Department of Mathematics

ARMA Model for Random Periodic Processes

by

Yujia Liu

A Doctoral Thesis

Submitted in partial fulfilment of the requirements for the award of

Doctor of Philosophy at Loughborough University

Abstract

In this thesis, we construct ARMA model for random periodic processes. We stress on the mixed periodicity and randomness of the model and redefined the definition of sample autocovariance function. We prove the asymptotic normality of Yule-Walker estimation and innovation estimation for coefficients in causal and invertible case. We also prove the central limit theorem for random periodic processes. Under this and ergodic theorem, we prove the asymptotic normality of maximum likelihood estimation for non-causal autoregressive model for random periodic processes. We simulate ARMA model for random periodic processes to two examples and compare the results with classical ARMA model.

Keywords: random periodic processes, ARMA model, central limit theorem, asymptotic normality, causal case, non-causal case, simulation.

Acknowledgements

First of all, I would like to express my sincerest gratitude to my supervisors, Prof. Huaizhong Zhao and Dr. Chunrong Feng. They guided me into the research in this following area. They helped and supported me continuously over the years. Their deep academic insights and hard-working spirit inspire me tremendously in my research and my life. I will always appreciate their encouragement and help.

I would like to thank my colleagues, Yu Liu, Johnny Zhong, Yan Luo and Baoyou Qu. We learned together and benefitted a lot from each other. Thanks for all the useful discussions and help.

I would like to show my thanks to all administrative and support staff at the School of Science. I have had a comfort and effective experience in the School during my PhD years.

Finally I would like to give my warm gratitude to my parents for their unconditional support and love.

Contents

Introduction

Periodic phenomena appear very often in our daily life such as the temperature cycle, business cycles, astronomical landscapes etc. People built many different kinds of models to capture periodic trends in order to help forecasting the future events and guiding social activities and material production. But real world systems are influenced by many internal or external fluctuations with uncertainty. The mixture of periodicity and randomness could be in a complex way and make forecasting much harder, especially when noise is very large and periodicity is covered up.

So far in time series analysis, the fundamental non-structural method to analyse a time series is to describe the observations as a path of a stationary process using different non-structural models to fit the data, e.g. the autoregressive moving average (ARMA) model. The classical ARMA model was described first in 1951 by Whittle in his thesis [27], and was popularized by Box and Jenkins in 1970 in [9]. It describes the time series as a stationary stochastic process which combines linearly with an autoregressive polynomial relating the current data to p past data and a moving average polynomial relating to history noises of lag q. For non-stationary time series, de-trended and de-seasoned preprocessing is need. Consider the well-used classical decomposition model of data [12],

$$X(t,\omega) = trend(t) + s(t) + \tilde{X}(t,\omega), \tag{0.0.1}$$

which describes the observed data $\{X(t,\omega)\}$ by a linear combination of a trend term, a noise term $\tilde{X}(t,\omega)$, which is a stationary process, and sometimes with a seasonal fluctuation component $s(t)$. Then ARMA model is applied to fit the process $\{\tilde{X}(t)\}$.

However this method to model seasonal fluctuations is quite a limited way due to the inflexible assumption of seasonal component. Recent work in Franses' book [21] recommends that the seasonal fluctuations should be paid more attention in econometric studies. One reason is that randomness in seasonal variation can explain more precisely the behaviour of economic agents, and the other is that in many de-trended time series the seasonal fluctuations take a dominate position in the remaining variation in the series. Franses also mentioned that the non-seasonal fluctuations in many quarterly or monthly observed macroeconomic time series do not appear to be a stationary process over time and that the dependence between seasonal fluctuations and non-seasonal fluctuations may well exist in some time series. This suggests the limitation of the classical decomposition model in the classical time series analysis.

In this thesis, we also provide an example of a sample set simulated by a SDE whose solution is random periodic path. The drift term and the diffusion term are both periodic functions of t. The seasonal fluctuations and randomness are mixed together. Here the decomposition model (0.0.1) is not correct.

