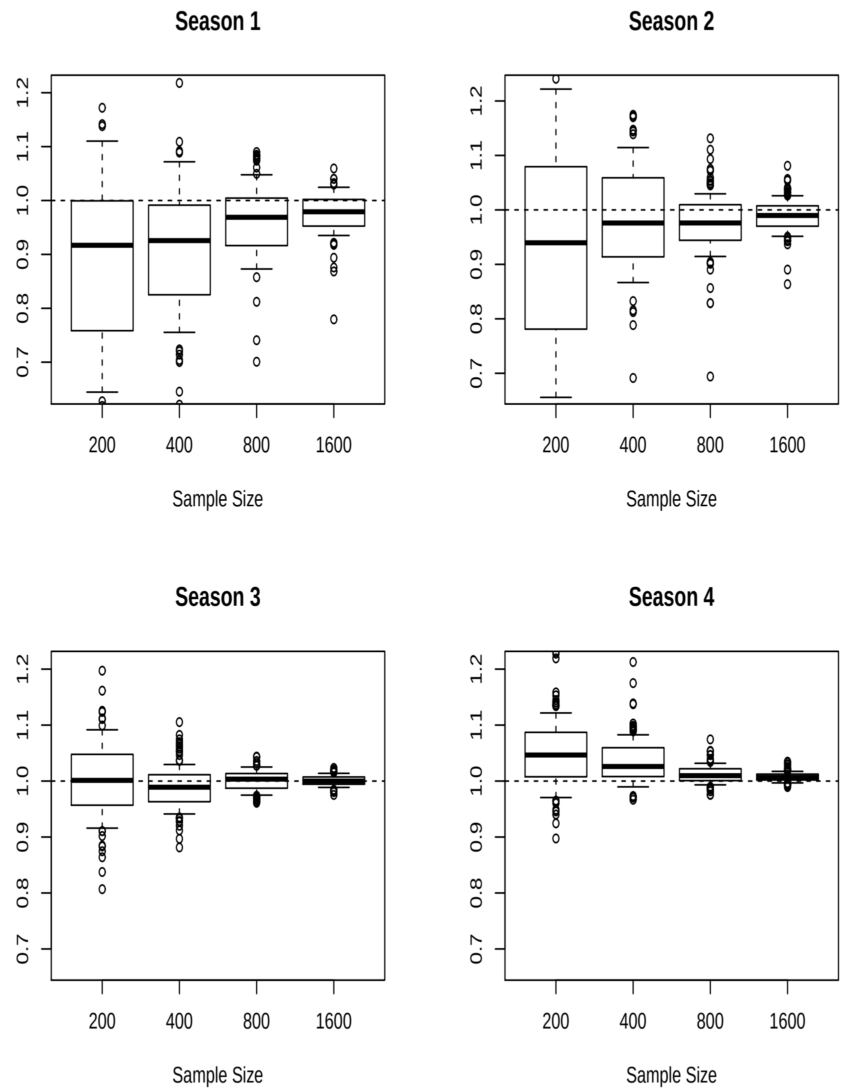

Figure 2-5: Model 5

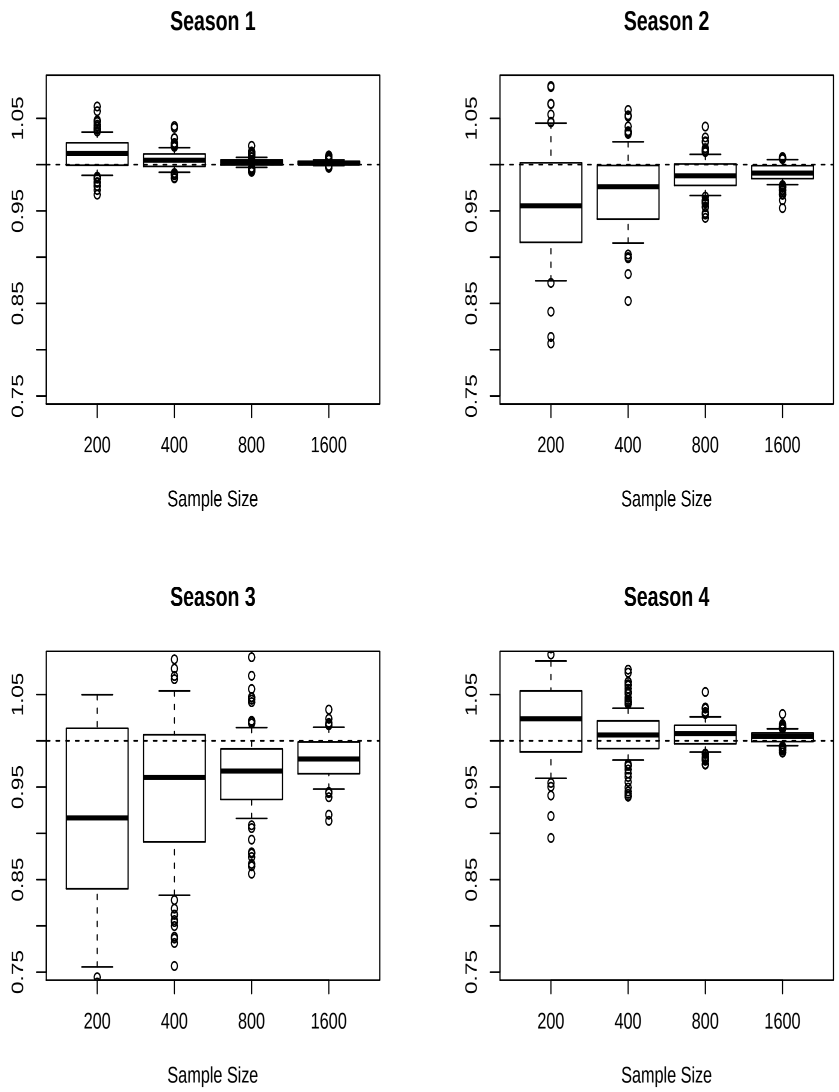

Figure 2-6: Model 6

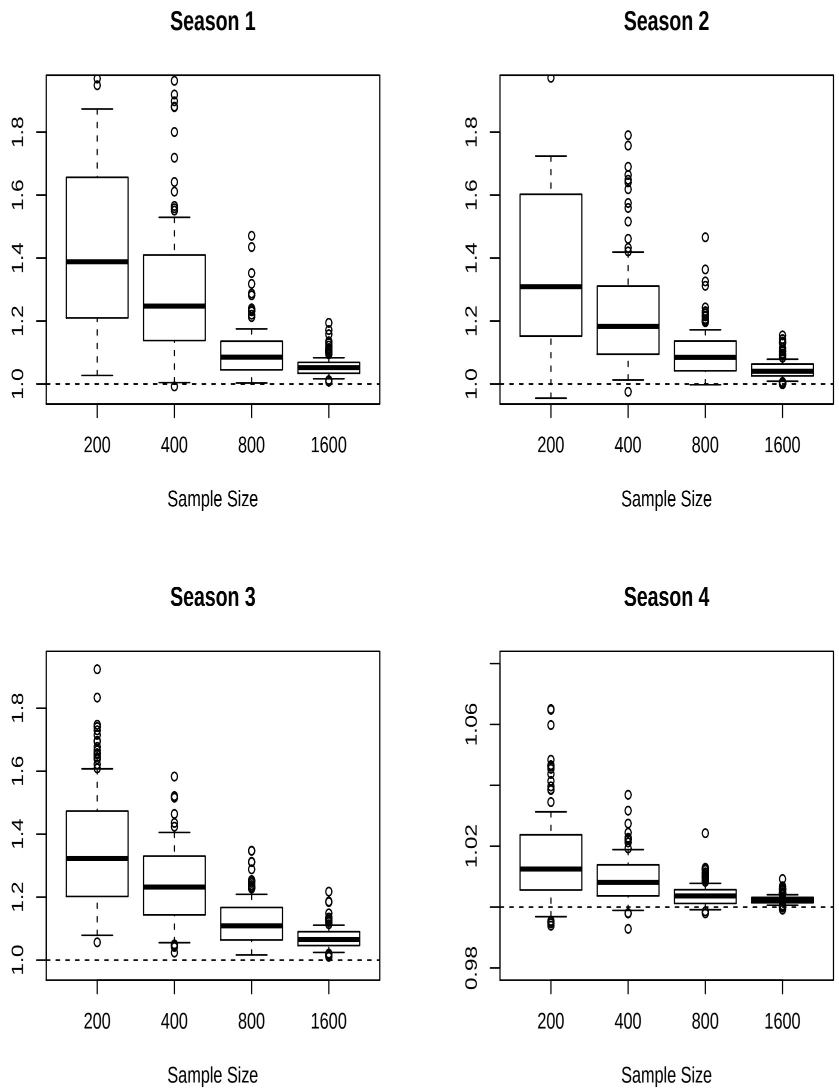

Figure 2-7: Model 7

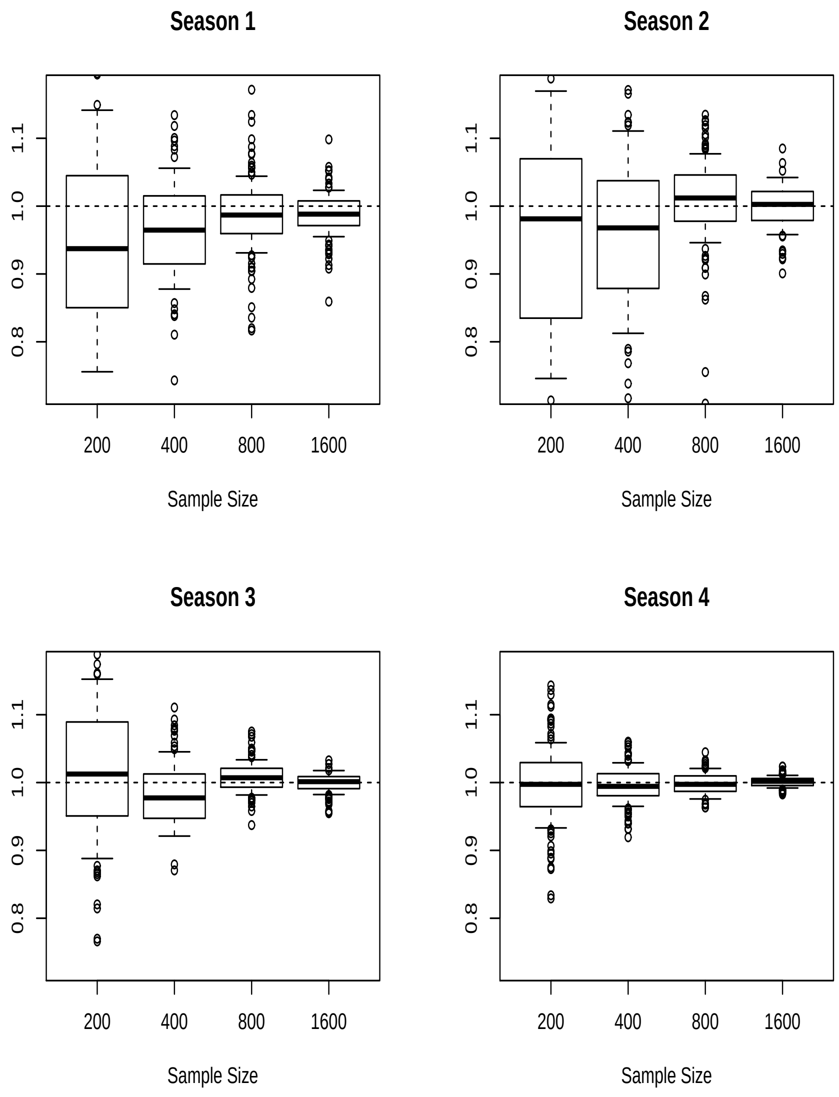

Figure 2-8: Model 8

The periodicity in the second moment was first noticed by Vecchia and Ballerini (1991). Since then this series has been adopted by many studies about periodic autoregressive moving-average with orders p and q (PARMA(p,q)) time series for the purpose of illustration. A PARMA(p,q) becomes a PAR(p) if $q = 0$. For example, McLeod (1994) discussed diagnostic checking of PARMA models with the monthly log transformed streamflows from 1913 to 1990; Tesfaye *et al* (2006) used the data from 1912 to 1984 to present how to identify PARMA models to capture the seasonal variations in river flow statistics; Anderson *et al* (2013) applied their forecasting procedures to the same time series. All these studies suggested that the streamflows exhibit periodicity in correlations which can be taken into account by PARMA models.

The observations under consideration are the quarterly data obtained from the monthly streamflows which can be downloaded at the website of the Canada of Environment. The data possibly contain not only a global trend function, but the seasonal means as well. The seasonal means could cause spikes which make the mean function not smooth enough to apply the proposed method. To eliminate the seasonality in the means, we subtract the quarterly data in 1913, which is chosen as benchmark, from the corresponding observations in 1914-2011. This removes the seasonal means, resulting in a smoother global trend function. Note that, unlike a seasonal AR model, subtracting the mean from a PAR model does not yield a stationary time series. The sample statistics in Table 2.6 exhibit non-stationary in the second moments. However, this preliminary analysis is based on the series which is possibly contaminated by the global trend.

We apply the proposed method to the quarterly data and estimate the global trend function by the linear B-spline. The trend estimates along with the differences of the quarterly streamflows are shown in Figure 2-9. The residuals are calculated by subtracting the trend estimates from the observations. These residuals play the role of the "observations" in the following analysis using "`pear`". We fit several

Table 2.6: Sample Means and Sample Standard Deviations of Quarterly Streamflows

	Season 1	Season 2	Season 3	Season 4
Mean	300	593	-1689	389
Standard Deviation	239	676	762	397

PAR(p) models, and PAR(1) is most appropriate according to the BIC criterion. Model adequacy checking is conducted using PAR(1) residual autocorrelations in Figure 2-10. The sample autocorrelation function at lags 0-25 shows no significant serial correlation. In other words, all the sample autocorrelations are within the 95% confidence interval, which indicates that the PAR(1) model is adequate. The coefficient $\boldsymbol{\phi}$ estimates with the standard errors from the residuals are $\hat{\phi}(1) = 0.305 \pm 0.052$, $\hat{\phi}(2) = 0.636 \pm 0.280$, $\hat{\phi}(3) = 0.429 \pm 0.105$, $\hat{\phi}(4) = 0.211 \pm 0.048$. We also modeled the residual sequence of the linear B-spline by AR(p) for the purpose of comparison. According to the BIC, the AR(20) is most appropriate, which is less parsimonious than PAR(1).

2.6 Concluding Remarks

In this chapter, we concentrate on time series with trend and periodic autoregressive random errors. To estimate the autoregressive coefficients of the unobservable error sequence, we propose a two-step method. The estimator we obtained by polynomial spline smoothing is $\sqrt{n}$-consistent estimators. It is equivalent to the Yule-Walker estimator from the observed periodic autoregressive time series in terms of efficiency. It is not only theoretically justified, but easy for practitioners to implement. The simulation studies show that the proposed method performs well. Compared with kernels and local linear method The polynomial spline smoothing has advantages of

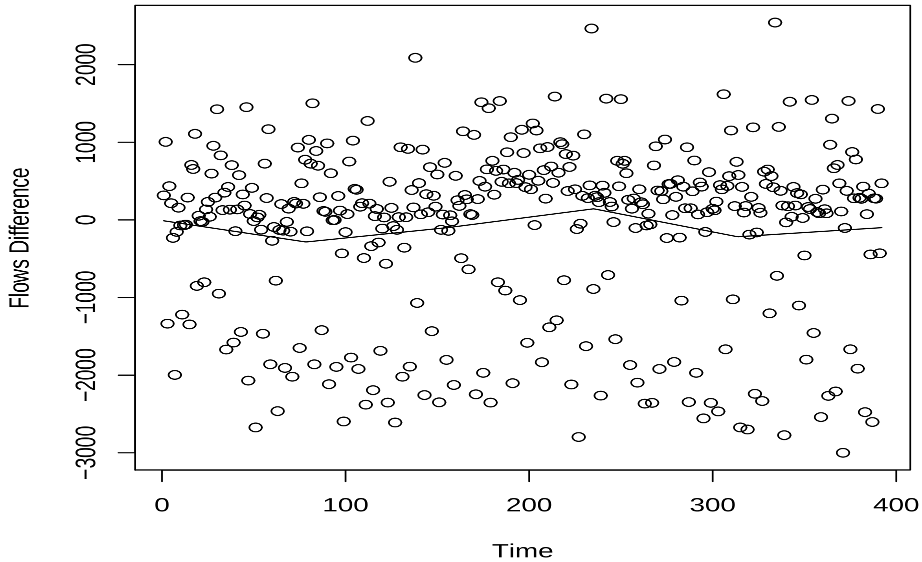

Figure 2-9: Differences of Quarterly Streamflows of Fraser River in m^3 with B-spline Trend Estimate

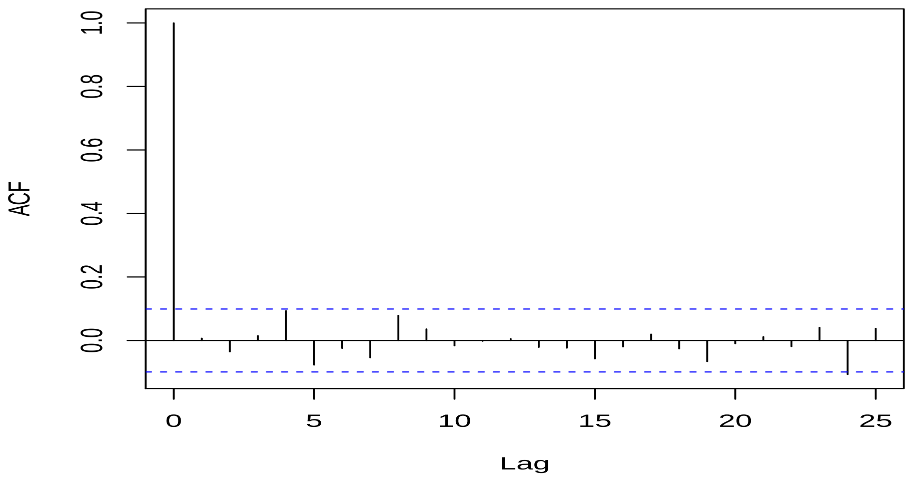

Figure 2-10: PAR(1) Residual Autocorrelations

computational expediency and mathematical simplicity. It is preferred when the data set is very large.

2.7 Proof

In this section, we will show Theorem 2.3.1 through several lemmas. Hereafter, $U(\cdot)$ and $u(\cdot)$ denote the uniform boundedness of a matrix and a scalar, respectively. Without loss of generality, we always assume that the number of observations is an multiple of the period T and the order of PAR is less than the period (i.e. $p < T$). Although the assumption is not necessary true, it will greatly simplify the notation without changing the asymptotic properties of a statistic under consideration.

According to (2.4), we have the following decomposition for the trend estimation:

$$\hat{\boldsymbol{g}} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\{\boldsymbol{g} + \boldsymbol{x}\} = \tilde{\boldsymbol{g}} + \tilde{\boldsymbol{x}}. \tag{2.6}$$

In (2.6), $\tilde{\boldsymbol{g}} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{g}$ and $\tilde{\boldsymbol{x}} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{x}$ are the projections of $\boldsymbol{g}$ and $\boldsymbol{x}$, respectively. Furthermore, we decompose the residuals $\hat{\boldsymbol{x}}$ into two components as follows:

$$\hat{\boldsymbol{x}} = \boldsymbol{y} - \hat{\boldsymbol{g}} = (\boldsymbol{g} - \tilde{\boldsymbol{g}}) + (\boldsymbol{x} - \tilde{\boldsymbol{x}}). \tag{2.7}$$

We will consider the asymptotics of the two terms in (2.7). First, under Assumptions 1 and 2, according to Theorem 5.1 of Huang (2003),

$$\|g - \tilde{g}\|_\infty = \sup_{u\in[0,1]} |g(u) - \tilde{g}(u)| = O(h^m). \tag{2.8}$$

where $\tilde{g}(u)$ is similarly defined by (2.2). The only difference is that we obtain $\hat{\boldsymbol{a}}$ using $\boldsymbol{g}$ in place of $\boldsymbol{y}$ in (2.3).