To solve the varying volatility problem in many real cases and to improve the model fitness, many sophisticated models are built based on the ARMA model.

Gladyshev [22] defined a class of stochastic processes called periodically correlated random processes, whose correlation function satisfies $B(s,t) = B(s+T, t+T)$ for some fixed number T. Later, some researches constructed periodic ARMA model based on periodically correlated random processes. The PARMA model assumes the coefficients of the standard ARMA model are varying with the season. There are several papers stated recent work on PARMA model. For instance, Vecchia studied the maximum likelihood estimation for periodic autoregressive model in [26]. Anderson *et al.* [4] studied the innovation algorithm asymptotics for causal PARMA model.

Another commonly used model in financial time series is the autoregressive conditionally heteroscedastic (ARCH) model. This is a stationary non-linear model for the data set. An ARCH(q) model first model the stationary process by an AR(q) model, and describes the variance of the residual term as a qth autoregressive polynomial relating to the history squared residuals back to lag q. Engle [17] first proposed this model in 1982, Bollerslev [8] and Taylor [25] independently generalised Engle's model in 1986 to make it more realistic, termed as generalized ARCH (GARCH) model. GARCH model describes the variance of the residual as an ARMA-type process which relates to not only the history squared residuals but also their variances.

These models have greatly improved the range of application to real world problems in time series. However for periodic phenomena, the estimation of randomness and periodicity is still limited and considered separately, for example in GARCH model, it only fits a periodic function such as Fourier series to the seasonal trend and adds it in the ARMA-type process of residuals linearly. The stochastic process presented in PARMA model may combine randomness and periodicity together, but it lacks of specific mathematical definition and description of random periodicity.

The concept of stationary solutions is the stochastic counterpart corresponding to the fixed points of deterministic dynamical systems. A fixed point is the simplest equilibrium and large time limiting set of a deterministic dynamical system. A periodic solution is a more complicated limiting set. Zhao and Zheng [28] first introduced the pathwise random periodic paths of random dynamical systems in 2009. Later, Feng, Zhao and Zhou [19], Feng and Zhao [20] established the concept of random periodicity for semi-flows of random dynamical systems. In [20], Feng and Zhao studied the periodic measure which describes random periodicity in the sense of distribution. They proved that the ergodic random periodic path and periodic measure are "equivalent" in some sense. They also obtained for the first time the ergodicity of periodic measure of the transition probability semigroup of Markovian random dynamical system and proved that random periodic processes satisfy the strong law of large numbers. This result suggests that the periodic measure gives a statistical description of random periodic processes in a long run.

As the concept of random periodic processes describes the randomness and periodicity in the revolution of the stochastic process simultaneously, it inspires us to apply random periodic process to classical time series analysis to help describing complex periodic phenomena.

The autocovariance of random periodic processes not only depends on the distance between two times, but also on the time itself. We redefine the sample covariance function as (2.3.2), which is estimated by averaging a moving window of product of history data. We stress on the mixed periodicity and randomness in the sample autocovariance. This definition is different from the one of the PARMA model, which is assumed by statistical experience.

We prove that the coefficients of the ARMA model for random periodic processes are deterministic periodic functions of time t from the periodicity of the autocovariance function of random periodic processes. This is consistent with the PARMA model.

For causal and invertible case, as the process of the same time point in each period can be considered as a sequence of stationary process, we obtain the convergence of Yule-Walker estimators at each time point for random periodic process model by using the idea in [12] for stationary process. In the proof for the convergence of Innovation estimators, we follow the steps in [11] and consider the innovation coefficients in a period as a vector. This idea is also used in [4] for the PARMA model.

For non-causal autoregressive case, we first prove the central limit theorem for random periodic processes. We then obtain the asymptotic normality of the coefficients for non-causal autoregressive model for random periodic processes using the ergodic theorem and the central limit theorem for random periodic processes.

We approach the ARMA model for random periodic processes from the ergodicity point of view. This is different from PARMA model. It is suggested that we can not only simulate the sample set, but also estimate periodic measure of data based on relative theorems of random periodic processes in [20]. We aim to study the method to estimate periodic measure for real world problem in the future.