Before we discuss the second term in (2.7), we introduce the following notation,

$$\boldsymbol{x}_\nu = (x_\nu, x_{T+\nu}, \ldots, x_{(n_T-1)T+\nu})',$$
$$\boldsymbol{g}_\nu = (g(u_\nu), g(u_{T+\nu}), \ldots, g(u_{(n_T-1)T+\nu}))'.$$

This notation sometimes is applied to $\hat{\boldsymbol{x}}_\nu$, $\tilde{\boldsymbol{x}}_\nu$, $\tilde{\boldsymbol{g}}_\nu$ and so on. Notice that the vector $\boldsymbol{x}_\nu$ includes almost all but a finite number of the observations, and the vectors $\boldsymbol{x}_{iT+\nu}$ and $(x_\nu, x_{T+\nu}, \ldots, x_{(n_T-1)T+\nu})'$ are equivalent in the sense of asymptotics for any fixed i. With this notation, from (2.8), it is apparent that for any fixed integer $1 \le k \le p$,

$$\frac{1}{n}(\boldsymbol{g}_\nu - \tilde{\boldsymbol{g}}_\nu)'(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k}) = O(n^{-2m/(2m+1)}), \tag{2.9}$$
$$\frac{1}{n}(\boldsymbol{g}_\nu - \tilde{\boldsymbol{g}}_\nu)'\boldsymbol{x}_{\nu-k} = O_p(n^{-(4m+1)/(4m+2)}). \tag{2.10}$$

Lemma 2.7.1. *$\boldsymbol{B}'\boldsymbol{B} = U(nh)$, where $\boldsymbol{B} = (\boldsymbol{b}_{-1}, \ldots, \boldsymbol{b}_N)$ and $\{\boldsymbol{b}_j, j = -1, 0, \ldots, N\}$ is the basis of $G_N^{(0)}$.*

Proof. We have $u_{j_t} = u_{j_{(iT+\nu)}} = j_t/n = j_{(iT+\nu)}/n$, where $t = iT + \nu$. Then According to Lemma 2 of Shao and Yang (2011) and (1.28), for $0 \le j = k \le N-1$,

$$\begin{aligned}\|\boldsymbol{b}_j\|^2 &= \sum_{t=1}^{n} b_{j,2}^2(u_t) = \sum_{t=1}^{n_j} b_{j,2}^2(u_{j_t}) + \sum_{t=1}^{n_{j+1}} b_{j,2}^2(u_{(j+1)_t}) \\ &= \sum_{t=1}^{n_j}\left(\frac{j_t}{nh} - j\right)^2 + \sum_{t=1}^{n_{j+1}}\left(j + 2 - \frac{(j+1)_t}{nh}\right)^2.\end{aligned}$$

We have for $1 \le t \le n_j$

$$\frac{t-1}{nh} \le \frac{j_t}{nh} - j \le \frac{t}{nh}, \tag{2.11}$$

$$\frac{n_j - t}{nh} \le j + 1 - \frac{j_t}{nh} \le \frac{n_j - t + 1}{nh}, \tag{2.12}$$

$$\frac{n_{j+1} - t}{nh} \le j + 2 - \frac{(j+1)_t}{nh} \le \frac{n_{j+1} - t + 1}{nh}. \tag{2.13}$$

Thus

$$\sum_{t=1}^{n_j} \left(\frac{j_t}{nh} - j \right)^2 \leq \frac{1}{6n^2h^2} n_j(n_j+1)(2n_j+1).$$

Then we obtain $\|\boldsymbol{b}_j\|^2 = \frac{2}{3}nh + u(nh)$ for $0 \leq j \leq N-1$. For $j = -1$ and N, we can similarly prove that $\|\boldsymbol{b}_j\|^2 = \frac{1}{3}nh + u(nh)$. In addition, according to (2.11), (2.12) and (2.13), $\boldsymbol{b}_j{}'\boldsymbol{b}_{j+1} = \frac{1}{6}nh + u(nh)$, for $-1 \leq j \leq N-1$. Apparently, for $\|j-k| > 1$, $\boldsymbol{b}_j{}'\boldsymbol{b}_k = 0$. Therefore, we have $\boldsymbol{B}'\boldsymbol{B} = U(nh)$.

Lemma 2.7.2. *Under Assumptions 1-4, if $g(\cdot) \in C^{(m)}[0,1]$ is estimated using the basis of the space $G_N^{(m-2)}$ ($m = 1, 2$), then $\|\tilde{\boldsymbol{x}}\| = O_p(h^{-1/2})$.*

Proof. We will show the Lemma for the case $m = 1$. Make a partition of $\boldsymbol{B}$ to obtain $(j-i+1) \times (N+m)$ submatrix consisting of the elements from the ith row to the jth row of $\boldsymbol{B}$. Then we have the following conclusion about $\boldsymbol{b}_j$,

$$\max_{0 \leq j \leq N} |\|\boldsymbol{b}_j\|^2 - nh| = o(nh)$$

Notice that under Assumption 2, according to (1.26), the (j,k)th entry of $(\boldsymbol{B}'\boldsymbol{B})^{-1}$ satisfies

$$(\boldsymbol{B}'\boldsymbol{B})^{-1}_{j,k} = \begin{cases} u(n^{-1}h^{-1}), & j = k, \\ 0, & j \neq k. \end{cases} \tag{2.14}$$

Then $(\boldsymbol{B}'\boldsymbol{B})^{-1} = U(n^{-1}h^{-1})$.

Then, according to Assumption 3 and Remark 2.3.1 we have

$$
\begin{aligned}
\frac{1}{n}E\|\boldsymbol{B}'\boldsymbol{x}\|^2 &= \frac{1}{n}\sum_{j=0}^{N} E\left\{\sum_{i=1}^{n} b_{j,1}(u_i)x_i\right\}^2 \\
&= \frac{1}{n}\sum_{j=0}^{N} E(\sum_{i=1}^{n_j} x_{j_i})^2 \\
&= \frac{1}{n}\sum_{j=0}^{N} 2\left\{\sum_{i=0}^{\kappa_j-1}\sum_{\nu=1}^{T}\sum_{k=0}^{n_j-(iT+\nu)} \gamma_\nu(-k) + \sum_{\nu=0}^{u_j}\sum_{k=0}^{u_j-\nu}\gamma_\nu(-k)\right\} \\
&\quad - \frac{1}{n}\sum_{j=0}^{N}\left\{\kappa_j\sum_{\nu=1}^{T}\gamma_\nu(0) + \sum_{\nu=0}^{u_j}\gamma_\nu(0)\right\} \\
&= \frac{1}{n}\sum_{j=0}^{N}\left\{2\left(\sum_{i=0}^{\kappa_j-1}\sum_{\nu=1}^{T}\sum_{k=0}^{n_j-(iT+\nu)} \gamma_\nu(-k) + O(1)\right)\right\} \\
&\quad - \frac{1}{n}\sum_{j=0}^{N}\left\{\kappa_j\sum_{\nu=1}^{T}\gamma_\nu(0) + O(1)\right\} < \infty
\end{aligned}
$$

where we denote $\kappa_j T + u_j = n_j$, κ_j is the largest integer such that $\kappa_j T \leq n_j$, $0 \leq u_j \leq T-1$. n_j is the number of observations in J_j. In addition we define $\gamma_0(k) = 0$, for $-T \leq k \leq T$.

The final step is obtained by Remark 2.3.1; that is, for each ν $(1 \leq \nu \leq T)$

$$\sum_{|k|=0}^{\infty} |\gamma_\nu(k)| < \infty.$$

So, $\|\boldsymbol{B}'\boldsymbol{x}\| = O_p(n^{1/2})$ and thus

$$\|(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{x}\| = O_p(n^{-1/2}h^{-1}),$$

Then according to (2.6),

$$\|\tilde{\boldsymbol{x}}\|^2 \leq \|\boldsymbol{B}'\boldsymbol{x}\|\|(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{x}\| = O_p(h^{-1}).$$

and thus,

$$\|\tilde{\boldsymbol{x}}\| = O_p(h^{-1/2}).$$

Next, I will show the Lemma for $m = 2$ as follows. For any fixed t, we can find integers ν and i such that $t = iT + \nu$. Hence, for the t-th time point in the j-th subinterval of B-spline, $u_{j_t} = u_{j_{(iT+\nu)}} = j_t/n = j_{(iT+\nu)}/n$. Consider

$$\frac{1}{n}E\|\boldsymbol{B}'\boldsymbol{x}\|^2 = \frac{1}{n}E(\boldsymbol{b}_{-1}'\boldsymbol{x})^2 + \frac{1}{n}E(\boldsymbol{b}_N'\boldsymbol{x})^2 + R_{1n} + R_{2n} + R_{3n}, \tag{2.15}$$

where

$$\begin{aligned}
R_{1n} &= \frac{1}{n}\sum_{j=0}^{N-1} E\left\{\sum_{t=1}^{n_j}\left(\frac{j_t}{nh} - j\right)x_{j_t}\right\}^2, \\
R_{2n} &= \frac{1}{n}\sum_{j=0}^{N-1} E\left\{\sum_{t=1}^{n_{j+1}}\left(j + 2 - \frac{(j+1)_t}{nh}\right)x_{(j+1)_t}\right\}^2, \\
R_{3n} &= \frac{2}{n}\sum_{j=0}^{N-1} E\left\{\sum_{t=1}^{n_j}\sum_{k=1}^{n_{j+1}}\left(\frac{j_t}{nh} - j\right)\left(j + 2 - \frac{(j+1)_t}{nh}\right)x_{j_t}x_{(j+1)_t}\right\}.
\end{aligned}$$

The number of observations in the j-th interval is denoted by n_j and $\kappa_j T + u_j = n_j$, where the number of cycles in the interval κ_j is the largest integer such that $\kappa_j T \le n_j$ and the remainder u_j $(0 \le u_j \le T-1)$ is the number of observations left. We rewrite the term R_{1n} as follows:

$$\begin{aligned}
R_{1n} &= \frac{1}{n}\sum_{j=0}^{N-1} E\left\{\sum_{i=0}^{\kappa_j-1}\sum_{\nu=1}^{T}\left(\frac{iT+\nu}{nh}\right)x_{j_{(iT+\nu)}} + \sum_{\nu=0}^{u_j}\left(\frac{\kappa_j T+\nu}{nh}\right)x_{j_{(\kappa_j T+\nu)}}\right\}^2 \\
&= \frac{1}{n}\sum_{j=0}^{N-1} E\left\{\sum_{i=0}^{\kappa_j-1}\sum_{\nu=1}^{T}\left(\frac{iT+\nu}{nh}\right)x_{j_{(iT+\nu)}}\right\}^2 \\
&\quad +\frac{1}{n}\sum_{j=0}^{N-1} E\left\{\sum_{\nu=0}^{u_j}\left(\frac{\kappa_j T+\nu}{nh}\right)x_{j_{(\kappa_j T+\nu)}}\right\}^2 \\
&\quad +\frac{2}{n}\sum_{j=0}^{N-1} E\left\{\sum_{i=0}^{\kappa_j-1}\sum_{\nu=1}^{T}\sum_{\nu=0}^{u_j}\left(\frac{iT+\nu}{nh}\right)\left(\frac{\kappa_j T+\nu}{nh}\right)x_{j_{(iT+\nu)}}x_{j_{(\kappa_j T+\nu)}}\right\} \\
&= \frac{1}{n}\sum_{j=0}^{N-1}\frac{2}{n^2h^2}\left\{\sum_{i=0}^{\kappa_j-1}\sum_{\nu=1}^{T}\sum_{k=0}^{n_j-(iT+\nu)}(iT+\nu)(k+iT+\nu)\gamma_\nu(-k)\right\} \\
&\quad -\frac{1}{n}\sum_{j=0}^{N-1}\frac{1}{n^2h^2}\left\{\sum_{i=0}^{\kappa_j-1}(iT+\nu)^2\gamma_\nu(0)\right\}+o(n^{-1}) \\
&= O(1).
\end{aligned}$$

The final step is obtained by the causality of $\{x_t\}_{t=-\infty}^{\infty}$ in Remark 2.3.1; that is $\sum_{|k|=0}^{\infty}|\gamma_\nu(k)| < \infty$ for any ν $(1 \le \nu \le T)$. Similarly, we can show that the other terms in (2.15) are finite. Therefore, $\|\boldsymbol{B}'\boldsymbol{x}\| = O_p(n^{1/2})$.

According to Lemma 2.7.1 and Lemma 2 of Shao and Yang (2011), $(\boldsymbol{B}'\boldsymbol{B})^{-1} = U(n^{-1}h^{-1})$. Therefore,

$$\|(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{x}\| = O_p(n^{-1/2}h^{-1}),$$

which implies

$$\|\tilde{\boldsymbol{x}}\|^2 \le \|\boldsymbol{B}'\boldsymbol{x}\|\|(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{x}\| = O_p(h^{-1}).$$

The proof is complete.

Lemma 2.7.3. *Under the same assumption in Lemma 2.7.2, for any fixed* ν $\hat{\gamma}_\nu(k) \xrightarrow{p}$

$\gamma_\nu(k)$ *and* $\hat{\gamma}_\nu(k) - \tilde{\gamma}_\nu(k) = O_p(h^{2m})$.

Proof. Without loss of generality, we assume that $\nu - k \geq 0$. According to (2.7), we have

$$\begin{aligned}
\hat{\gamma}_\nu(k) - \tilde{\gamma}_\nu(k) &= \frac{1}{n_T}\hat{\boldsymbol{x}}'_\nu \hat{\boldsymbol{x}}_{\nu-k} - \frac{1}{n_T}\boldsymbol{x}'_\nu \boldsymbol{x}_{\nu-k} \\
&= \frac{1}{n_T}(\boldsymbol{g}_\nu - \tilde{\boldsymbol{g}}_\nu)'\boldsymbol{x}_{\nu-k} - \frac{1}{n_T}(\boldsymbol{g}_\nu - \tilde{\boldsymbol{g}}_\nu)'\tilde{\boldsymbol{x}}_{\nu-k} \\
&\quad -\frac{1}{n_T}\boldsymbol{x}'_{\nu-k}\tilde{\boldsymbol{x}}_\nu - \frac{1}{n_T}\tilde{\boldsymbol{x}}'_{\nu-k}\boldsymbol{x}_\nu + \frac{1}{n_T}\tilde{\boldsymbol{x}}'_\nu\tilde{\boldsymbol{x}}_{\nu-k}. \\
&\quad +\frac{1}{n_T}(\boldsymbol{g}_\nu - \tilde{\boldsymbol{g}}_\nu)'(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k}) \\
&\quad +\frac{1}{n_T}(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k})'\boldsymbol{x}_\nu \\
&\quad -\frac{1}{n_T}(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k})'\tilde{\boldsymbol{x}}_\nu. \qquad (2.16)
\end{aligned}$$

According to (2.8), (2.9), (2.10), and Lemma 2.7.2 we have

$$\begin{aligned}
&\frac{1}{n_T}(\boldsymbol{g}_\nu - \tilde{\boldsymbol{g}}_\nu)'(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k}) = O_p(n^{-2m/(2m+1)}), \\
&\frac{1}{n_T}(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k})'\boldsymbol{x}_\nu = O_p(n^{-2m/(2m+1)}), \\
&\frac{1}{n_T}(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k})'\tilde{\boldsymbol{x}}_\nu \leq \frac{1}{n_T}\|\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k}\|\|\tilde{\boldsymbol{x}}_\nu\| = O_p(n^{-2m/(2m+1)}), \\
&\frac{1}{n_T}\tilde{\boldsymbol{x}}'_\nu\tilde{\boldsymbol{x}}_{\nu-k} = O_p(n^{-2m/(2m+1)}), \\
&\frac{1}{n_T}\boldsymbol{x}'_{\nu-k}\tilde{\boldsymbol{x}}_\nu \leq \frac{1}{n_T}\|\boldsymbol{B}'\boldsymbol{x}\|\|(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{x}\| = O_p(n^{-2m/(2m+1)}).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
&\frac{1}{n_T}(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k})'\tilde{\boldsymbol{x}}_\nu = O_p(n^{-2m/(2m+1)}), \\
&\frac{1}{n_T}\boldsymbol{x}'_{\nu-k}\tilde{\boldsymbol{x}}_\nu = O_p(n^{-2m/(2m+1)}).
\end{aligned}$$

Therefore, the order of the dominant terms in (2.16) is $O_p(n^{-2m/(2m+1)})$, which implies

$\hat{\gamma}_\nu(k) - \tilde{\gamma}_\nu(k) = O_p(h^{2m})$ for any fixed $1 \le \nu \le T$. According to $\tilde{\gamma}_\nu(k) \xrightarrow{p} \gamma_\nu(k)$, we have $\hat{\gamma}_\nu(k) \xrightarrow{p} \gamma_\nu(k)$ for any fixed $1 \le \nu \le T$. The proof is complete.

Proof of Theorem 2.3.1 Notice the difference of Yule-Walker estimators $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ is

$$\hat{\boldsymbol{\phi}} - \tilde{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}^{-1}(\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}) + \hat{\boldsymbol{\Gamma}}^{-1}(\tilde{\boldsymbol{\Gamma}} - \hat{\boldsymbol{\Gamma}})\tilde{\boldsymbol{\Gamma}}^{-1}\tilde{\boldsymbol{\gamma}}.$$

According to Lemmas 2.7.1 and 2.7.2, $\hat{\boldsymbol{\phi}} - \tilde{\boldsymbol{\phi}} = o_p(n^{-1/2})$. Notice that for any fixed $1 \le \nu \le T$

$$\begin{aligned}
& \hat{\sigma}_\nu^2 - \tilde{\sigma}_\nu^2 \\
= \; & \left\{ \hat{\gamma}_\nu(0) - \sum_{k=1}^{p} \hat{\phi}_k(\nu)\, \hat{\gamma}_{\nu-k}(-k) \right\} - \left\{ \tilde{\gamma}_\nu(0) - \sum_{k=1}^{p} \tilde{\phi}_k(\nu)\, \tilde{\gamma}_{\nu-k}(-k) \right\} \\
= \; & \{\hat{\gamma}_\nu(0) - \tilde{\gamma}_\nu(0)\} - \left\{ \sum_{k=1}^{p} \hat{\phi}_k(\nu)\, \hat{\gamma}_{\nu-k}(-k) - \sum_{k=1}^{p} \tilde{\phi}_k(\nu)\, \tilde{\gamma}_{\nu-k}(-k) \right\} \\
= \; & o_p(1)
\end{aligned}$$

The proof is complete.

Chapter 3

Semiparametric Estimation for Periodic Autoregressive Coefficients in Partially Linear Models

3.1 Introduction

Time series of environmental sciences often contain trend, seasonality, and periodically correlated random components due to the seasonal and periodic nature of dynamical systems. The primary interest of this topic is inference for periodically correlated time series in the presence of both a nonlinear trend and seasonal effects. To emphasize such a cycling pattern, the time index t is often written as $iT+\nu$, where i is the number of cycle, and it is an integer, T is the period, and ν ($1 \leq \nu \leq T$) is called the season. Hipel and McLeod (1994) provided a comprehensive framework for modeling such time series. In particular, if the mean of such a time series $\{x_t\}_{t=1}^n$ is zero, the periodic correlated random terms $\{x_t\}_{t=1}^n$ are often well described by the periodic autoregressive model with order p (PAR(p)) defined in Section 1.4, which is as follows

$$x_{iT+\nu} - \sum_{k=1}^{p} \phi_k(\nu) x_{iT+\nu-k} = \sigma_\nu \epsilon_{iT+\nu}, \tag{3.1}$$

where $\{\epsilon_{iT+\nu}\}$ is white noise with $\mathrm{E}(\epsilon_{iT+\nu}) = 0$ and $\mathrm{Var}(\epsilon_{iT+\nu}) = 1$. A PAR($p$) becomes an autoregressive time series (AR(p)) when $T = 1$. Denote the model coefficients by $\boldsymbol{\phi} = (\boldsymbol{\phi}'(1), \cdots, \boldsymbol{\phi}'(T))'$ with $\boldsymbol{\phi}(\nu) = (\phi_1(\nu), \cdots, \phi_p(\nu))'$. The inference for $\boldsymbol{\phi}$ based on the observations $\{x_t\}_{t=1}^{n}$ or $\{x_{iT+\nu}, 1 \leq \nu \leq T\}_{i=0}^{n_T-1}$, where n is the number of observations and n_T is the number of periods, was considered by many researchers.

This topic is motivated by the analysis of the monthly global temperature data from January 1979 to December 2014 provided by the National Space Science and Technology Center. It seems from the time series plot in Figure 3-9 that the data set contains a trend and possibly seasonal means. A great deal of research has been done from the parametric approach when the trend $g(\cdot)$ follows a known analytic function with some unknown parameters and the error sequence $\{x_{iT+\nu}\}$ is time series, for example, in Anderson and Meerschaert (1998), Davis and Dunsmuir (1997), Lund, Shao, and Basawa (2006), and Tsay (1984). In particular, Anderson and Meerschaert (1998) took the seasonal-varying constant trend into account for periodic autoregressive error terms; Lund, Shao, and Basawa (2006) modeled the trend using a trigonometric function. Although parametric trend estimation works well when the shape of a trend can be approximated by a known analytical function, its major drawback is that the assumed model is usually subjective and is likely misspecified.