In a ARMA(p, q) model, good estimation result can be obtained with relative small values of p, q. In fact, bigger values than the theoretical ones of p, q will cause overfitting problem. To specify suitable values for p, q, we follow the criteria for model selection stated in [12] in this thesis, i.e. AIC, AICc and BIC.

Nowadays, many software programmes such as R, Matlab, Maple etc have corresponding statistical packages which contain functions to realize ARMA model. We programme the ARMA model for random periodic processes with language R throughout this thesis and attach the main codes in the Appendix.

In Chapter 1, we introduce the classical ARMA model and some important properties and theorems which we will use later. Then we introduce the concept of random periodic processes and properties relevant to this thesis. We also introduce the method to eliminate the trend and seasonal components. In Chapter 2, we construct the ARMA model for random periodic processes. In Chapter 3, we deduce the asymptotic behaviour of the coefficients and obtain the algorithm of estimation for causal and invertible ARMA model.

In Chapter 4, we discuss some properties about non-causal autoregressive model for random periodic processes. We hope we will study non-invertible moving-average model for random periodic processes later. In Chapter 5, we introduce the model criteria and list the procedure to obtain a suitable model for a time series. Finally we give two simulation examples in Chapter 5.

Chapter 1

Background

1.1 The Elimination of Trend and Seasonal Component

Almost all time series in the real world are non-stationary, which is difficult to simulate and study. The main idea of the classical decomposition model (0.0.1) is to separate a stationary process from the original series. First we note that the classical decomposition model assumes time series is separable to three independent parts. However this assumption may not be suitable to all observed data. Trend component is defined as a slowly changing function of time t, which will obviously make time series non-stationary. Seasonal fluctuations describe the similar pattern cycle to cycle with approximate period τ, which is also non-stationary.

Two main approaches to eliminate the trend and seasonal components are introduced in the monograph of Brockwell and Davis [12]. One is to estimate the independent trend and seasonal parts so that $\tilde{X}(t) := X(t) - trend(t) - s(t)$ is stationary. The other one is to apply differencing operator to the observed data until the differenced observations resemble a realization of some stationary process.

First we assume the seasonal fluctuations are absent. Without loss of generality, we assume that the stationary part $\tilde{X}(t)$ has mean 0.

Method 1. ***(Least Squares Estimation)*** *Assume the trend is a second order function* *of* t,

$$trend(t) = a_0 + a_1 t + a_2 t^2, \tag{1.1.1}$$

and minimize the residual $\sum_t (x(t) - trend(t))^2$ *by choosing appropriate parameters.*

This is the most well-used method in many kinds of fitting examples. It requires that in application the trend is less possible a higher than second order function of t.

Method 2. ***(Smoothing by Means of Moving Average)***
For $q \in \mathbb{Z}$, $q \geq 0$, *assume*

(i) $trend(t)$ *is approximately linear over* $[t-q, t+q]$,

(ii) the average of the residual is close to 0.

Then the estimate of trend of two-sided moving average is defined as

$$\hat{trend}(t) = \frac{1}{2q+1} \sum_{i=-q}^{q} x(t+i),$$

one-sided moving average is as

$$\hat{trend}(t) = ax(t) + (1-a)\hat{trend}(t-1) = \sum_{j=0}^{t-2} a(1-a)^j x(t-j) + (1-a)^{t-1} x(1),$$

for $t = 2, \cdots, n$, *and*

$$\hat{trend}(1) = x(1).$$

This kind of estimation uses the corrected term to estimate next term, which will attenuate the effect of the noise term. We should note that if the trend term is not linear with time t, the two-sided moving average will automatically not be accurate for estimation.

If we regard the coefficients in the moving average as weights, then we could obtain the general equation of moving average estimation method. Define the weighted moving average trend as,

$$\hat{trend}(t) = \sum_{j=-\infty}^{\infty} a_j x(t+j).$$

where $\sum_{j=-\infty}^{\infty} a_j = 1$. If we choose the weights carefully, this moving average may describe high order polynomial-type trend undisortedly.

Method 3. ***(Differencing to Generate Stationary Data)*** *Define the differencing operator* $\triangledown$ *as* $\triangledown X(t) = X(t) - X(t-1) = (1-B)X(t)$*, where* B *presents the backward shift operator* $BX(t) = X(t-1)$*. The main idea of this method is to apply the differencing operator* $\triangledown$ *on the series repeatedly until the remaining data appears a stationary pattern. For example, if* $trend(t)$ *has the form* $trend(t) = \sum_{j=0}^{k} a_j t^j$ *in* $X(t) = trend(t) + \hat{X}(t)$*, we could apply* $\triangledown^k$ *to the series to obtain* $\triangledown^k X(t) = k!a_k + \triangledown^k \hat{X}(t)$*, a stationary process with mean* $k!a_k$*. In general,* k *only needs to be quite small, frequently 1 or 2.*

Generally, these methods all use polynomials to estimate the trend component.

Now we turn to the general decomposition equation $X(t) = trend(t) + s(t) + \hat{X}(t)$ containing the seasonal fluctuations with period τ. First assume

$$s(t+\tau) = s(t), \sum_{j=1}^{\tau} s(t+j) = 0 \text{ for all } t \in T.$$

This is a very strong assumption for seasonality. It requires the seasonal fluctuations to be the same cycle to cycle and the sum of the fluctuations in one period is zero. In other word this assumption omits the possible randomness in season fluctuations and limits the randomness of the original time series only in the stationary noise. It is very hard to realize this condition in the practical condition. Here there is a possibly fundamental breakthrough for us to improve the classical models. We will discuss it further in later chapters.

In order to distinguish season and cycle, it is convenient to index the data as $x_{j,k}$, where the first index j represents the j^{th} cycle, and the second index k represents the k^{th} season.

Similar to the methods estimating only trend term, there are methods for both trend and seasonal components as follows.

Method 4. ***(Small trend)*** *Assume that there are* N *cycles of a data sample and the period of the seasonal component is* τ*. For some small trend time series, consider the trend in each cycle as a constant. Then the estimation of trend is*

$$\hat{trend}(j) = \frac{1}{\tau} \sum_{k=1}^{\tau} x_{j,k},$$

since $\sum_{k=1}^{\tau} s(k) = 0$. *Then the seasonal component could be estimated as*

$$\hat{s}(k) = \frac{1}{N}\sum_{j=1}^{N}(x_{j,k} - \hat{trend}(j)) = \hat{s}(k \pm \tau),$$

and the estimated error term is

$$\tilde{x}_{j,k} = x_{j,k} - \hat{trend}(j) - \hat{s}(k).$$

Method 5. ***(Moving Average Estimation)*** *First we use a moving average method to estimate the trend component for each data. If* $\tau = 2q$,

$$\tilde{trend}(t) = \frac{1}{\tau}(0.5x(t-q) + x(t-q+1) + \ldots + x(t+q-1) + 0.5x(t+q)),$$

since $\sum_{k=1}^{\tau} s(k) = 0$. *If* $\tau = 2q+1$,

$$\tilde{trend}(t) = \frac{1}{\tau}\sum_{k=-q}^{q} x(t+k).$$

Then we estimate the k^{th}*-season component,*

$$w_k = \frac{1}{N}\sum_{j=1}^{N}\left(x_{j,k} - \tilde{trend}(k+(j-1)\tau)\right), \qquad \hat{s}(k) = w_k - \frac{1}{\tau}\sum_{i=1}^{\tau} w_i.$$

The deseasonalized data $d_t = x_t - \hat{s}(t)$ *is supposed to contain only the trend component and noise term. In the last step we apply the methods introduced before to the deseasonalized data* $\{d_t\}$ *to re-estimate the trend component, denoted by* $\hat{trend}(t)$. *Therefore we could get the estimation for the stationary component*

$$\hat{x}(t) = x(t) - \hat{trend}(t) - \hat{s}(t).$$

Re-estimation for the trend component could make the influence of seasonal fluctuations less significant in the first estimation.