The temperature data in Figure 3-9, however, do not reveal that the trend follows the shape of any well-known analytic function. Moreover, given the seasonal nature of the observations, it is possible that the monthly means are not constant. The semiparametric approach we propose in this chapter aims at a more flexible modeling alternative which does not require a trend function to have an explicit format, but only a certain degree of smoothness. We generalize the partially linear models proposed by Engle, Granger, Rice, and Weiss (1986) to the following periodically correlated error terms:

$$y_{iT+\nu} = \sum_{k=1}^{T-1} I_k\left(\nu\right)\beta_k + g(u_{iT+\nu}) + x_{iT+\nu}, \tag{3.2}$$

where $u_t = t/n$; $g(\cdot)$ represents the unknown smooth trend function defined in the interval $[0, 1]$; We consider $g(\cdot)$ to be the smooth part of the (3.2) and assume that it represents a smooth unparameterized functional relationship. $I_k\left(\nu\right)$ is the indicator function for season ν with $I_k\left(\nu\right) = 1$ if $k = \nu$ and $I_k\left(\nu\right) = 0$ otherwise; $\{\beta_k\}_{k=1}^{T-1}$ are the seasonal effects compared with the reference level $\nu = T$; $\{x_t\}_{t=1}^{n}$ is a PAR(p) time series defined by (3.1). In (3.2) time series $\{y_t\}_{t=1}^{n}$ is assumed to be linearly related to one or more variables, but the relation to an additional variable or variables is not assumed to be easily parameterized. The model (3.2) of time series $\{y_t\}_{t=1}^{n}$ is more general than the one that only contains the trend function because it contains both trend function and seasonal effects. Therefore it can be used widely in practice.

Define an $n \times (T-1)$ matrix $\boldsymbol{D} = (\boldsymbol{D}_0, \boldsymbol{D}_0, \ldots, \boldsymbol{D}_0)'$, where a $(T-1) \times T$ matrix $\boldsymbol{D}_0 = (\boldsymbol{I}, \boldsymbol{0})$ with $\boldsymbol{I}$ being the $(T-1) \times (T-1)$ identity matrix and $\boldsymbol{0}$ being the $(T-1)$-dimensional zero vector $(0, 0, \ldots, 0)'$. The partially linear model (3.2) can be rewritten in vector format as

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{\beta} + \boldsymbol{g} + \boldsymbol{x}, \tag{3.3}$$

where $\boldsymbol{y} = (y_1, \ldots, y_n)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{T-1})'$, $\boldsymbol{g} = (g(u_1), g(u_2), \ldots, g(u_n))'$ with $u_i =$

i/n, and $\boldsymbol{x} = (x_1, \ldots, x_n)'$. According to $\boldsymbol{D}_0 = (\boldsymbol{I}, \boldsymbol{0})$, $\boldsymbol{D}_0'$ is as follows

$$\boldsymbol{D}_0' = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \end{pmatrix},$$

Our proposed semiparametric three-step approach not only can be used to estimate all the components in model (3.2), but can be applied to making inferences about them as well. In particular, we provide a theoretical justification for residual based Yule-Walker estimators for the time series model coefficients $\boldsymbol{\phi}$ by generalizing the results in Shao and Yang (2011) and Qiu, Shao, and Yang (2013) for the oracle efficiency of Yule-Walker estimators for autoregressive coefficients when time series observations contain a trend. It is straightforward to construct a 95% confidence interval for $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ based on Theorems 3.2.1-3.2.2, and a 95% confidence band for $g(\cdot)$ from the B-spline residuals using the method in Shao and Yang (2012) for AR(p). We will illustrate how to apply the procedure by analyzing three real time series data in Section 3.4.

The essence of the proposed three-step method is to replace the unobservable time series $\{x_t\}_{t=1}^n$ by the residuals of the B-spline estimates in partially linear models. The method is as follows: in the first step, the seasonal effects $\boldsymbol{\beta}$ is estimated based on B-splines; in the second step, the unknown trend $g(u)$ is estimated and the residuals are calculated by subtracting the estimated trend $\hat{g}(u)$ and estimated seasonal effects $\hat{\boldsymbol{\beta}}$ from the observations $\{y_t\}_{t=1}^n$; in the third step, the periodic autoregressive coefficients $\boldsymbol{\phi}$ are estimated from the residuals in the second step. We estimate the trend and seasonal effects $\boldsymbol{\beta}$ nonparametrically by B-spline method in partially linear models.

One of the advantages of the procedure is that it is very easy to implement. Practitioners can utilize any software package that has built-in functions for AR and PAR. For example, the simulation studies and data analysis in this chapter are accomplished using "`ar`" for autoregressive time series and "`pear`" (Mcleod and Balcilar (2011)) for periodic autoregressive time series. They are packages in R which is an open access environment for statistical computing and graphics developed by the R Core Team (2014). This package utilizes the Yule-Walker estimation method for periodic autoregressive coefficients of PAR(p) models. When calculating the Yule-Walker estimates according to the proposed method by "`pear` ", one simply needs to substitute the residuals for the time series in the package. However, an interesting and critical question is whether the estimator is asymptotically equivalent to the one based on $\{x_t\}_{t=1}^n$. The oracle efficiency of the proposed Yule-Walker estimator ensures that this replacement is appropriate. Shao and Yang (2011) were among the first to consider such an approach. They considered that observations contain trend plus autoregressive error terms, and the trend was estimated by polynomial spline smoothing. They concluded that under certain conditions, the difference between these two estimators is negligible. In this chapter we will generalize their results to a periodic correlated time series in partially linear models and establish the oracle efficiency of the proposed estimator.

The chapter is organized as follows. In Section 3.2, we will consider how to estimate the trend function $g\,(u)$ and seasonal effects $\boldsymbol{\beta}$ by B-splines smoothing in partially linear models, and how to calculate the Yule-Walker estimates for $\boldsymbol{\phi}$ from the residuals. In particular, we will provide the asymptotic properties of the estimators for the constant B-spline. In Section 3.3, we will illustrate implementation of the procedure and the performance of the estimators by simulation studies. In Section

3.4 we will apply the procedure to the monthly global temperature data from January 1979 to December 2014, the monthly global land-ocean temperature anomaly indices from January 1880 to December 2007, and the quarterly streamflows of the Fraser River at Hope, BC from 1913 to 2011. In Section 3.5, we will summarize this chapter and provide some concluding remarks. The proofs of the theoretical results in Section 3.2 are given in Section 3.6.

3.2 Construction of Yule-Walker Estimators

In this section, we will discuss the details of how to estimate $g\left(u\right)$ and $\boldsymbol{\beta}$ by B-spines, and how to calculate $\hat{\boldsymbol{\phi}}$ from residuals $\{\hat{x}_t\}_{t=1}^n$.

We consider a sequence of equally spaced points $\{(-m+1)h,\ (-m+2)h,\ \ldots\ ,$ $-h,\ 0,\ h,\ 2h,\ \ldots\ ,\ Nh,\ 1\}$, where m is a positive integer. The interval $[0,1]$ is divided into $N+1$ subintervals of equal length $h=(N+1)^{-1}$ as $J_j=[jh,(j+1)h), j=0,1,2,\ldots,N-1$ and $J_N=[Nh,1]$. We have the same subintervals as in Section 1.4. Then Let $G_N^{(m-2)}=G_N^{(m-2)}[0,1]$ denote the space of polynomial functions of degree $m-1$ on each J_j and let these polynomial functions have continuous $(m-2)$-th derivatives. Then the B-spline basis of $G_N^{(m-2)}$ is $\{b_{j,m}(u), j=-m+1,\ldots,N\}$.

We are going to discuss the same two cases as in Section 1.4: when $m=1$, we have the space of functions $G_N^{(-1)}$, which is constant on each J_j; when $m=2$, we have the space of functions $G_N^{(0)}$, which is linear on each J_j and continuous on $[0,1]$. For case $G_N^{(-1)}$, the B-spline basis is $\{b_{j,1}(u)\}_{j=0}^N$, where $b_{j,1}(u)$ is defined in equation (1.26). For case $G_N^{(0)}$, the B-spline basis is $\{b_{j,2}(u)\}_{j=-1}^N$, where $b_{j,2}(u)$ is defined in equations (1.27) and (1.28).

Then for observations $\{y_t\}_{t=1}^n$, we define a vector $\boldsymbol{b}_j = (b_{j,m}(u_1), \ldots, b_{j,m}(u_n))'$ with $u_i = i/n$ and $n \times (N+m)$ matrix

$$\boldsymbol{B} = (\boldsymbol{b}_{-m+1}, \ldots, \boldsymbol{b}_N),$$

where $b_{j,m}(u_1), \ldots, b_{j,m}(u_n)$ are defined in equation (1.26) or (1.28), which depends on the value of m.

n_j is the number of observations in J_j, and $j_i = \sum_{k=0}^{j-1} n_k + i$. Thus x_{j_i} is the i-th observation in the j-th interval. Therefore, $\sum_{j=0}^{N} n_j = n$. The spline smoother is

$$\boldsymbol{P_B} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'. \tag{3.4}$$

The seasonal effects $\boldsymbol{\beta}$ and the trend $g(\cdot)$ in the partially linear model (3.3) are respectively estimated by

$$\hat{\boldsymbol{\beta}} = \{\boldsymbol{D}'(\boldsymbol{I} - \boldsymbol{P_B})\boldsymbol{D}\}^{-1}\boldsymbol{D}'(\boldsymbol{I} - \boldsymbol{P_B})\boldsymbol{y}. \tag{3.5}$$

and

$$\hat{\boldsymbol{g}} = \boldsymbol{P_B}(\boldsymbol{y} - \boldsymbol{D}\hat{\boldsymbol{\beta}}). \tag{3.6}$$

The residual sequence $\hat{\boldsymbol{x}} = (\hat{x}_1, \ldots, \hat{x}_n)'$ is calculated by subtracting estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{g}}$ from the observations $\{y_t\}_{t=1}^n$:

$$\hat{\boldsymbol{x}} = \boldsymbol{y} - \hat{\boldsymbol{g}} - \boldsymbol{D}\hat{\boldsymbol{\beta}}. \tag{3.7}$$

We will replace $\{x_t\}_{t=1}^n$ by $\{\hat{x}_t\}_{t=1}^n$ in the calculation of the Yule-Walker estimate $\hat{\boldsymbol{\phi}}$.

The autocovariance function of a PAR(p) time series $\{x_t\}_{t=-\infty}^{\infty}$ with mean zero is defined by $\gamma_\nu(k) = \mathrm{E}(x_{iT+\nu}x_{iT+\nu-k})$, which is determined by both the lag k and the season ν. Without loss of generality, we only consider the case where $n_T = n/T$ is an integer, that is, there are n_T cycles of observations. When $\{x_t\}_{t=1}^n$ were observable, for such a time series with mean zero, we can obtain the sample autocovariance at lag k, that is defined in equation (1.21).

Let $\tilde{\mathbf{\Gamma}}_\nu$ be a $p \times p$ symmetric matrix with $1 \le \nu \le T$, the (i,j)-th entry $(\tilde{\mathbf{\Gamma}}_\nu)_{i,j}$ of which is described in equation (1.22).

Let $\tilde{\mathbf{\Gamma}}$ be a $pT \times pT$ diagonal partitioned matrix with $\mathrm{diag}\left(\tilde{\mathbf{\Gamma}}\right) = (\tilde{\mathbf{\Gamma}}_1, \cdots, \tilde{\mathbf{\Gamma}}_T)$, and a pT dimensional vector $\tilde{\boldsymbol{\gamma}} = (\tilde{\boldsymbol{\gamma}}_1', \tilde{\boldsymbol{\gamma}}_2', \ldots, \tilde{\boldsymbol{\gamma}}_T')'$ with $\tilde{\boldsymbol{\gamma}}_\nu = (\tilde{\gamma}_\nu(1), \tilde{\gamma}_\nu(2), \cdots, \tilde{\gamma}_\nu(p))'$. Then the Yule-Walker estimators from $\{x_t\}_{t=1}^n$ are defined in equation (1.23).

When the time series contains trend and seasonal effects, we propose to estimate the periodic autoregressive coefficients by residuals of the partial linear model described in equation (3.7). We use $\hat{\mathbf{\Gamma}}$ and $\hat{\boldsymbol{\gamma}}$ to denote $\tilde{\mathbf{\Gamma}}$ and $\tilde{\boldsymbol{\gamma}}$ when $\{x_t\}_{t=1}^n$ is replaced by the residuals $\{\hat{x}_t\}_{t=1}^n$ in equation (3.7). The proposed Yule-Walker estimators are

$$\hat{\boldsymbol{\phi}} = \hat{\mathbf{\Gamma}}^{-1}\hat{\boldsymbol{\gamma}} \text{ and } \hat{\sigma}_\nu^2 = \hat{\gamma}_\nu(0) - \sum_{k=1}^{p} \hat{\phi}_k(\nu)\,\hat{\gamma}_{\nu-k}(-k)\,. \tag{3.8}$$

Formulas (1.23) and (3.8) are very similar except that the time series $\{x_t\}_{t=1}^n$ is replaced by the residuals $\{\hat{x}_t\}_{t=1}^n$.

These estimators are not only computationally simple, but they have desirable asymptotic properties under very general conditions. For example, Theorem 3.2.2 below indicates that although these two estimators are different, $\hat{\boldsymbol{\phi}}$ is oracally or asymptotically equivalent to $\tilde{\boldsymbol{\phi}}$. We summarize these conditions and results below.

1. The trend function $g(\cdot) \in C^{(m)}[0,1]$, $m = 1$; that is, the trend function has m continuous derivatives. The first derivative of $g(\cdot)$ is finite on the interval $[0,1]$; that is, $g'(u) < \infty$ for every $u \in [0,1]$.

2. The subinterval length $h \sim n^{-1/(2m+1)}$; that is, the number of interior knots $N \sim n^{1/(2m+1)}$.

3. The time series $\{x_{iT+\nu}, 1 \leq \nu \leq T\}_{i=0}^{n_T-1}$ is causal; that is, for each fixed ν $(1 \leq \nu \leq T)$, there exists a sequence of constants $\{\psi_j(\nu)\}_{j=0}^{\infty}$ such that $\sum_{j=0}^{\infty} |\psi_j(\nu)| < \infty$ and
$$x_{iT+\nu} = \sum_{j=0}^{\infty} \psi_j(\nu) \epsilon_{iT+\nu-j}.$$

4. $\mathrm{E}(\epsilon^4) < \infty$.

These conditions are very typical either for periodically stationary time series or for B-spline estimator. For example, Assumption 3 ensures $\sum_{|k|=0}^{\infty} |\gamma_\nu(k)| < \infty$ for each season ν, which is necessary for the asymptotic normality of the Yule-Walker estimator of $\boldsymbol{\phi}$ from $\{x_t\}_{t=1}^{n}$. Although we anticipate the estimation procedure and the following theorems to hold for $m \geq 2$, we only provide here the results for $m = 1$ for simplicity.

Remark 3.2.1. *Under Assumption 3,* $\sum_{|k|=0}^{\infty} |\gamma_\nu(k)| < \infty$ *for each season* ν*.*

Under the above assumptions the primary conclusions of this chapter are the asymptotic property of $\hat{\boldsymbol{\beta}}$ and asymptotic equivalence of $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ in partially linear models, which are summarized in the following theorems:

Theorem 3.2.1. *Under Assumptions 1-4, as* $n_T \to \infty$*,*

$$\sqrt{n_T}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \stackrel{D}{\Longrightarrow} N\left(\mathbf{0}, \boldsymbol{V}\right), \tag{3.9}$$

where the (i,j)-th entry of the $(T-1)\times(T-1)$ matrix $\boldsymbol{V}$ is

$$\begin{aligned}(\boldsymbol{V})_{i,j} &= \sum_{k=-\infty}^{\infty}\{\gamma_i(kT+i-j)-\gamma_i(kT+T-j)\\ &\quad -\gamma_i(kT+i-T)+\gamma_i(kT)\}.\end{aligned}$$

Theorem 3.2.2. *Under Assumptions 1-4, the Yule-Walker estimator $\hat{\boldsymbol{\phi}}$ defined in (3.8) and $\tilde{\boldsymbol{\phi}}$ defined in (1.23) satisfy*

$$\hat{\boldsymbol{\phi}}-\tilde{\boldsymbol{\phi}} = o_p(n^{-1/2}),$$

and

$$\hat{\sigma}_\nu^2-\tilde{\sigma}_\nu^2 = o_p(1).$$

Under Assumptions 3 and 4, Pagano (1978) showed $\sqrt{n_T}(\tilde{\boldsymbol{\phi}}-\boldsymbol{\phi}) \overset{D}{\Longrightarrow} N(\mathbf{0},\boldsymbol{\Sigma}^{-1})$ and $\tilde{\sigma}_\nu^2 \xrightarrow{P} \sigma_\nu^2$ as $n_T\to\infty$, where $\boldsymbol{\Sigma}$ is the diagonal partitioned matrix with $\operatorname{diag}(\boldsymbol{\Sigma}) = (\boldsymbol{\Sigma}_1,\ldots,\boldsymbol{\Sigma}_T)$ and the (i,j)-th entry of $\boldsymbol{\Sigma}_\nu$ is $(\boldsymbol{\Sigma}_\nu)_{i,j} = \gamma_{\nu-i}(i-j)/\sigma_\nu^2$. Theorem 3.2.2 implies that $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ are equivalent in terms of efficiency and that both are $\sqrt{n}$-consistent estimators, which is given in the corollary below:

Corollary 3.2.1. *Under Assumptions 1-4, $\sqrt{n_T}(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}) \overset{D}{\Longrightarrow} N(\mathbf{0},\boldsymbol{\Sigma}^{-1})$ and $\hat{\sigma}_\nu^2 \xrightarrow{P} \sigma_\nu^2$ as $n_T\to\infty$.*

We omit the proof of Corollary 3.2.1, as it is obvious according to Slusky's Theorem, while the detailed proofs of Theorems 3.2.1-3.2.2 will be given in the Section 3.6.

3.3 Simulation Studies

In this section, we will conduct simulations to illustrate the performance of the proposed procedure for the linear B-spline with $m=2$ using PAR(1) time series.

All the statistical computing is carried out in the R environment and the R package "`pear`" is used for periodic autoregressive time series modeling. The sample sizes are $n = 200, 400, 800, 1600$ or $n_T = 50, 100, 200, 400$ with $T = 4$ so that the performance of the proposed estimators are illustrated for relatively small and large sample sizes. We simulate 100 sample paths of the time series in (3.2) with error terms from eight different PAR(1) models with period $T = 4$, $\boldsymbol{\beta} = (0.8, 1.5, 2.3)'$, and variances of periodic white noise $\sigma_1^2 = 0.5$, $\sigma_2^2 = 0.7$, $\sigma_3^2 = 0.85$, $\sigma_4^2 = 1$. The model coefficients $\boldsymbol{\phi} = (\phi(1), \phi(2), \phi(3), \phi(4))'$ of these eight different PAR(1) models are given in Table 3.1. These models are the same as in Table 2.1. We put it here to make this section more readable. We omit the subscripts of $\boldsymbol{\phi}$, as there is only one coefficient for each season of PAR(1). These models in Table 3.1 represent a wide range of PAR(1) time series and some models in the simulation studies are particularly chosen to illustrate the performance of $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\beta}}$ when the absolute values of the coefficients are close to or even larger than 1, for example, Model 4. For the trend function, we use the same one as in Section 2.4,

$$g(u) = \sin(2\pi u), u \in [0, 1].$$

We estimate the trend $g(\cdot)$ from (3.6), $\boldsymbol{\beta}$ from (3.5), and $\boldsymbol{\phi}$ from (3.8) nonparametrically by B-spline smoothing in partially linear models where $m = 2$. The number of knots here as well as in the real data analysis in the next section is $N = [n^{1/5}]$ which is adopted from Shao and Yang (2011). We obtain the estimates of PAR(1) model coefficients based on the residuals in partially linear models. For different models, the sample means and sample standard deviations of the estimates of $\boldsymbol{\phi}$ and $\boldsymbol{\beta}$ are summarized in Tables 3.2-3.5. It is worth pointing out that when the absolute values of the true coefficients $\boldsymbol{\phi}$ are close to or larger than one, $\hat{\boldsymbol{\phi}}$ tend to be smaller and unstable with relatively large standard deviations, but the estimates of $\boldsymbol{\beta}$ perform well for all of these models. As shown in these tables, the estimates are close to the

Table 3.1: PAR(1) Models for Simulation Studies

Model	True Value of $(\phi(1), \phi(2), \phi(3), \phi(4))$
1	(0.3, 0.6, 0.4, 0.2)
2	(0.2, 0.4, 0.6, 0.9)
3	(0.2, 0.9, 0.6, 0.9)
4	(0.2, 2, 1.5, 0.9)
5	$(-0.1, 0.2, -0.6, 0.4)$
6	$(0.6, -0.4, 0.2, -0.9)$
7	$(0.2, -2, -1.5, 0.9)$
8	$(-0.1, -0.2, -0.4, -0.6)$

true values and are robust as well. Overall, the bias and variability of both $\hat{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\beta}}$ decrease when the sample size becomes larger.

To demonstrate the performance of $\hat{\boldsymbol{\phi}}$, We also calculate $\tilde{\boldsymbol{\phi}}$ based on time series $\{x_t\}_{t=1}^n$ without trend and seasonality and ratios $\{\tilde{\phi}(\nu)/\hat{\phi}(\nu)\}_{\nu=1}^4$. Figures 3-1-3-8 are the box plots of the ratios of four seasons for the eight models. According to the box plots, we conclude that these boxes become narrower and narrower and closer and closer to 1 as sample sizes increases from 200 to 1600, which is in accordance with Theorem 3.2.2. We also notice that for Models 1-4 and Model 7, $\hat{\boldsymbol{\phi}}$ tends to underestimate the positive coefficients. For other models $\hat{\boldsymbol{\phi}}$ estimate the coefficients evenly.

3.4 Applications

We will demonstrate our proposed method by means of the following three time series, the monthly global temperature data from January 1979 to December 2014, the monthly global land-ocean temperature anomaly indices from January 1880 to December 2007, and the quarterly streamflows of the Fraser River at Hope, B.C. from

Table 3.2: $n = 200$

M	P	Mean Estimate of $(\phi(1), \phi(2), \phi(3), \phi(4))\pm$ SD
1	$\boldsymbol{\phi}$	$(0.282 \pm 0.086,\ 0.569 \pm 0.163,\ 0.361 \pm 0.166,\ 0.169 \pm 0.162)$
	$\boldsymbol{\beta}$	$(0.776 \pm 0.148,\ 1.458 \pm 0.175,\ 2.272 \pm 0.181)$
2	$\boldsymbol{\phi}$	$(0.168 \pm 0.071,\ 0.364 \pm 0.158,\ 0.572 \pm 0.176,\ 0.858 \pm 0.153)$
	$\boldsymbol{\beta}$	$(0.781 \pm 0.185,\ 1.477 \pm 0.186,\ 2.275 \pm 0.141)$
3	$\boldsymbol{\phi}$	$(0.165 \pm 0.076,\ 0.845 \pm 0.146,\ 0.552 \pm 0.130,\ 0.856 \pm 0.143)$
	$\boldsymbol{\beta}$	$(0.776 \pm 0.188,\ 1.485 \pm 0.203,\ 2.284 \pm 0.153)$
4	$\boldsymbol{\phi}$	$(0.119 \pm 0.052,\ 1.292 \pm 0.326,\ 1.439 \pm 0.089,\ 0.880 \pm 0.056)$
	$\boldsymbol{\beta}$	$(0.845 \pm 0.591,\ 1.527 \pm 0.316,\ 2.290 \pm 0.150)$
5	$\boldsymbol{\phi}$	$(-0.119 \pm 0.088,\ 0.222 \pm 0.171,\ -0.560 \pm 0.140,\ 0.373 \pm 0.143)$
	$\boldsymbol{\beta}$	$(0.769 \pm 0.185,\ 1.482 \pm 0.220,\ 2.288 \pm 0.166)$
6	$\boldsymbol{\phi}$	$(0.582 \pm 0.078,\ -0.412 \pm 0.126,\ 0.218 \pm 0.146,\ -0.886 \pm 0.156)$
	$\boldsymbol{\beta}$	$(0.773 \pm 0.143,\ 1.470 \pm 0.287,\ 2.280 \pm 0.321)$
7	$\boldsymbol{\phi}$	$(0.121 \pm 0.044,\ -1.360 \pm 0.382,\ -1.110 \pm 0.154,\ 0.878 \pm 0.048)$
	$\boldsymbol{\beta}$	$(0.826 \pm 0.546,\ 1.554 \pm 1.269,\ 2.276 \pm 0.150)$
8	$\boldsymbol{\phi}$	$(-0.122 \pm 0.101,\ -0.218 \pm 0.181,\ -0.424 \pm 0.157,\ -0.626 \pm 0.133)$
	$\boldsymbol{\beta}$	$(0.750 \pm 0.188,\ 1.480 \pm 0.173,\ 2.230 \pm 0.253)$

Table 3.3: $n = 400$

M	P	Mean Estimate of $(\phi(1), \phi(2), \phi(3), \phi(4))\pm$ SD
1	$\boldsymbol{\phi}$	$(0.285 \pm 0.079,\ 0.584 \pm 0.093,\ 0.379 \pm 0.092,\ 0.180 \pm 0.090)$
	$\boldsymbol{\beta}$	$(0.790 \pm 0.104,\ 1.475 \pm 0.123,\ 2.283 \pm 0.135)$
2	$\boldsymbol{\phi}$	$(0.188 \pm 0.055,\ 0.385 \pm 0.114,\ 0.582 \pm 0.093,\ 0.873 \pm 0.082)$
	$\boldsymbol{\beta}$	$(0.789 \pm 0.136,\ 1.487 \pm 0.138,\ 2.292 \pm 0.108)$
3	$\boldsymbol{\phi}$	$(0.177 \pm 0.058,\ 0.862 \pm 0.108,\ 0.578 \pm 0.085,\ 0.874 \pm 0.092)$
	$\boldsymbol{\beta}$	$(0.816 \pm 0.132,\ 1.492 \pm 0.116,\ 2.295 \pm 0.093)$
4	$\boldsymbol{\phi}$	$(0.152 \pm 0.037,\ 1.480 \pm 0.248,\ 1.470 \pm 0.050,\ 0.882 \pm 0.039)$
	$\boldsymbol{\beta}$	$(0.815 \pm 0.393,\ 1.518 \pm 0.217,\ 2.309 \pm 0.100)$
5	$\boldsymbol{\phi}$	$(-0.109 \pm 0.069,\ 0.210 \pm 0.128,\ -0.590 \pm 0.113,\ 0.386 \pm 0.110)$
	$\boldsymbol{\beta}$	$(0.785 \pm 0.132,\ 1.483 \pm 0.143,\ 2.291 \pm 0.118)$
6	$\boldsymbol{\phi}$	$(0.588 \pm 0.056,\ -0.407 \pm 0.074,\ 0.190 \pm 0.100,\ -0.893 \pm 0.103)$
	$\boldsymbol{\beta}$	$(0.787 \pm 0.098,\ 1.483 \pm 0.216,\ 2.282 \pm 0.230)$
7	$\boldsymbol{\phi}$	$(0.153 \pm 0.027,\ -1.677 \pm 0.221,\ -1.218 \pm 0.127,\ 0.889 \pm 0.033)$
	$\boldsymbol{\beta}$	$(0.784 \pm 0.390,\ 1.468 \pm 0.934,\ 2.318 \pm 0.093)$
8	$\boldsymbol{\phi}$	$(-0.106 \pm 0.056,\ -0.210 \pm 0.115,\ -0.411 \pm 0.110,\ -0.612 \pm 0.100)$
	$\boldsymbol{\beta}$	$(0.783 \pm 0.138,\ 1.485 \pm 0.116,\ 2.285 \pm 0.185)$

Table 3.4: $n = 800$

M	P	Mean Estimate of $(\phi(1), \phi(2), \phi(3), \phi(4))\pm$ SD
1	$\boldsymbol{\phi}$	$(0.293 \pm 0.045,\ 0.594 \pm 0.082,\ 0.390 \pm 0.078,\ 0.184 \pm 0.075)$
	$\boldsymbol{\beta}$	$(0.796 \pm 0.073,\ 1.495 \pm 0.087,\ 2.291 \pm 0.088)$
2	$\boldsymbol{\phi}$	$(0.190 \pm 0.039,\ 0.388 \pm 0.078,\ 0.588 \pm 0.074,\ 0.891 \pm 0.067)$
	$\boldsymbol{\beta}$	$(0.805 \pm 0.095,\ 1.508 \pm 0.096,\ 2.302 \pm 0.072)$
3	$\boldsymbol{\phi}$	$(0.188 \pm 0.034,\ 0.878 \pm 0.076,\ 0.589 \pm 0.055,\ 0.894 \pm 0.059)$
	$\boldsymbol{\beta}$	$(0.795 \pm 0.100,\ 1.499 \pm 0.097,\ 2.296 \pm 0.073)$
4	$\boldsymbol{\phi}$	$(0.179 \pm 0.020,\ 1.690 \pm 0.175,\ 1.490 \pm 0.033,\ 0.895 \pm 0.023)$
	$\boldsymbol{\beta}$	$(0.793 \pm 0.290,\ 1.489 \pm 0.152,\ 2.293 \pm 0.064)$
5	$\boldsymbol{\phi}$	$(-0.104 \pm 0.042,\ 0.203 \pm 0.094,\ -0.610 \pm 0.072,\ 0.401 \pm 0.067)$
	$\boldsymbol{\beta}$	$(0.789 \pm 0.105,\ 1.485 \pm 0.112,\ 2.308 \pm 0.086)$
6	$\boldsymbol{\phi}$	$(0.596 \pm 0.041,\ -0.396 \pm 0.058,\ 0.205 \pm 0.066,\ -0.897 \pm 0.069)$
	$\boldsymbol{\beta}$	$(0.790 \pm 0.061,\ 1.491 \pm 0.138,\ 2.286 \pm 0.151)$
7	$\boldsymbol{\phi}$	$(0.176 \pm 0.018,\ -1.818 \pm 0.123,\ -1.331 \pm 0.092,\ 0.897 \pm 0.023)$
	$\boldsymbol{\beta}$	$(0.790 \pm 0.273,\ 1.482 \pm 0.645,\ 2.305 \pm 0.076)$
8	$\boldsymbol{\phi}$	$(-0.106 \pm 0.039,\ -0.207 \pm 0.079,\ -0.392 \pm 0.068,\ -0.608 \pm 0.073)$
	$\boldsymbol{\beta}$	$(0.809 \pm 0.100,\ 1.512 \pm 0.100,\ 2.306 \pm 0.126)$

Table 3.5: $n = 1600$

M	P	Mean Estimate of $(\phi(1), \phi(2), \phi(3), \phi(4))\pm$ SD
1	$\boldsymbol{\phi}$	$(0.295 \pm 0.035,\ 0.597 \pm 0.056,\ 0.400 \pm 0.052,\ 0.198 \pm 0.047)$
	$\boldsymbol{\beta}$	$(0.797 \pm 0.051,\ 1.498 \pm 0.064,\ 2.300 \pm 0.058)$
2	$\boldsymbol{\phi}$	$(0.198 \pm 0.024,\ 0.393 \pm 0.056,\ 0.605 \pm 0.052,\ 0.897 \pm 0.043)$
	$\boldsymbol{\beta}$	$(0.801 \pm 0.061,\ 1.502 \pm 0.065,\ 2.302 \pm 0.046)$
3	$\boldsymbol{\phi}$	$(0.196 \pm 0.022,\ 0.883 \pm 0.059,\ 0.596 \pm 0.039,\ 0.902 \pm 0.048)$
	$\boldsymbol{\beta}$	$(0.796 \pm 0.064,\ 1.499 \pm 0.066,\ 2.301 \pm 0.048)$
4	$\boldsymbol{\phi}$	$(0.190 \pm 0.013,\ 1.805 \pm 0.106,\ 1.495 \pm 0.021,\ 0.897 \pm 0.015)$
	$\boldsymbol{\beta}$	$(0.801 \pm 0.208,\ 1.497 \pm 0.107,\ 2.296 \pm 0.052)$
5	$\boldsymbol{\phi}$	$(-0.102 \pm 0.029,\ 0.200 \pm 0.063,\ -0.595 \pm 0.052,\ 0.400 \pm 0.049)$
	$\boldsymbol{\beta}$	$(0.798 \pm 0.069,\ 1.503 \pm 0.071,\ 2.300 \pm 0.060)$
6	$\boldsymbol{\phi}$	$(0.597 \pm 0.023,\ -0.401 \pm 0.040,\ 0.203 \pm 0.050,\ -0.900 \pm 0.053)$
	$\boldsymbol{\beta}$	$(0.801 \pm 0.049,\ 1.500 \pm 0.096,\ 2.303 \pm 0.108)$
7	$\boldsymbol{\phi}$	$(0.186 \pm 0.012,\ -1.900 \pm 0.085,\ -1.389 \pm 0.066,\ 0.899 \pm 0.015)$
	$\boldsymbol{\beta}$	$(0.793 \pm 0.205,\ 1.489 \pm 0.473,\ 2.298 \pm 0.053)$
8	$\boldsymbol{\phi}$	$(-0.102 \pm 0.029,\ -0.198 \pm 0.061,\ -0.400 \pm 0.053,\ -0.596 \pm 0.053)$
	$\boldsymbol{\beta}$	$(0.801 \pm 0.076,\ 1.503 \pm 0.062,\ 2.293 \pm 0.094)$

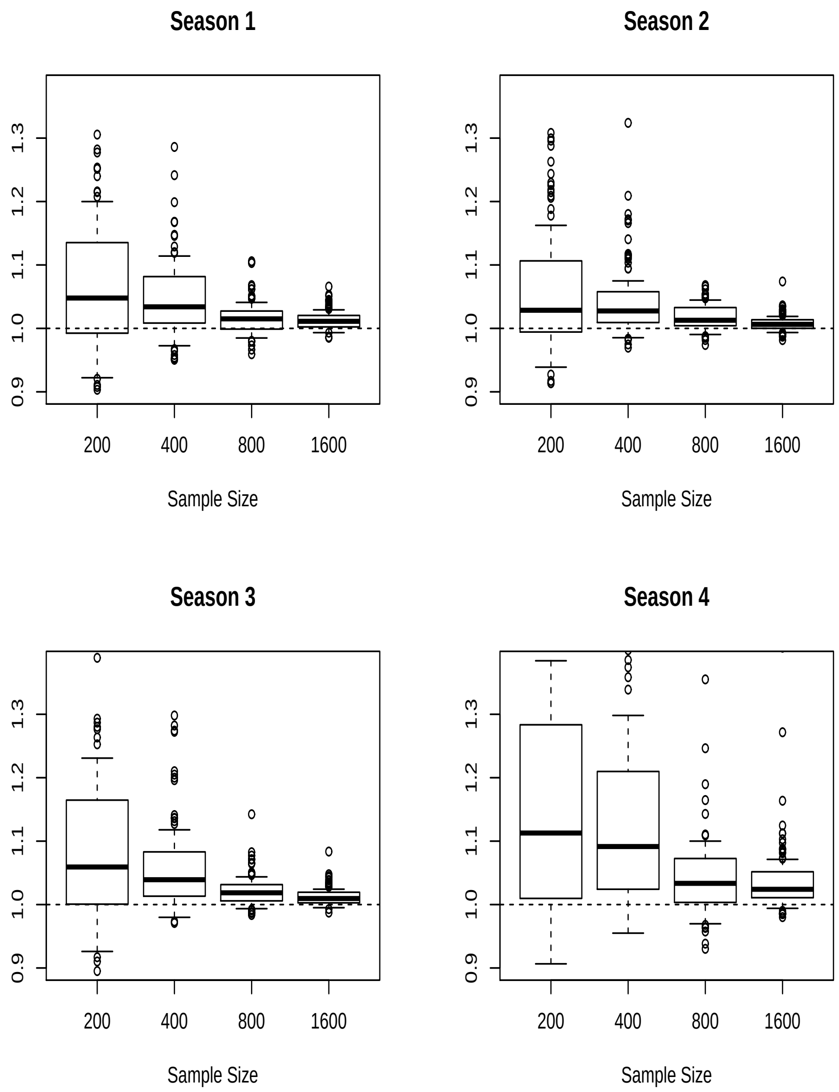

Figure 3-1: Model 1

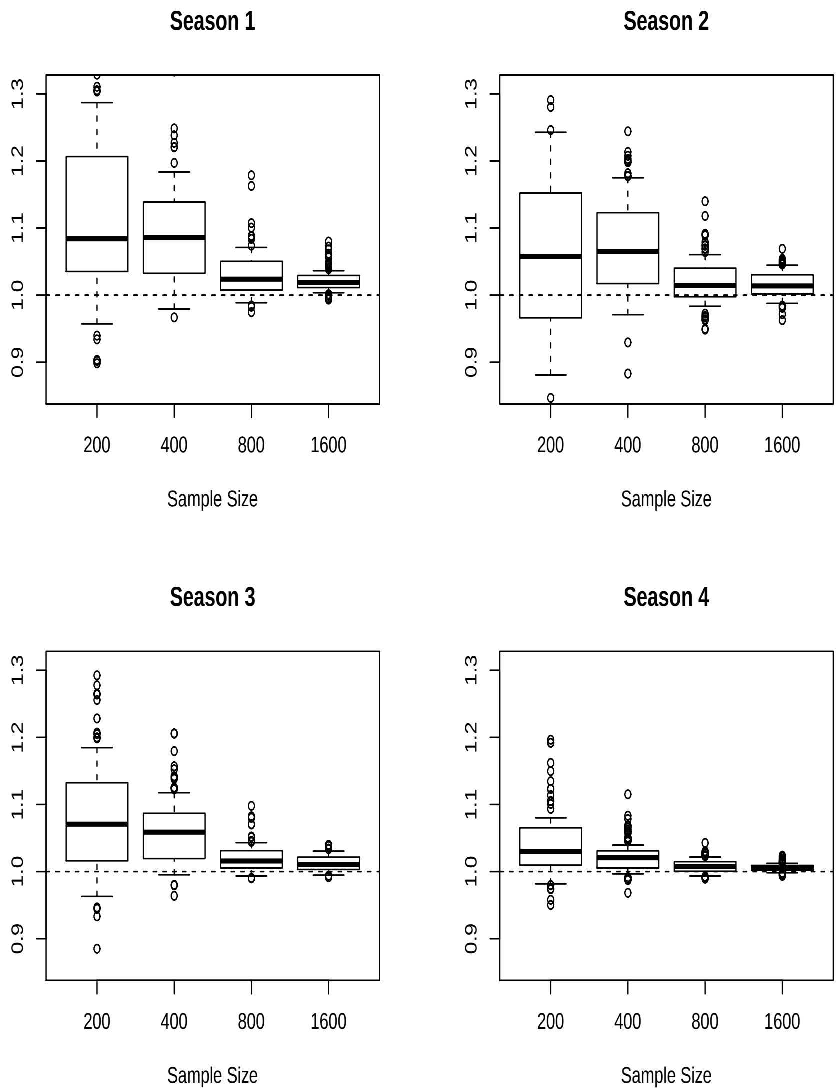

Figure 3-2: Model 2

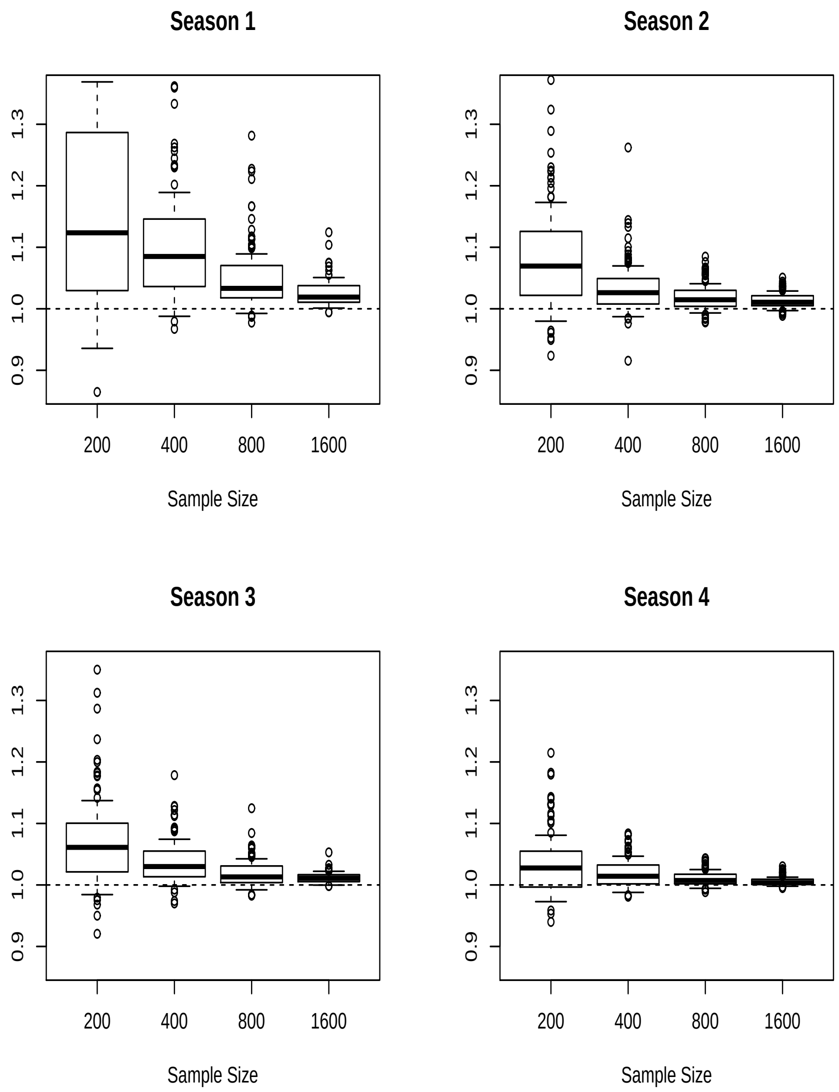

Figure 3-3: Model 3

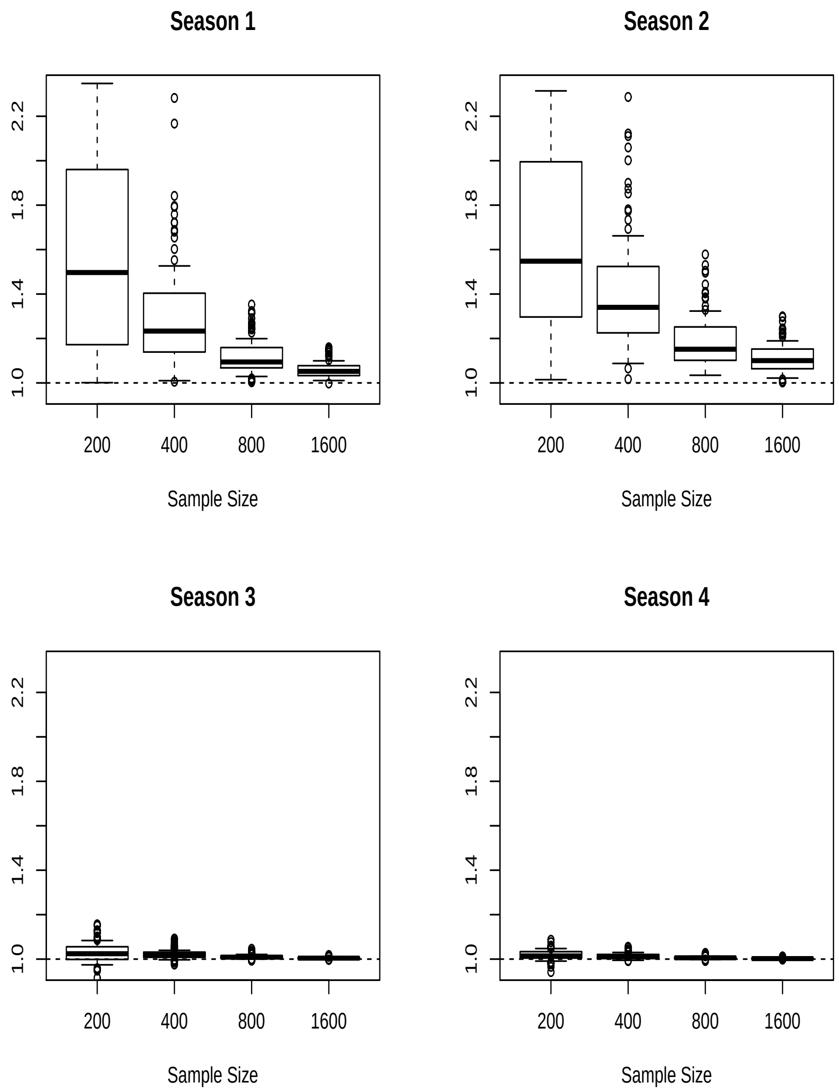

Figure 3-4: Model 4

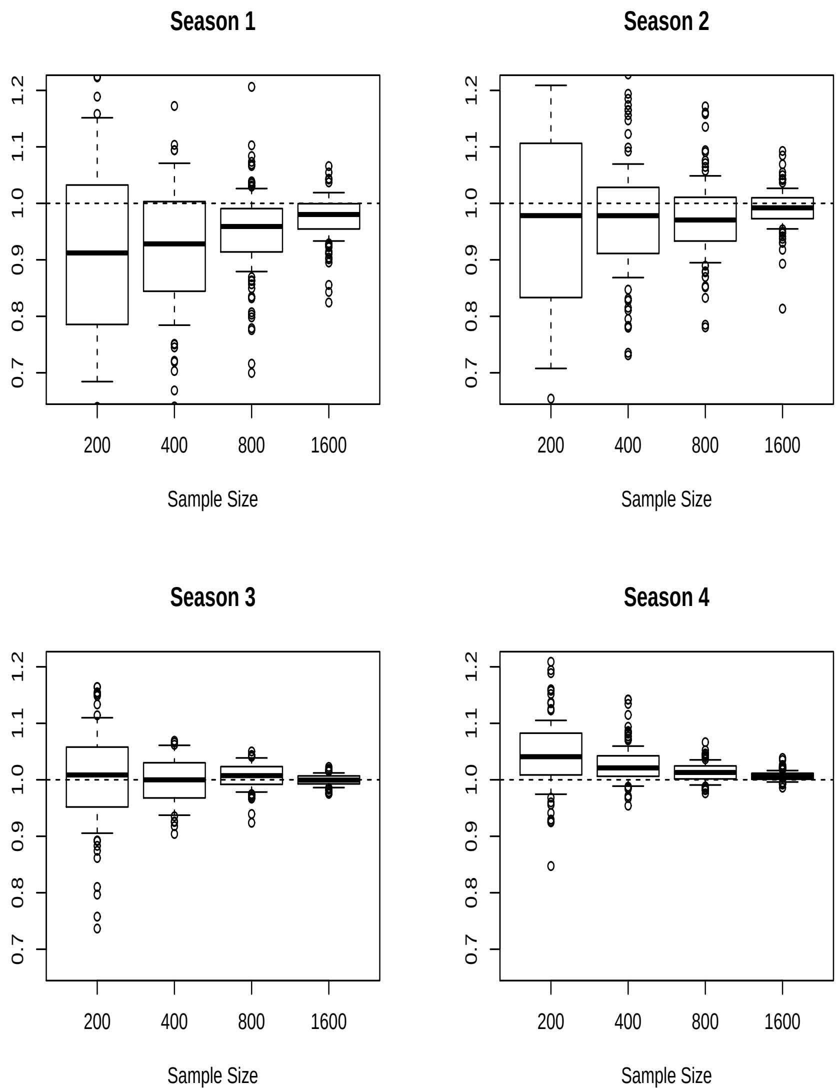

Figure 3-5: Model 5

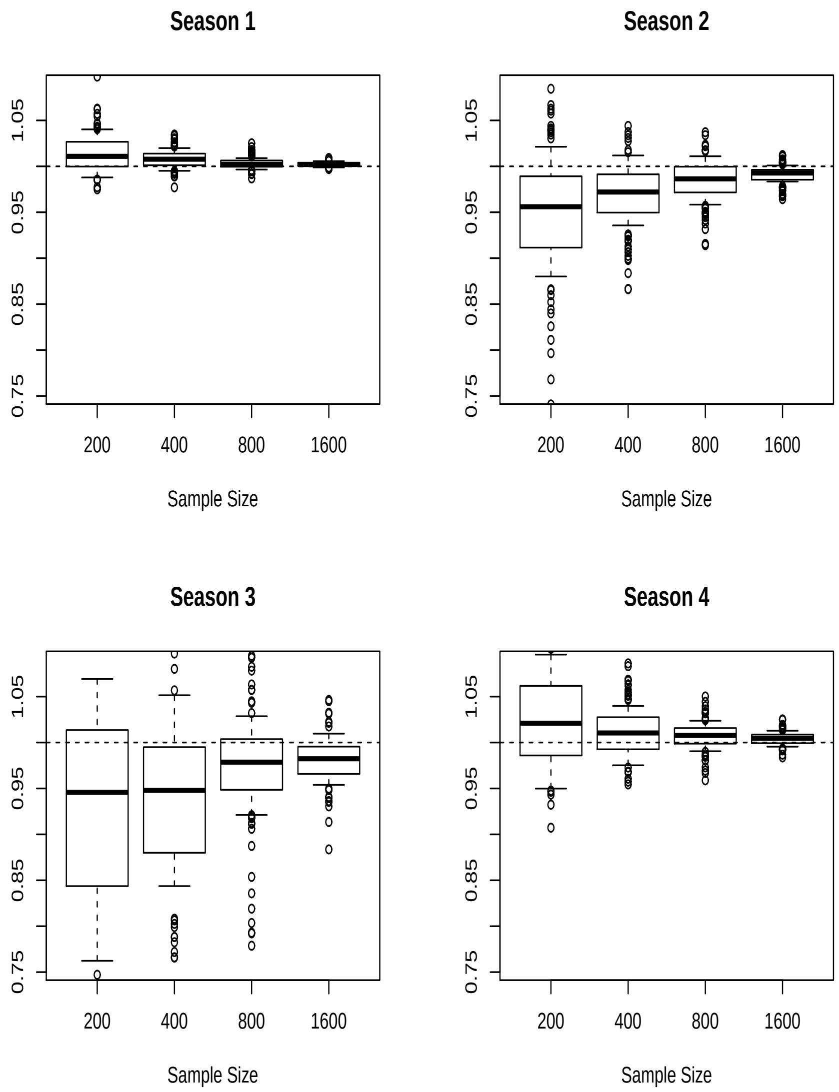

Figure 3-6: Model 6

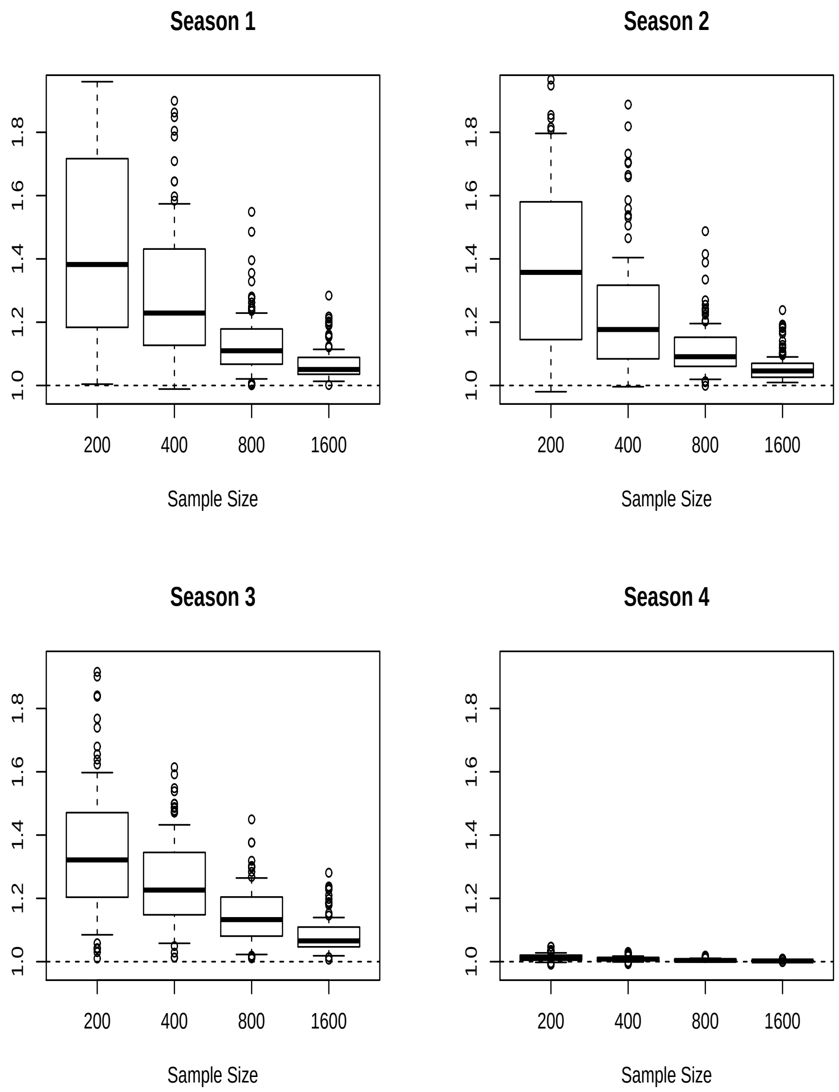

Figure 3-7: Model 7

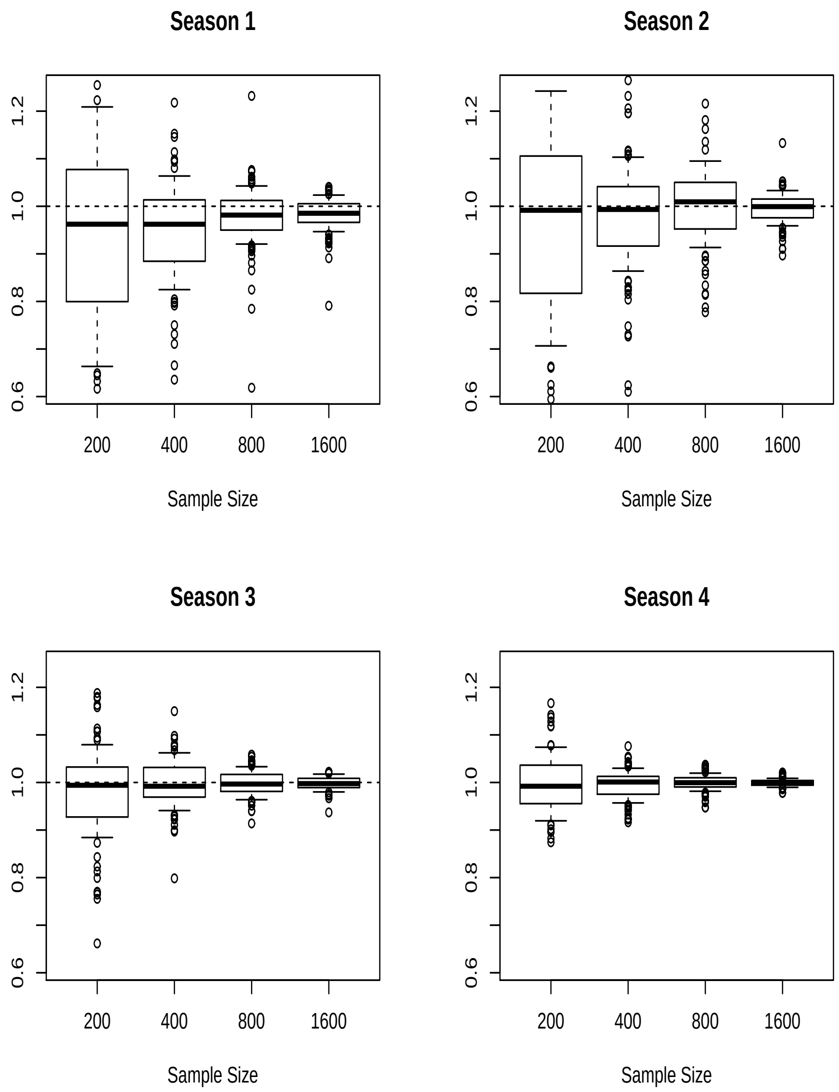

Figure 3-8: Model 8

1913 to 2011.

We will start with the monthly global temperature data from January 1979 to December 2014. This data is provided by the National Space Science and Technology Center. There are a total of $n_T = 36$ cycles for the $n = 432$ monthly temperatures with the period $T = 12$. We start with the full model (3.2), which includes a deterministic trend, seasonal components and a random error term. The seasonal effects and trend are respectively estimated by (3.5) and (3.6) with the linear B-spline ($m = 2$). The number of knots is $N = 4$ which is obtained by $N = [n^{1/5}]$, the same formula as in the simulation studies. We explore several AR models with different orders using the residuals $\{\hat{x}_t\}_{t=1}^{432}$, and an AR(2) appears to be the most appropriate one with the smallest Akaike information criterion (AIC). For the full model (3.2), the estimates of the AR(2) model coefficients with the standard errors are $\hat{\phi}_1 = 0.565 \pm 0.047$, $\hat{\phi}_2 = 0.234 \pm 0.047$, and the estimates of the seasonal components with the standard deviations are $\hat{\beta}_1 = -0.002 \pm 0.124$, $\hat{\beta}_2 = -0.012 \pm 0.139$, $\hat{\beta}_3 = -0.014 \pm 0.156$, $\hat{\beta}_4 = -0.005 \pm 0.165$, $\hat{\beta}_5 = -0.001 \pm 0.170$, $\hat{\beta}_6 = 0.011 \pm 0.171$, $\hat{\beta}_7 = -0.007 \pm 0.169$, $\hat{\beta}_8 = 0.0049 \pm 0.163$, $\hat{\beta}_9 = 0.019 \pm 0.154$, $\hat{\beta}_{10} = 0.011 \pm 0.136$, $\hat{\beta}_{11} = 0.002 \pm 0.120$. Since none of these seasonal effects are significantly different from zero according to Theorem 3.2.1, we reduce the model to

$$y_{iT+\nu} = g(u_{iT+\nu}) + x_{iT+\nu}. \tag{3.10}$$

The observations $\{y_t\}_{t=1}^{n}$ contain a trend in the reduced model. We use the two-step method proposed in Chapter 2 to analyze $\{y_t\}_{t=1}^{n}$. An AR(2) appears to be the best again for the residual sequence according to AIC. Model adequacy checking is conducted for the reduced model (3.10) with AR(2) residual autocorrelations. The residual sample autocorrelation function at lags 0-20 in Figure 3-10 shows no significant serial correlation. Therefore, the reduced model (3.10) with AR(2) is the most appropriate model for the monthly global temperature data. The Yule-Walker estimates are $\hat{\phi}_1 = 0.565 \pm 0.047$, $\hat{\phi}_2 = 0.233 \pm 0.047$. It is worth mentioning that the AR(2) model is not only adequate for the data, but also parsimonious compared with a PAR, as even a PAR(1) model with $T = 12$ has 12 time series model parameters.

Moreover, we obtain 90% and 95% confidence bands using the method in Shao and Yang (2012) based on the reduced model (3.10). These confidence bands which have at least those significance levels are conservative. In Figure 3-9, the dashed curve is the linear B-spline estimate of the trend, the dashed and dotted bands respectively have the significance levels 90% and 95%, and the dash-dot line is $y = 0$. It is interesting to note that the upward trend statistically is not very significant in that part of $y = 0$ is very close to the 95% confidence band or overlapped with the 90% confidence band. However, since there are $n_T = 36$ cycles in the data set, further research is needed when more observations are available.

Then our proposed method will be applied to the monthly global land-ocean temperature anomaly indices. A temperature anomaly is the difference between the temperature and the average temperature of a base period. There are totally $n = 1536$ observations in the data set. Hansen *et al.* (1981) and Jones *et al.* (1999) provided a detailed information about the data. From the scatter plot of Figure 3-11, the trend exhibits two increasing patterns with a relatively stable period in the middle. Shao (2009) analyzed the data set using partially linear models with AR(2) time series error terms. In particular, she estimated the trend function by the local linear kernel method and studied the possible seasonal effects for the data. She concluded that AR(2) appeared to be sufficient, and there is no seasonal effect in the means. We explore the partially linear models with AR(2) errors using the proposed procedure. Notice that an autoregressive time series is a special PAR with period $T = 1$. In Figure 3-11, the dashed curve is the linear B-spline estimate of the global trend function in partially linear models with AR(2) error terms. Model adequacy checking is conducted using AR(2) residual autocorrelations. The sample autocorrelation function at lags 0-30 shows no significant serial correlation, which is shown in Figure 3-12. The estimates for the coefficient $\boldsymbol{\phi}$ with the standard errors from the residuals are $\hat{\phi}_1 = 0.509 \pm 0.025$, $\hat{\phi}_2 = 0.225 \pm 0.025$. The estimates for seasonal effects $\boldsymbol{\beta}$ with the standard deviations in partially linear models are $\hat{\beta}_1 = -0.010 \pm 0.073$, $\hat{\beta}_2 = -0.013 \pm 0.082$, $\hat{\beta}_3 = -0.001 \pm 0.092$, $\hat{\beta}_4 = -0.004 \pm 0.097$, $\hat{\beta}_5 = -0.007 \pm 0.100$, $\hat{\beta}_6 = -0.007 \pm 0.101$, $\hat{\beta}_7 = -0.007 \pm 0.100$, $\hat{\beta}_8 = -0.008 \pm 0.097$, $\hat{\beta}_9 = -0.006 \pm 0.092$, $\hat{\beta}_{10} = -0.008 \pm 0.082$, $\hat{\beta}_{11} = -0.011 \pm 0.073$. None of these seasonal effects are significantly different from zero according to Theorem 3.2.1.

Next we will illustrate the proposed method by the streamflows of the Fraser River. The periodicity in the second moment was first noticed by Vecchia and Ballerini (1991). Since then this series has been adopted by many studies about periodic autoregressive moving-average (PARMA) time series for the purpose of illustration. A PARMA(p,q) becomes a PAR(p) if $q = 0$. For example, McLeod (1994) discussed diagnostic checking of PARMA models with the monthly log transformed streamflows from 1913 to 1990; Tesfaye *et al* (2006) used the data from 1912 to 1984 to present how to identify PARMA models to capture the seasonal variations in river flow statistics; Anderson *et al* (2013) applied their forecasting procedures to the same time series. All these studies suggested that the streamflows exhibit periodicity in correlations which can be taken into account by PARMA models. The observations under consideration are the quarterly data obtained from the monthly streamflows which can be downloaded at the website of the Canada of Environment.

We apply the proposed method to the quarterly streamflows data of the Fraser River and estimate the seasonal effects and global trend function by (3.5) and (3.6) with the linear B-spline in partially linear models. The trend estimates along with the quarterly streamflows are shown in Figure 3-13. The residuals are calculated by subtracting the estimates $\hat{g}(u)$ and $\hat{\boldsymbol{\beta}}$ from the observations. These residuals play the role of the "observations" in the following analysis using "`pear`". We fit several PAR models with different orders, and a PAR(1) appears to be the most appropriate one according to the BIC criterion. Model adequacy checking is conducted using PAR(1) residual autocorrelations. The sample autocorrelation function at lags 0-25 shows no significant serial correlation, which is shown in Figure 3-14. In other words, all the sample autocorrelations are within the 95% confidence interval, which indicates that the PAR(1) model is adequate. The estimates of $\boldsymbol{\phi}$ with the standard errors from the residuals are $\hat{\phi}(1) = 0.300 \pm 0.052$, $\hat{\phi}(2) = 0.663 \pm 0.277$, $\hat{\phi}(3) = 0.399 \pm 0.107$, $\hat{\phi}(4) = 0.188 \pm 0.048$. The estimates of seasonal effects $\boldsymbol{\beta}$ with the standard deviations in partially linear models are $\hat{\beta}_1 = -659 \pm 34$, $\hat{\beta}_2 = 3017 \pm 70$, $\hat{\beta}_3 = 2241 \pm 71$. According to Theorem 3.2.1, these seasonal effects are significantly different from zero.

3.5 Concluding Remarks

In this chapter, we proposed a semiparametric three-step method for analyzing periodic time series with trend and seasonality. The proposed procedure is not only computationally accessible but theoretically well-justified for the constant B-spline. We anticipate that it can be generalized to higher order B-splines. We expect that the oracle efficiency showed in this chapter can be extended to partially linear models of which trends are estimated by other nonparametric methods, such as smoothing splines (for example in Wang (1998) and Kohn *et al.* (1992)), under general conditions.

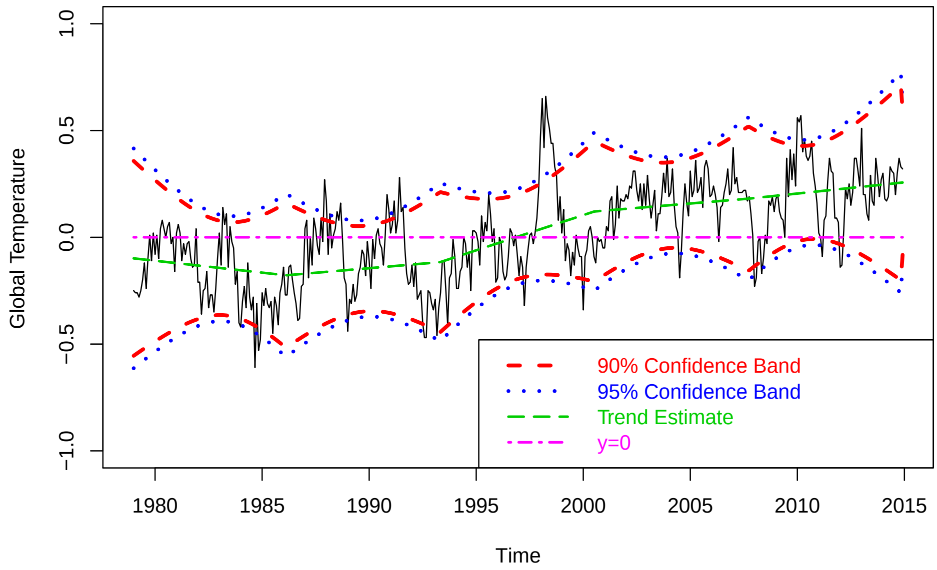

Figure 3-9: Confidence Band for Monthly Global Temperature Data, January 1979-December 2014

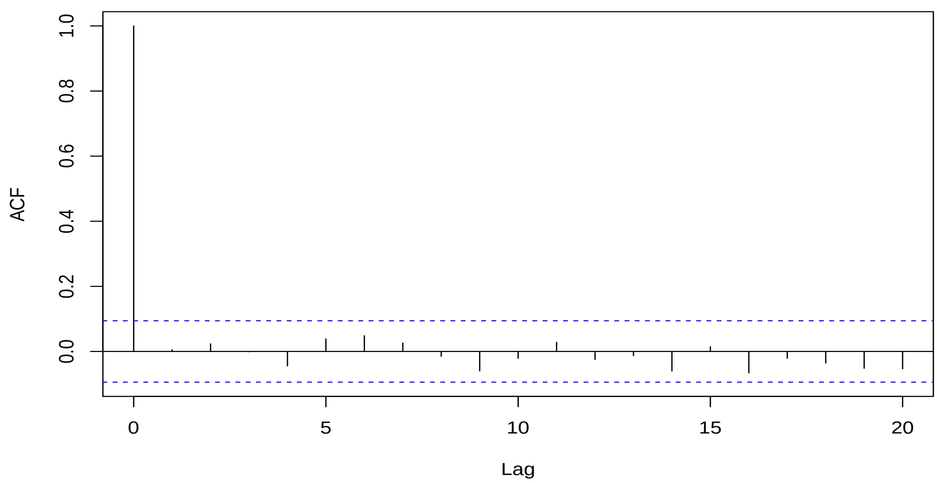

Figure 3-10: AR(2) Residual Autocorrelations for Monthly Global Temperature Data

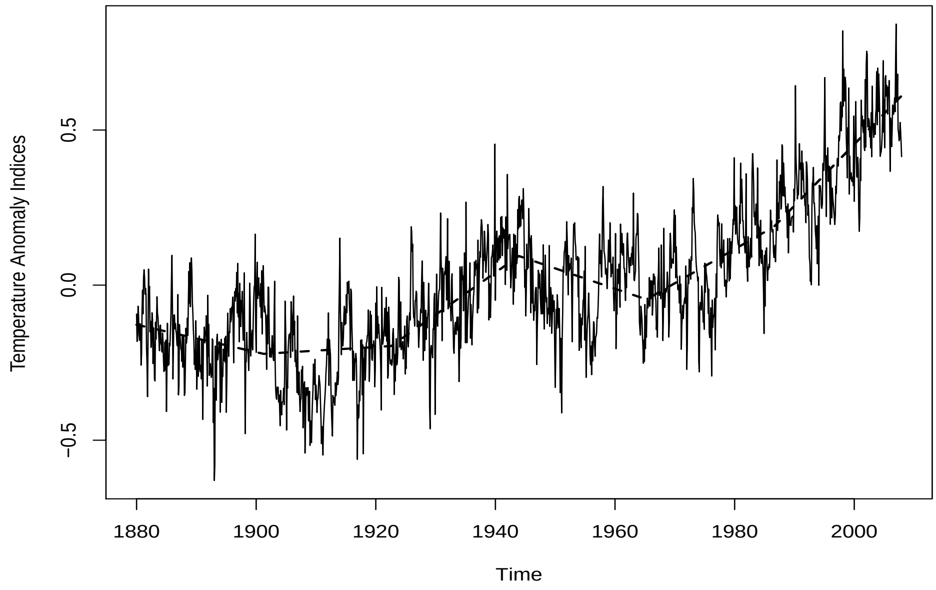

Figure 3-11: Monthly Global Land and Ocean Combined Anomaly Indices with B-spline Trend Estimate

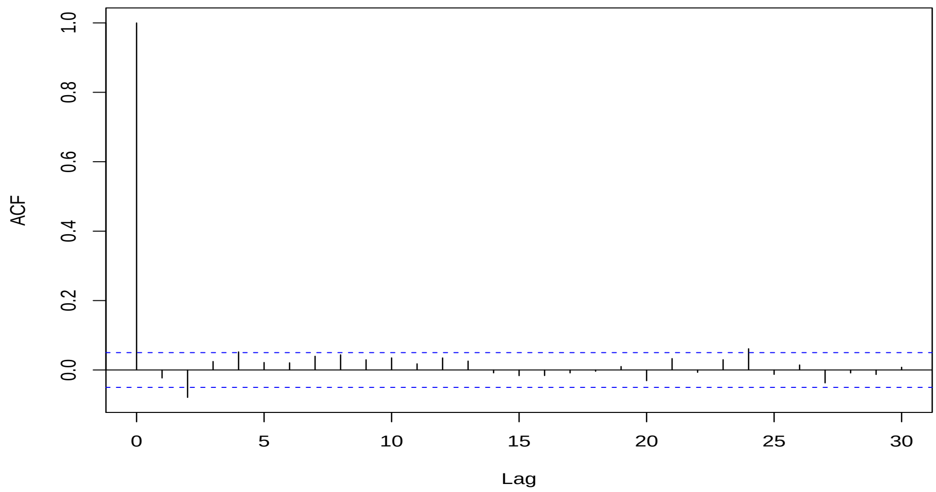

Figure 3-12: AR(2) Residual Autocorrelations for Monthly Global Land and Ocean Combined Anomaly Indices

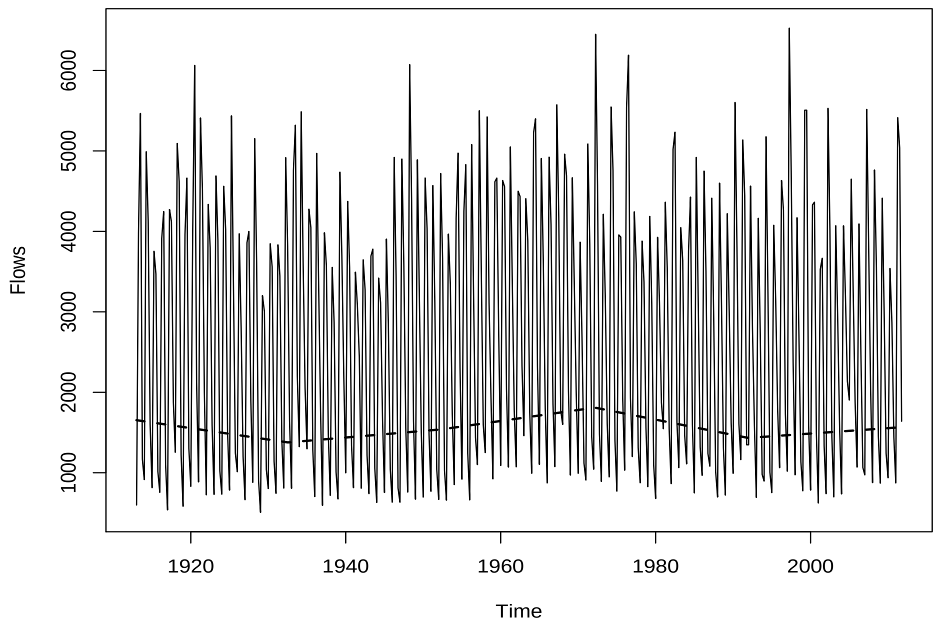

Figure 3-13: Quarterly Streamflows of Fraser River in m^3 with B-spline Trend Estimate

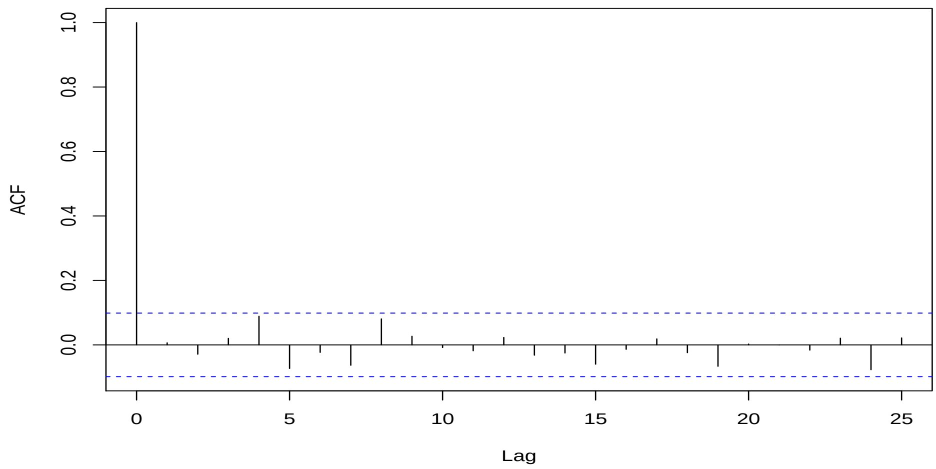

Figure 3-14: PAR(1) Residual Autocorrelations for Quarterly Streamflows of Fraser River

Wang and Yang (2007) and Xue and Yang (2006) compared how faster B-splines is than kernel estimation for functions. The simulation studies shows that our method works well for the linear B-spline estimation. Replacing unobservable error terms by residuals has been widely applied in time series analysis, for example, see Brockwell and Davis (1991) and Shumway and Stoffer (2011). However, an interesting and critical question is whether the estimators are oracally efficient; in other words, whether the analysis from the residuals $\{\hat{x}_t\}_{t=1}^n$ is asymptotically equivalent to the one based on $\{x_t\}_{t=1}^n$. For the partially linear model in (3.2),we established the oracle efficiency of the model parameters, which is a desirable property and ensures that replacement is appropriate.

3.6 Proof

In this section, we will show Theorem 3.2.1-3.2.2 through three lemmas. Hereafter, $\|\cdot\|$ refers to the Euclidean norm, $\|g-\tilde{g}\|_\infty$ stands for the supremum norm, and $U(\cdot)$ and $u(\cdot)$ denote the uniform boundedness of a matrix and a scalar, respectively. Without loss of generality, we always assume that the number of observations is an multiple of the period T and the order of PAR is less than the period (i.e. $p < T$). Although the assumption is not necessarily true, it will greatly simplify the notation without changing the asymptotic properties of a statistic under consideration.

We decompose $\hat{\boldsymbol{g}}$ in (3.6) into three terms:

$$\hat{\boldsymbol{g}} = \tilde{\boldsymbol{g}} + \tilde{\boldsymbol{x}} + \boldsymbol{P_B}(\boldsymbol{D}\boldsymbol{\beta} - \boldsymbol{D}\hat{\boldsymbol{\beta}}), \tag{3.11}$$

where $\tilde{\boldsymbol{g}} = \boldsymbol{P_B}\boldsymbol{g}$ and $\tilde{\boldsymbol{x}} = \boldsymbol{P_B}\boldsymbol{x}$ are the projections of $\boldsymbol{g}$ and $\boldsymbol{x}$, respectively. Furthermore, we decompose the residuals $\hat{\boldsymbol{x}}$ in (3.7) into three components as follows:

$$\hat{\boldsymbol{x}} = (\boldsymbol{g} - \tilde{\boldsymbol{g}}) + (\boldsymbol{x} - \tilde{\boldsymbol{x}}) + (\boldsymbol{I} - \boldsymbol{P_B})\boldsymbol{D}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}). \tag{3.12}$$

We will consider the asymptotics of the three terms in (3.12).

Under Assumptions 1 and 2, according to Theorem 5.1 of Huang (2003),

$$\|g - \tilde{g}\|_\infty = \sup_{u \in [0,1]} |g(u) - \tilde{g}(u)| = O(h^m). \tag{3.13}$$

According to Shao and Yang (2011),

$$\frac{1}{n}(\boldsymbol{g} - \tilde{\boldsymbol{g}})'(\boldsymbol{g} - \tilde{\boldsymbol{g}}) = O(n^{-2m/(2m+1)}), \tag{3.14}$$

$$\frac{1}{n}(\boldsymbol{g} - \tilde{\boldsymbol{g}})'\boldsymbol{x} = O_p(n^{-(4m+1)/(4m+2)}), \tag{3.15}$$

$$\frac{1}{n}(\boldsymbol{g} - \tilde{\boldsymbol{g}})'\tilde{\boldsymbol{x}} = O_p(n^{-2m/(2m+1)}). \tag{3.16}$$

Lemma 3.6.1. *Under Assumptions 1-4, $\frac{1}{n_T}\boldsymbol{D}'(\boldsymbol{I} - \boldsymbol{P_B})\boldsymbol{D} = \boldsymbol{A}_0 + o(1)$, where the $(T-1) \times (T-1)$ matrix $\boldsymbol{A}_0$ is defined as follows:*

$$\boldsymbol{A}_0 = \begin{pmatrix} 1 - \frac{1}{T} & -\frac{1}{T} & \cdots & -\frac{1}{T} \\ -\frac{1}{T} & 1 - \frac{1}{T} & \cdots & -\frac{1}{T} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{T} & -\frac{1}{T} & \cdots & 1 - \frac{1}{T} \end{pmatrix}.$$

Proof. In this case where $m = 1$,

$$\boldsymbol{D}'(\boldsymbol{I} - \boldsymbol{P_B}) = (\boldsymbol{A}_1, \cdots, \boldsymbol{A}_1) + u\left\{(nh)^{-1}\right\},$$

where the $(T-1)\times T$ matrix $\boldsymbol{A}_1$ is as follows:

$$\boldsymbol{A}_1 = \begin{pmatrix} 1-\frac{1}{T} & -\frac{1}{T} & \cdots & -\frac{1}{T} & -\frac{1}{T} \\ -\frac{1}{T} & 1-\frac{1}{T} & \cdots & -\frac{1}{T} & -\frac{1}{T} \\ \vdots & \vdots & & \vdots & \vdots \\ -\frac{1}{T} & -\frac{1}{T} & \cdots & 1-\frac{1}{T} & -\frac{1}{T} \end{pmatrix}.$$

and thus

$$\frac{1}{n_T}\boldsymbol{D}'\left(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}}\right)\boldsymbol{D} = \boldsymbol{A}_0 + o(1).$$

The proof is complete.

Lemma 3.6.2. *Under Assumptions 1-4,* $\|E(\hat{\boldsymbol{\beta}})-\boldsymbol{\beta}\| = O(n^{-1})$

Proof. We have

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + \{\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}})\boldsymbol{D}\}^{-1}\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}})\boldsymbol{g}.$$

Therefore,

$$E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} = \{\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}})\boldsymbol{D}\}^{-1}\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}})\boldsymbol{g} \tag{3.17}$$

According to Lemma 3.6.1,

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} &= \left\{\boldsymbol{D}'\left(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}}\right)\boldsymbol{D}\right\}^{-1}\boldsymbol{D}'\left(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}}\right)\boldsymbol{g} \\ &= \frac{1}{n_T}(\boldsymbol{A},\cdots,\boldsymbol{A})\boldsymbol{g} + o(n_T^{-1}), \end{aligned}$$

where the $(T-1)\times T$ matrix $\boldsymbol{A} = (\boldsymbol{I}, \boldsymbol{-1})$ with $\boldsymbol{I}$ being the identity matrix and

$\mathbf{-1} = (-1, -1, \cdots, -1)'$. It is as follows

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & \cdots & 0 & -1 \\ 0 & 1 & \cdots & 0 & -1 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -1 \end{pmatrix}.$$

According to Assumption 1, we have

$$\|E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}\| = O(n^{-1}).$$

The proof is complete.

Proof of Theorem 3.2.1. From Lemma 3.6.1,

$$\left\{\frac{1}{n_T}\boldsymbol{D}'\left(\boldsymbol{I} - \boldsymbol{P_B}\right)\boldsymbol{D}\right\}^{-1} = \boldsymbol{A}_0^{-1} + o(1),$$

where

$$\boldsymbol{A}_0^{-1} = \begin{pmatrix} 2 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 2 \end{pmatrix}.$$

Let $\boldsymbol{x}_i = (x_{iT+1}, \cdots, x_{iT+T})'$. Then $\{\boldsymbol{x}_i\}$ is a multivariate stationary time series with $\mathrm{E}(\boldsymbol{x}_i) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{x}_i, \boldsymbol{x}_{i-k}) = \boldsymbol{W}_k$ $(k = 0, \pm 1, \cdots)$, where the $T \times T$ matrix $\boldsymbol{W}_k = (\gamma_i(kT + i - j))_{i=1,\cdots,T; j=1,\cdots,T}$. For any T dimension vector $\boldsymbol{a}$, define $u_i = \boldsymbol{a}'\boldsymbol{x}_i$. Then $\{u_i\}$ is a stationary time series with $\mathrm{E}(u_i) = 0$ and $\gamma_u(k) = \mathrm{Cov}(u_i, u_{i-k}) = \boldsymbol{a}'\boldsymbol{W}_k\boldsymbol{a}$. Since $\{x_{iT+\nu}, 1 \le \nu \le T\}_{i=0}^{n_T-1}$ is causal, $\{u_i\}$ satisfies Theorem 7.1.2 of Brockwell and

Davis (1991), which implies

$$\sqrt{n_T}\left(\frac{1}{n_T}\sum_{i=0}^{n_T-1}\boldsymbol{a}'\boldsymbol{x}_i\right)\overset{D}{\Longrightarrow}N\left(0,\boldsymbol{a}'\boldsymbol{W}\boldsymbol{a}\right), \tag{3.18}$$

where $\boldsymbol{W}=\sum\limits_{k=-\infty}^{\infty}\boldsymbol{W}_k=\left(\sum_{k=-\infty}^{\infty}\gamma_i(kT+i-j)\right)_{i=1,\cdots,T;j=1,\cdots,T}$. Since $\boldsymbol{a}$ is arbitrary, from the Cramér-Wold device,

$$\sqrt{n_T}\left(\frac{1}{n_T}\sum_{i=0}^{n_T-1}\boldsymbol{x}_i\right)\overset{D}{\Longrightarrow}N(0,\boldsymbol{W}). \tag{3.19}$$

Notice that

$$\begin{aligned}
&\sqrt{n_T}\left\{\hat{\boldsymbol{\beta}}-\mathrm{E}\left(\hat{\boldsymbol{\beta}}\right)\right\}\\
=\;&\sqrt{n_T}\left\{\frac{1}{n_T}\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}})\boldsymbol{D}\right\}^{-1}\frac{1}{n_T}\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}})\boldsymbol{x}\\
=\;&\frac{1}{\sqrt{n_T}}\boldsymbol{A}_0^{-1}\sum_{i=0}^{n_T-1}\boldsymbol{A}_1\boldsymbol{x}_i+o_p(n_T^{-1/2})\\
=\;&\sqrt{n_T}\left(\frac{1}{n_T}\boldsymbol{A}_0^{-1}\boldsymbol{A}_1\sum_{i=0}^{n_T-1}\boldsymbol{x}_i\right)+o_p(n_T^{-1/2})\\
=\;&\sqrt{n_T}\left(\frac{1}{n_T}\boldsymbol{A}\sum_{i=0}^{n_T-1}\boldsymbol{x}_i\right)+o_p(n_T^{-1/2}).
\end{aligned}$$

where simple algebra yields

$$\boldsymbol{A}_0^{-1}\boldsymbol{A}_1=\boldsymbol{A}.$$

$\boldsymbol{A}$ is defined in the proof of Lemma 3.6.2 and $\boldsymbol{A}_1$ is defined in the proof of Lemma 3.6.1.

Then according (3.19), we have

$$\sqrt{n_T}\left\{\hat{\boldsymbol{\beta}}-E\left(\hat{\boldsymbol{\beta}}\right)\right\}\overset{D}{\Longrightarrow}N\left(\mathbf{0},\boldsymbol{V}\right), \tag{3.20}$$

where the $(T-1)\times(T-1)$ matrix $\boldsymbol{V} = \boldsymbol{AWA}'$. The matrix $\boldsymbol{V}$ is $(\boldsymbol{V})_{i,j} = \left[(\boldsymbol{W})_{i,j} - (\boldsymbol{W})_{T,j}\right] - \left[(\boldsymbol{W})_{i,T} - (\boldsymbol{W})_{T,T}\right]$, $(\boldsymbol{W})_{i,j} = \sum_{k=-\infty}^{\infty}\gamma_i(kT+i-j)$ is the (i,j)-th entry of matrix $\boldsymbol{W}$. Thus the (i,j)-th entry of $\boldsymbol{V}$ is

$$
\begin{aligned}
(\boldsymbol{V})_{i,j} &= \sum_{k=-\infty}^{\infty}\{\gamma_i(kT+i-j) - \gamma_i(kT+T-j) \\
&\quad -\gamma_i(kT+i-T) + \gamma_i(kT)\}.
\end{aligned}
$$

Then, according to Lemma 3.6.2 we have

$$E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta} = o_p(n^{-1/2}).$$

Slusky's Theorem together with (3.20) implies (3.9). The proof is complete.

Notice the difference of Yule-Walker estimators $\hat{\boldsymbol{\phi}}$ and $\tilde{\boldsymbol{\phi}}$ is

$$\hat{\boldsymbol{\phi}} - \tilde{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}^{-1}(\hat{\boldsymbol{\gamma}} - \tilde{\boldsymbol{\gamma}}) + \hat{\boldsymbol{\Gamma}}^{-1}(\tilde{\boldsymbol{\Gamma}} - \hat{\boldsymbol{\Gamma}})\tilde{\boldsymbol{\Gamma}}^{-1}\tilde{\boldsymbol{\gamma}}.$$

The proof of Theorem 3.2.2 is complete if we can show that $\hat{\boldsymbol{\phi}} - \tilde{\boldsymbol{\phi}} = o_p(n^{-1/2})$ or equivalently

$$\hat{\gamma}_\nu(k) - \tilde{\gamma}_\nu(k) = o_p(n^{-1/2}) \tag{3.21}$$

We will show (3.21) by Lemma 3.6.3 below. Define the following vectors

$$
\begin{aligned}
\boldsymbol{x}_\nu &= (x_\nu, x_{T+\nu}, \ldots, x_{(n_T-1)T+\nu})', \\
\boldsymbol{g}_\nu &= (g(u_\nu), g(u_{T+\nu}), \ldots, g(u_{(n_T-1)T+\nu}))'.
\end{aligned}
$$

This notation sometimes is applied to $\hat{\boldsymbol{x}}_\nu$, $\tilde{\boldsymbol{x}}_\nu$, $\tilde{\boldsymbol{g}}_\nu$ and so on. Notice that the vector $\boldsymbol{x}_\nu$ includes almost all but a finite number of the observations, and the vectors $\boldsymbol{x}_{iT+\nu}$ and $(x_\nu, x_{T+\nu}, \ldots, x_{(n_T-1)T+\nu})'$ are equivalent in the sense of asymptotics for any fixed i.

We will show (3.21) by the following lemma.

Lemma 3.6.3. *Under Assumptions 1-4, for any fixed* ν $\hat{\gamma}_\nu(k) - \tilde{\gamma}_\nu(k) = O_p(h^2)$ *and* $\hat{\gamma}_\nu(k) \xrightarrow{p} \gamma_\nu(k)$.

Proof. Without loss of generality, we assume that $\nu - k \geq 0$. According to (3.12), we have

$$\begin{aligned}
&\hat{\boldsymbol{\gamma}}_\nu(k) - \tilde{\boldsymbol{\gamma}}_\nu(k) = \frac{1}{n_T}\hat{\boldsymbol{x}}_\nu'\hat{\boldsymbol{x}}_{\nu-k} - \frac{1}{n_T}\boldsymbol{x}_\nu'\boldsymbol{x}_{\nu-k} \\
= \;& \frac{1}{n_T}\left(\boldsymbol{g}_\nu - \tilde{\boldsymbol{g}}_\nu\right)'\boldsymbol{x}_{\nu-k} - \frac{1}{n_T}\left(\boldsymbol{g}_\nu - \tilde{\boldsymbol{g}}_\nu\right)'\tilde{\boldsymbol{x}}_{\nu-k} \\
&+\frac{1}{n_T}\tilde{\boldsymbol{x}}_\nu'\tilde{\boldsymbol{x}}_{\nu-k} - \frac{1}{n_T}\boldsymbol{x}_{\nu-k}'\tilde{\boldsymbol{x}}_\nu - \frac{1}{n_T}\tilde{\boldsymbol{x}}_{\nu-k}'\boldsymbol{x}_\nu \\
&+\frac{1}{n_T}\left(\boldsymbol{g}_\nu - \tilde{\boldsymbol{g}}_\nu\right)'\left(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k}\right) \\
&+\frac{1}{n_T}\left(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k}\right)'\boldsymbol{x}_\nu - \frac{1}{n_T}\left(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k}\right)'\tilde{\boldsymbol{x}}_\nu \\
&+\frac{1}{n_T}\left(\boldsymbol{g}_\nu - \tilde{\boldsymbol{g}}_\nu\right)'\left\{\left(\boldsymbol{I} - \boldsymbol{P_B}\right)\boldsymbol{D}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right\}_{\nu-k} \\
&+\frac{1}{n_T}\left(\boldsymbol{g}_{\nu-k} - \tilde{\boldsymbol{g}}_{\nu-k}\right)'\left\{\left(\boldsymbol{I} - \boldsymbol{P_B}\right)\boldsymbol{D}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right\}_{\nu} \\
&+\frac{1}{n_T}\left(\boldsymbol{x}_\nu - \tilde{\boldsymbol{x}}_\nu\right)'\left\{\left(\boldsymbol{I} - \boldsymbol{P_B}\right)\boldsymbol{D}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right\}_{\nu-k} \\
&+\frac{1}{n_T}\left(\boldsymbol{x}_{\nu-k} - \tilde{\boldsymbol{x}}_{\nu-k}\right)'\left\{\left(\boldsymbol{I} - \boldsymbol{P_B}\right)\boldsymbol{D}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right\}_{\nu} \\
&+\frac{1}{n_T}\left\{\left(\boldsymbol{I} - \boldsymbol{P_B}\right)\boldsymbol{D}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right\}_{\nu}'\left\{\left(\boldsymbol{I} - \boldsymbol{P_B}\right)\boldsymbol{D}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)\right\}_{\nu-k}.
\end{aligned} \tag{3.22}$$

We will consider the orders of the terms above. First

$$\begin{aligned}
&\frac{1}{n_T}\left\{\left(\boldsymbol{I} - \boldsymbol{P_B}\right)\boldsymbol{D}\right\}'\left(\boldsymbol{x} - \tilde{\boldsymbol{x}}\right) \\
= \;& \frac{1}{n_T}\boldsymbol{D}'\left(\boldsymbol{I} - \boldsymbol{P_B}\right)'\left(\boldsymbol{I} - \boldsymbol{P_B}\right)\boldsymbol{x} \\
= \;& \frac{1}{n_T}\boldsymbol{D}'\left(\boldsymbol{I} - \boldsymbol{P_B}\right)\boldsymbol{x} \\
= \;& \frac{1}{n_T}\sum_{i=0}^{n_T-1}\boldsymbol{A}_1\boldsymbol{x}_i + u_p(n_T^{-1}).
\end{aligned}$$

By following a similar discussion to derivation of (3.18), we obtain $\frac{1}{n_T}\sum_{i=0}^{n_T-1}\boldsymbol{A}_1\boldsymbol{x}_i = O_p(n^{-1/2})$. Thus

$$\frac{1}{n_T}\left\|\left\{\left(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}}\right)\boldsymbol{D}\right\}'\left(\boldsymbol{x}-\tilde{\boldsymbol{x}}\right)\right\| = O_p(n^{-1/2}). \tag{3.23}$$

Next, we note $\boldsymbol{D}'\left(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}}\right) = u\left(1\right)$, and thus the (i,j)th entry of $\boldsymbol{D}'\left(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}}\right)\left(\boldsymbol{g}-\tilde{\boldsymbol{g}}\right)$ is $O\left(n_T\right)\|g-\tilde{g}\|_\infty$. Therefore,

$$\begin{aligned}
&\frac{1}{n_T}\left\|\boldsymbol{D}'\left(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}}\right)\left(\boldsymbol{g}_\nu-\tilde{\boldsymbol{g}}_\nu\right)\right\| \\
&\leq \frac{1}{n_T}O\left(n_T\right)\|g-\tilde{g}\|_\infty \\
&= O(n^{-m/(2m+1)}).
\end{aligned} \tag{3.24}$$

According to Theorem 3.2.1, we have

$$\|\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\| = O_p(n^{-1/2}). \tag{3.25}$$

According to (3.23), (3.24), and (3.25), we have

$$\begin{aligned}
\frac{1}{n_T}\left(\boldsymbol{g}_\nu-\tilde{\boldsymbol{g}}_\nu\right)'\left\{\left(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}}\right)\boldsymbol{D}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right\}_{\nu-k} &= O_p(n^{-5/6}), \\
\frac{1}{n_T}\left(\boldsymbol{x}_\nu-\tilde{\boldsymbol{x}}_\nu\right)'\left\{\left(\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{B}}\right)\boldsymbol{D}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right\}_{\nu-k} &= O_p(n^{-1}).
\end{aligned}$$

According to Lemmas 2.7.2 and 2.7.3 in Chapter 2,

$$\begin{aligned}
&\frac{1}{n_T}\tilde{\boldsymbol{x}}_\nu'\tilde{\boldsymbol{x}}_{\nu-k} = O_p(n^{-2m/(2m+1)}), \\
&\frac{1}{n_T}\boldsymbol{x}_{\nu-k}'\tilde{\boldsymbol{x}}_\nu = O_p(n^{-2m/(2m+1)}).
\end{aligned} \tag{3.26}$$

At last,

$$
\begin{aligned}
& \left\{(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{D}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right\}'\left\{(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{D}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)\right\} \\
= & \left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)'\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P_B})'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{D}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right) \\
= & \left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right)'\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{D}\left(\boldsymbol{\beta}-\hat{\boldsymbol{\beta}}\right) \\
= & \boldsymbol{g}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{D}\left\{\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{D}\right\}^{-1}\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{g} \\
& +\boldsymbol{x}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{D}\left\{\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{D}\right\}^{-1}\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{x} \\
= & \left\|\left\{\frac{1}{n_T}\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{D}\right\}^{-1}\left\{\frac{1}{n_T}\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{g}\right\}\right\|^2 \\
& +\left\|\left\{\frac{1}{n_T}\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{D}\right\}^{-1}\left\{\frac{1}{n_T}\boldsymbol{D}'(\boldsymbol{I}-\boldsymbol{P_B})\boldsymbol{x}\right\}\right\|^2 \\
= & O(1)+O_p\left(n_T^{-1}\right)=O_p(1).
\end{aligned}
\tag{3.27}
$$

According to (3.14), (3.15), (3.16), (3.23), (3.24), (3.25), (3.26), (3.27), and Theorem 3.2.1, the order of the dominant terms in (3.22) is $O_p(n^{-2/3})$, which implies $\hat{\gamma}_\nu(k)-\tilde{\gamma}_\nu(k)=O_p(h^2)=O_p(n^{-2/3})$ and $\hat{\gamma}_\nu(k)\xrightarrow{p}\gamma_\nu(k)$ for any fixed $1\le\nu\le T$. Therefore $\hat{\gamma}_\nu(k)-\tilde{\gamma}_\nu(k)=o_p(n^{-1/2})$. The proof is complete.

Notice that for any fixed $1\le\nu\le T$

$$
\begin{aligned}
& \hat{\sigma}_\nu^2-\tilde{\sigma}_\nu^2 \\
= & \left\{\hat{\gamma}_\nu(0)-\sum_{k=1}^{p}\hat{\phi}_k(\nu)\hat{\gamma}_{\nu-k}(-k)\right\}-\left\{\tilde{\gamma}_\nu(0)-\sum_{k=1}^{p}\tilde{\phi}_k(\nu)\tilde{\gamma}_{\nu-k}(-k)\right\} \\
= & \left\{\hat{\gamma}_\nu(0)-\tilde{\gamma}_\nu(0)\right\}-\left\{\sum_{k=1}^{p}\hat{\phi}_k(\nu)\hat{\gamma}_{\nu-k}(-k)-\sum_{k=1}^{p}\tilde{\phi}_k(\nu)\tilde{\gamma}_{\nu-k}(-k)\right\} \\
= & o_p(1)
\end{aligned}
$$

The proof of Theorem 3.2.2 is complete.

Chapter 4

Future Research

4.1 Introduction

In this chapter, we will make the inference for the logistic regression models using the nonparametric estimation method. Logistic regression models are widely used in applications. In these models, the outcome variable y is binary with a value of either 0 or 1, and the covariates $\boldsymbol{x}$ can be both discrete and continuous. The primary interest of this topic is the estimation of the conditional mean $\mathrm{E}[y|\boldsymbol{x} = \boldsymbol{z}]$ for the logistic regression models.

The coefficients can be estimated using a parametric method such as the maximum likelihood estimation. However, the parametric estimation method implies some functional form assumptions that are restrictive in practical applications. Hyslop (1999) indicated that the parametric models often perform poorly in binary choice context compared to nonlinear models such as probit or logit. Thus nonparametric estimation, such as local likelihood logit, might be better for binary dependent variables. Tibshirani and Hastie (1987) and Fan *et al.* (1995) showed exemplary biometric applications of local likelihood logit for one-dimensional $\boldsymbol{x}$. Fan *et al.* (1998) did further research and presented a good performance of the local likelihood logit estimator by a simulation study. However, they only considered one-dimensional $\boldsymbol{x}$. In this chapter, we generalize their models from one-dimensional x to higher-dimensional $\boldsymbol{x}$, which contain both continuous and discrete variables. In particular, we use the linear B-spline to estimate the coefficients of logistic regression models and the conditional mean.

For a logistic regression model, the local likelihood logit estimator is

$$\hat{E}[y|\boldsymbol{x} = \boldsymbol{z}] = \frac{1}{1 + e^{-\boldsymbol{z}'\hat{\boldsymbol{\theta}}}} \tag{4.1}$$

where

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \sum_{i=1}^{n} \left\{ y_i \ln\left(\frac{1}{1 + e^{-\boldsymbol{x}_i'\boldsymbol{\theta}}}\right) + (1 - y_i) \ln\left(\frac{1}{1 + e^{\boldsymbol{x}_i'\boldsymbol{\theta}}}\right) \right\} \boldsymbol{P_B} y_i \tag{4.2}$$

with $\boldsymbol{P_B} = \boldsymbol{B}(\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'$ being the spline smoother defined in equation (3.4).

4.2 Simulation Studies

In this section, we will conduct simulations to illustrate the performance of the proposed local likelihood logit estimator in the logistic regression model.

The simulation studies are designed with ten covariates where the first four variables are continuous and the other six variables are binary; that is, we have covariates $\boldsymbol{x} = (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10})'$, where x_1, x_2, x_3, x_4 are continuous and the others are binary. The outcome variable y is binary with a value of either 0 and 1. The true value of the model coefficients is $\boldsymbol{\theta} = (-1, 2, -3, 4, -2, 2, -2, 2, -2, 2)'$. The four continuous variables x_1, x_2, x_3, x_4 are generated from different χ^2 distributions. The six binary variables are generated from Bernoulli distributions. We consider three different designs (see Frölich (2006)) for covariates. In these designs for the studies, the dependent variable y is simulated in the same fashion and the covariates are different. The observations y are generated as follows:

$$y = \begin{cases} 1, & x^* > 0, \\ 0, & \text{otherwise.} \end{cases}$$

where $x^* = \boldsymbol{x}'\boldsymbol{\theta} - 8 + \epsilon > 0$, ϵ is error generated from a logistic distribution.

Next, three designs for the covariates are specified. In the first design (Design 1), the continuous variables are independent and binary variables are independent as well. The continuous variables are independently generated from χ^2 distributions with degrees of freedom 1, 2, 3, 4, respectively. The binary variables are generated from Bernoulli distributions with $p = 0.5$. All the variables are independent of each other. In the second design (Design 2), the continuous variables are generated in the same way as in this first design; the six binary variables are correlated and are generated as follows: x_5 is generated from $Ber(p = 0.5)$ and x_k is generated from $Ber(p = 0.3 + 0.4\bar{x}_{k-1})$, where k is integer and $6 \leq k \leq 10$, $\bar{x}_{k-1} = \frac{1}{k-5}\sum_{j=5}^{k-1} x_j$ is the mean of generated values of the previous variables. In the third design (Design 3), the binary variables are generated in the same way as in the first design. The continuous variables are correlated and are generated as follows: x_1 is generated from χ^2 with degree of freedom 1, x_2 is generated as x_1 plus an independent χ^2 with degree of freedom 1, x_3 is generated as x_2 plus an independent χ^2 with degree of freedom 1, x_4 is generated as x_3 plus an independent χ^2 with degree of freedom 1.

The number of simulations is 100. The simulations are conducted for sample sizes 200, 400, 800. First, the data samples $\{(y_i, \boldsymbol{x}_i)\}_{i=1}^n$ are generated according to three designs of the covariates. The model coefficients are estimated based on these samples using the linear B-spline method. Second, validation samples $\{\boldsymbol{x}_j\}_{j=1}^n$ are

Table 4.1: Average Prediction Performance for Conditional Mean

Design	Size	MAE	MSE
1	200	0.166	0.119
	400	0.143	0.090
	800	0.134	0.076
2	200	0.146	0.090
	400	0.128	0.079
	800	0.123	0.068
3	200	0.163	0.083
	400	0.151	0.071
	800	0.147	0.070

generated. The validation samples and data sample are generated from the same population. Third, the conditional mean is predicted at each point of validation samples on the basis of the estimates of model coefficients using local likelihood logit method and the predicted value is compared to the true conditional mean. Thus the out-of-sample prediction performance is checked for the conditional mean. The out-of-sample prediction performance for the conditional mean is assessed by the mean absolute error (MAE) and the mean squared error (MSE). The performance is measured relative to the true value of the conditional mean. With different designs and sample sizes, the values of MAE and MSE are summarized in Table 4.1. It is worth pointing out that when sample sizes increases from 200 to 800, the values of MAE and MSE decrease.

4.3 Concluding Remarks

In this chapter, we concentrate on the nonparametric estimation of the conditional mean in the logistic regression models. We propose the local likelihood logit estimation with linear B-spline to estimate the conditional mean. We conduct the simulation studies to illustrate the proposed method. It shows that our method works well. It would be interesting yet challenging to reduce the values of MAE and MSE by improving the estimation method and extend the estimation method to more complicated models, which will be the focus of the future research.

References

Anderson, P. L. and Meerschaert, M. M. (1998). Modeling river flows with heavy tails. *Water Resources Research* 34, 2271–2280.

Anderson, P. L. and Meerschaert, M. M. (2005). Parameter estimation for periodically stationary time series. *Journal of Time Series Analysis* 26, 489–518.

Anderson, P. L., Meerschaert, M. M., and Zhang, K. (2013). Forecasting for periodic ARMA models. *Journal of Time Series Analysis* 34, 187–193.

Basawa, I. V. and Lund, R. (2001). Large sample properties of parameter estimates for periodic ARMA models. *Journal of Time Series Analysis* 22, 651–663

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.

Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (2008). *Time Series Analysis. Forecasting and Control*, Fourth Edition. Wiley, Hoboken, NJ.

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

Brockwell, P. J. and Davis, R. A. (2002). *Introduction to Time Series and Forecasting*, Second Edition. Springer, New York.

Chan, N. H. (2010). *Time Series: Applications to Finance with R and S-Plus*, Second Edition. Wiley, Hoboken, NJ.

Davis, R. A. and Dunsmuir, W. T. (1997). Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability* 10, 481–497.

Engle, R., Granger, C., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American Statistical Association* 81, 310–320.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modeling and Its Applications.* Chapman & Hall, Boca Raton, London, and New York.

Fan, J., Farmen, M., and Gijbels, I. (1998). Local maximum likelihood estimation and inference. *Journal of Royal Statistical Society Series B* 60, 591–608.

Fan, J., Heckman, N., and Wand, M. (1995). Local Polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association* 90, 141–150.

Ferreira, E., Núñez-Antón, V., and Rodríguez-Póo, J. (2000). Semiparametric approaches to signal extraction problems in economic time series. *Computational Statistics and Data Analysis* 33, 315–333.

Frölich, M. (2006). Non-parametric regression for binary dependent variables. *Journal of Econometrics* 9, 511–540.

Gao, J. T. (1995). Asymptotic theory for partly linear models. *Communications in Statistics: Theory Methods* 24, 1985–2009.

Hansen, J., Johnson, D., Lacis, A., Lebedeff, S., Lee, P., Rind, D., and Russell, G. (1981). Climate impact of increasing atmospheric carbon dioxide. *Science* 213, 957–966.

Heckman, H. (1986). Spline smoothing in partly linear model. *Journal of Royal Statistical Society Series B* 48, 244–248.

Hipel, W. K. and McLeod, I. A. (1994). *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, Amsterdam.

Kohn, R., Ansley, C. F., and Wong, C. M. (1992). Nonparametric spline regression with autoregressive moving average errors. *Biometrika* 79, 335–346.

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *Annals of Statistics* 31, 1600–1635.

Huang J. and Yang L. (2004). Identification of nonlinear additive autoregressive models. *Journal of the Royal Statistical Society Series B* 66, 463–477.

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.

Hyslop, D. (1999). State dependence, serial correlation and heterogeneity in intertemporal labor force participation of married women. *Econometrica* 67, 1255–1294.

Jones, P. D., New M., Parker, D. E., Martin, S., and Rigor, I. G. (1999). Surface air temperature and its changes over the past 150 years. *Reviews of Geophysics* 37, 173–199.

Lund, R. B. (2006). A seasonal analysis of riverflow trends. *Journal of Statistical Computation and Simulation* 76, 397–405.

Lund, R. B. 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., Shao, Q., and Basawa, I. V. (2006). Parsimonious periodic time series modeling. *Australian & New Zealand Journal of Statistics* 48, 33–47.

McLeod, A. I. and Balcilar, M. (2011). pear: Package for Periodic Autoregression Analysis, R package. Available from: `http://cran.r-project.org/web/packages/pear/` [Accessed on 18 May 2011].

McLeod, A. I. (1994). Diagnostic checking of periodic autoregression. *Journal of Time Series Analysis* 15, 221–233.

Pagano, M. (1978). On periodic and multiple autoregressions. *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.

Qiu, D., Shao, Q., and Yang, L. (2013). Efficient inference for autoregressive coefficient in the presence of trend. *Journal of Multivariate Analysis* 114, 40–53.

R Core Team (2013). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. Available from: `http://www.R-project.org/`.

Seymour, L. (2002). Modeling plant growth using regression with periodically correlated errors. *Journal of Agricultural, Biological and Environmental Statistics* 7, 1–17.

Shao, Q. (2009). Seasonality analysis of time series in partial linear models. *Journal of Nonparametric Statistics* 21, 827–837.

Shao, Q. (2006). Mixture autoregressive periodic time series. *Statistics & Probability Letters* 76, 609–618.

Shao, Q. and Lund, R. B. (2004). Computation and characterization of autocorrelations and partial autocorrelations in periodic ARMA models. *Journal of Time Series Analysis* 25, 359–372.

Shao, Q. and Yang, L. (2013). Oracally efficient estimation of ARMA coefficients and consistent model selection in the presence of trend, manuscript.

Shao, Q. and Yang, L. (2012). Polynomial spline confidence bands for time series trend. *Journal of Statistical Planning and Inference* 142, 1678–1689.

Shao, Q. and Yang, L. (2011). Autoregressive coefficient estimation in nonparametric analysis. *Journal of Time Series Analysis* 32, 587–597.

Shumway, R. H. and Stoffer, D. S. (2011). *Time Series Analysis and Its Applications with R Examples*, Third Edition. Springer, New York.

Song, Q. and Yang, L. (2010). Oracally efficient spline smoothing of nonlinear additive autoregression model with simultaneous confidence band. *Journal of Multivariate Analysis* 101, 2008–2025.

Speckman, P. (1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society. Series B (Methodological)* 50, 413–436.

Tang, L. and Shao, Q. (2014). Efficient Estimation for Periodic Autoregressive Coefficients via Residuals. *Journal of Time Series Analysis* 35, 378–389.

Tesfaye, Y. G., Meerschaert, M. M., and Anderson, P. L. (2006). Identification of PARMA models and their application to the modeling of riverflows. *Water Resources Research* 42, W01419.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association* 82, 559–567.

Truong, Y. K. (1991). Nonparametric curve estimation with time series errors. *Journal of Statistical Planning and Inference* 28, 167–183.

TSAY, R. S. (2010). *Analysis of Financial Time Series*, Third Edition. Wiley, Hoboken, NJ.

Tsay, R. S. (1984). Regression models with time series errors. *Journal of the American Statistical Association* 79, 118–124.

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

Wang, L. and Yang, L. (2010). Simultaneous confidence bands for time series prediction function. *Journal of Nonparametric Statistics* 22, 999–1018.

Wang, L. and Yang, L. (2009). Spline estimation of single index model. *Statistica Sinica* 19, 765–783.

Wang, L. and Yang, L. (2007). Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Annals of Statistics* 35, 2474–2503.

Wang, Y. (1998). Smoothing spline models with correlated random errors. *Journal of the American Statistical Association* 93, 341–348.

Xue, L. and Yang, L. (2006). Additive coefficient modelling via polynomial spline. *Statistica Sinica* 16, 1423–1446.

You, J., and Chen, G. (2007). Semiparametric generalized least squares estimation in partially linear regression models with correlated errors. *Journal of Statistical Planning and Inference* 137, 117–132.

Appendix A

R Program for Nonparametric Efficient Estimation for Periodic Autoregressive Coefficients via Residuals

```
# R Program for Simulation #
library(pear)
S=100; n=200; T=4; loop=n/T
ro=matrix(c(0), nrow=S, ncol=T)
X=rep(0,n); Y=rep(0,n); r=rep(0,T); v=rep(0,T)
err=rep(0,n); N=n^(1/5); M=ceiling(N)+2
h=1/(ceiling(N)+1); B=matrix(c(0), nrow=n, ncol=M)
phihat=matrix(c(0),nrow=S,ncol=T)
phitld=matrix(c(0),nrow=S,ncol=T)
ghat=rep(0,n); Xhat=rep(0,n)
avphi=rep(0,T); sdphi=rep(0,T)
r[1]=0.2; r[2]=0.4; r[3]=0.6; r[4]=0.9
```

```
v[1]=0.5; v[2]=0.7; v[3]=0.85; v[4]=1
t=0; t1=0; X0=0; u=0
for(k in 1:S)
{ err=rnorm(n,0,1)
  for(i in 1:loop)
  { for(j in 1:T)
    { t=(i-1)*T+j
      t1=t-1
      if(t==1)
        X[1]=r[1]*X0+sqrt(v[1])*err[t]
      else
        X[t]=r[j]*X[t1]+sqrt(v[j])*err[t]
    }
  }
  for(i in 1:n)
  { Y[i]=X[i]+sin(2*pi*i/n) }
  for(q in 1:(M-1))
  { for(i in 1:n)
    { u=i/n
      if(u>=(q-2)*h & u<(q-1)*h){B[i,q]=u/h-(q-2)} else
      if(u>=(q-1)*h & u<q*h){B[i,q]=q-u/h} else
        {B[i,q]=0}
    }
  }
  for(i in 1:n)
  { u=i/n
    if(u>=(M-2)*h & u<=(M-1)*h){B[i,M]=u/h-(M-2)} else
```

```
    {B[i,M]=0}
  }
  ghat=B%*%solve(t(B)%*%B)%*%t(B)%*%Y
  Xhat=Y-ghat
  z=ts(data=Xhat, start=c(1,1), end=c(n/4,4), frequency=4)
  a1=pear(z, 1, ic="none")
  phihat[k,]=a1$phi
  z1=ts(data=X, start=c(1,1), end=c(n/4,4), frequency=4)
  a11=pear(z1, 1, ic="none")
  phitld[k,]=a11$phi
  }
for(j in 1:T)
{ avphi[j]=mean(phihat[,j])
  sdphi[j]=sd(phihat[,j])
}
for(j in 1:T)
{ ro[,j]=phitld[,j]/phihat[,j] }
par(mfrow=c(2,2))
boxplot(ro[,1], ro[,1], ro[,1], ro[,1], main='Season 1', ylim=c(0.91,1.32),
range=0.5, names=c(200,400,800,1600), xlab="Sample Size")
abline(h=1,lty=2)
boxplot(ro[,2], ro[,2], ro[,2], ro[,2], main='Season 2', ylim=c(0.885,1.305),
range=0.5, names=c(200,400,800,1600), xlab="Sample Size")
abline(h=1,lty=2)
boxplot(ro[,3], ro[,3], ro[,3], ro[,3], main='Season 3', ylim=c(0.91,1.30),
range=0.5, names=c(200,400,800,1600), xlab="Sample Size")
abline(h=1,lty=2)
```

```
boxplot(ro[,4], ro[,4], ro[,4], ro[,4], main='Season 4', ylim=c(0.91,1.3),
range=0.5, names=c(200,400,800,1600), xlab="Sample Size")
abline(h=1,lty=2)

# R Program for Application #
Fra=as.matrix(read.table("Fraser.txt",header=T)[2:100,])
Fra=as.matrix(read.table("Fraser.txt")[2:100,][,2:13])
Newfra=matrix(c(0),nrow=98,ncol=12)
Newfra1=matrix(c(0),nrow=98,ncol=4)
for(j in 1:98)
{ j1=j+1
  Newfra[j,]=Fra[j1,]-Fra[1,]
}
for(j in 1:98)
{ Newfra1[j,1]=(Newfra[j,1]+Newfra[j,2]+Newfra[j,3])/3
  Newfra1[j,2]=(Newfra[j,4]+Newfra[j,5]+Newfra[j,6])/3
  Newfra1[j,3]=(Newfra[j,7]+Newfra[j,8]+Newfra[j,9])/3
  Newfra1[j,4]=(Newfra[j,10]+Newfra[j,11]+Newfra[j,12])/3
}
n=392; T=4; Y=rep(0,n)
N=n^(1/5); M=ceiling(N)+2; h=1/(ceiling(N)+1)
B=matrix(c(0), nrow=n, ncol=M); t=0; u=0
phihat0=rep(0,T); ghat=rep(0,n); Xhat=rep(0,n)
for(i in 1:98)
{ for(j in 1:4)
  { t=(i-1)*T+j
    Y[t]=Newfra1[i,j]
  }
```

```
}
for(q in 1:(M-1))
{ for(i in 1:n)
  { u=i/n
    if(u>=(q-2)*h & u<(q-1)*h){B[i,q]=u/h-(q-2)} else
    if(u>=(q-1)*h & u<q*h){B[i,q]=q-u/h} else
      {B[i,q]=0}
  }
}
for(i in 1:n)
{ u=i/n
  if(u>=(M-2)*h & u<=(M-1)*h){B[i,M]=u/h-(M-2)} else
    {B[i,M]=0}
}
ghat=B%*%solve(t(B)%*%B)%*%t(B)%*%Y
Xhat=Y-ghat
z=ts(data=Xhat, start=c(1,1), end=c(n/4,4), frequency=4)
a0=pear(z, ic="bic")
phihat0=a0$phi
```

Appendix B

R Program for Semiparametric Estimation for Periodic Autoregressive Coefficients in Partially Linear Models

```
library(pear)
S=100; n=200; T=4; loop=n/T;
ro=matrix(c(0),nrow=S,ncol=T)
X=rep(0,n); Y=rep(0,n); r=rep(0,T);
v=rep(0,T); err=rep(0,n)
z=matrix(c(0),nrow=T-1, ncol=1);
bta=matrix(c(0),nrow=T-1,ncol=1)
N=n^(1/5); M=ceiling(N)+2; h=1/(ceiling(N)+1)
B=matrix(c(0),nrow=n,ncol=M)
XDesign=matrix(c(0),nrow=n,ncol=T-1)
pphihat=matrix(c(0),nrow=S,ncol=T)
pbta=matrix(c(0),nrow=S,ncol=T-1)
```

```
pphitld=matrix(c(0),nrow=S,ncol=T)
pghat=rep(0,n); pXhat=rep(0,n); E=diag(1,n)
avphi=rep(0,T); sdphi=rep(0,T)
avbta=rep(0,T-1); sdbta=rep(0,T-1)
r[1]=0.3; r[2]=0.6; r[3]=0.4; r[4]=0.2
v[1]=0.5; v[2]=0.7; v[3]=0.85; v[4]=1
t=0; t1=0; X0=0
z[1]=0.8; z[2]=1.5; z[3]=2.3
for(k in 1:S)
{ err=rnorm(n,0,1)
  for(i in 1:loop)
  { for(j in 1:T)
    { t=(i-1)*T+j
      t1=t-1
      if(t==1)
        X[1]=r[1]*X0+sqrt(v[1])*err[t]
      else
        X[t]=r[j]*X[t1]+sqrt(v[j])*err[t]
    }
  }
  t1=0; t2=0; t3=0; t4=0
  for(i in 1:loop)
  { t1=(i-1)*T+1
    t2=(i-1)*T+2
    t3=(i-1)*T+3
    t4=(i-1)*T+4
    XDesign[t1,]=c(1,0,0)
```

```
  XDesign[t2,]=c(0,1,0)
  XDesign[t3,]=c(0,0,1)
  XDesign[t4,]=c(0,0,0)
}
pl=XDesign%*%z
for(i in 1:n)
{ Y[i]=pl[i]+X[i]+sin(2*pi*i/n) }
for(q in 1:(M-1))
{ for(i in 1:n)
  { u=i/n
    if(u>=(q-2)*h & u<(q-1)*h){B[i,q]=u/h-(q-2)} else
    if(u>=(q-1)*h & u<q*h){B[i,q]=q-u/h} else
      {B[i,q]=0}
  }
}
for(i in 1:n)
{ u=i/n
  if(u>=(M-2)*h & u<=(M-1)*h){B[i,M]=u/h-(M-2)} else
    {B[i,M]=0}
}
promat=B%*%solve(t(B)%*%B)%*%t(B)
bta=solve(t(XDesign)%*%(E-promat)%*%XDesign)%*%t(XDesign)
    %*%(E-promat)%*%Y
pghat=promat%*%(Y-XDesign%*%bta)
pXhat=Y-pghat81-XDesign%*%bta
pz=ts(data=pXhat, start=c(1,1), end=c(n/4,4), frequency=4)
pa=pear(pz, 1, ic="none")
```

```
  pphihat[k,]=pa$phi
  pz1=ts(data=X, start=c(1,1), end=c(n/4,4), frequency=4)
  pa1=pear(pz1, 1, ic="none")
  pphitld[k,]=pa1$phi
  pbta[k,1]=bta[1,1]; pbta[k,2]=bta[2,1]; pbta[k,3]=bta[3,1]
}
for(j in 1:T)
{ avphi[j]=mean(pphihat[,j])
  sdphi[j]=sd(pphihat[,j])
}
for(j in 1:T)
{ ro[,j]=pphitld[,j]/pphihat[,j] }
avbta=apply(pbta, 2, mean)
sdbta=apply(pbta, 2, sd)
```

Appendix C

R Program for Future Research

```
Y=rep(0,n); ln1=rep(0,n)
ln2=rep(0,n); spl=rep(0,n)
f1=rep(0,n); f2=rep(0,n)
spl=B%*%solve(t(B)%*%B)%*%t(B)%*%Y
fr<-function(th){
    th1=th[1]
    th2=th[2]
    th3=th[3]
    th4=th[4]
    th5=th[6]
    th6=th[6]
    th7=th[7]
    th8=th[8]
    th9=th[9]
    th10=th[10]
    theta=c(th1,th2,th3,th4,th5,th6,th7,th8,th9,th10)
    for(i in 1:n)
     {
```

```
      f1[i]=1/(1+exp(-sum(X[i,]*theta)))
      ln1[i]=log(f1[i], base=exp(1))
      f2[i]=1/(1+exp(sum(X[i,]*theta)))
      ln2[i]=log(f2[i], base=exp(1))
    }
   sum(Y*ln1*spl+(1-Y)*ln2*spl)
 }

# Design 1 #
n=200
for(j in 1:10)
 { if(j<=cont)
     X[,j]=rchisq(n, df=j)
   else
     X[,j]=rbinom(n,1,0.5)
 }
lgnoise=rlogis(n, location =0, scale =1); t=0
for(i in 1:n)
 { t=sum(YDes1*X[i,])+lgnoise[i]-8
   if(t>0)
     Y[i]=1
   else
     Y[i]=0
 }

# Design 2 #
for(i in 1:n)
```

```
 { X[i,1]=rchisq(1, df=1)
   X[i,2]=rchisq(1, df=2)
   X[i,3]=rchisq(1, df=3)
   X[i,4]=rchisq(1, df=4)
   X[i,5]=rbinom(1,1,0.5)
   p6=0.3+0.4*X[i,5]; X[i,6]=rbinom(1,1,p6)
   p7=0.3+0.4*(X[i,5]+X[i,6])/2; X[i,7]=rbinom(1,1,p7)
   p8=0.3+0.4*(X[i,5]+X[i,6]+X[i,7])/3
   X[i,8]=rbinom(1,1,p8)
   p9=0.3+0.4*(X[i,5]+X[i,6]+X[i,7]+X[i,8])/4
   X[i,9]=rbinom(1,1,p9)
   p10=0.3+0.4*(X[i,5]+X[i,6]+X[i,7]+X[i,8]+X[i,9])/5
   X[i,10]=rbinom(1,1,p10)
   t=sum(YDes1*X[i,])+lgnoise[i]-8
   if(t>0)
     Y[i]=1
   else
     Y[i]=0
 }

# Design 3 #
for(i in 1:n)
 { X[i,1]=rchisq(1, df=1)
   X[i,2]=X[i,1]+rchisq(1, df=1)
   X[i,3]=X[i,2]+rchisq(1, df=1)
   X[i,4]=X[i,3]+rchisq(1, df=1)
   X[i,5]=rbinom(1,1,0.5)
```

```
  X[i,6]=rbinom(1,1,0.5)
  X[i,7]=rbinom(1,1,0.5)
  X[i,8]=rbinom(1,1,0.5)
  X[i,9]=rbinom(1,1,0.5)
  X[i,10]=rbinom(1,1,0.5)
  t=sum(YDes1*X[i,])+lgnoise[i]-8
  if(t>0)
    Y[i]=1
  else
    Y[i]=0
}
```