Method 6. ***(Differencing at Lag*** τ***)*** *Since we first assume that* $s(t+\tau) = s(t)$, *after applying the differencing operator of lag* τ, *the differenced data will contain only the trend and noise part,*

$$\nabla_\tau X(t) = (1-B^\tau)X(t) = X(t)-X(t-\tau) = (trend(t) - trend(t-\tau)) + \left(\hat{X}(t) - \hat{X}(t-\tau)\right),$$

then we apply differencing operator to the deseasonalized data to eliminate trend component as introduced before.

1.2 The Autoregressive Moving Average Model

Definition 1.2.1. ***(White Noise)*** *A white noise is a process $\{Z(t)\}$ with zero mean and autocovariance function*

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

written as $\{Z(t)\} \sim WN(0, \sigma^2)$.

Definition 1.2.2. ***(The ARMA(p, q) Process)*** *The process $\{X(t), t = 0, \pm 1, \pm 2, \cdots\}$ is said to be an ARMA(p, q) process if $\{X(t)\}$ is stationary and for each t,*

$$X(t) - \phi_1 X(t-1) - \cdots - \phi_p X(t-p) = Z(t) + \theta_1 Z(t-1) + \cdots + \theta_q Z(t-q)$$

where $\{Z(t)\} \sim WN(0, \sigma^2)$.
If $\{X(t) - \mu\}$ is an ARMA(p, q) process, then $\{X(t)\}$ is an ARMA(p, q) process with mean μ.

Definition 1.2.3. ***(Sample Autocovariance Function and Sample autucorrelation of data*** $\{x(1), \cdots, x(n)\}$***)*** *The sample autocovariance function of data $\{x(1), \cdots, x(n)\}$ is defined as*

$$\hat{\gamma}(h) := \frac{1}{n} \sum_{j=1}^{n-h} (x(j+h) - \bar{x})(x(j) - \bar{x}), 0 \leq h \leq n.$$

where $\bar{x}$ is the sample mean of the observed data $\bar{x} = \frac{1}{n} \sum_{j=1}^{n} x(j)$.
The sample autocorrelation function is defined as

$$\hat{\rho}(h) := \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}, |h| < n$$

Another characteristic indicator of ARMA process is the partial autocorrelation function.

Definition 1.2.4. *Denote span$\{1, X(1)X(2), \ldots, X(k)\}$ to be the closed span of the finite set $\{1, X(1)X(2), \ldots, X(k)\}$ and $P_{span\{1, X(1), X(2), \ldots, X(k)\}} x$ to be the unique projection of x onto span$\{1, X(1), X(2), \ldots, X(k)\}$ such that for any $X(j), j = 1, 2, \cdots, k$,*

$$\langle P_{span\{1, X(1), X(2), \ldots, X(k)\}} x, X(j) \rangle = \langle x, X(j) \rangle.$$

The partial autocorrelation function of ARMA process $\{X(t)\}$ is defined as

$$\alpha(1) = Corr(X(2), X(1)) = \rho(1)$$

and

$$\alpha(k) = Corr(X(k+1) - P_{span\{1,X(2),\ldots,X(k)\}}X(k+1), X(1) - P_{span\{1,X(2),\cdots,X(k)\}}X(1)), k \geq 2$$

Remark 1.2.5. *(i) The autocorrelation function conveys the dependence structure of a stationary process. $\alpha(k)$ is the correlation of the two residuals obtained after regressing $X(k+1)$ and $X(1)$ on the intermediate observations $X(2), \cdots, X(k)$.*

(ii) $\alpha(k)$ only depends on the second order properties of the process.

(iii) If the stationary process has zero mean, then

$$P_{span\{1,X(2),\cdots,X(k)\}}(\cdot) = P_{span\{X(2),\cdots,X(k)\}}(\cdot)$$

(iv) The partial autocorrelation of an AR(p) process vanishes for large lags.

(v) The partial autocorrelation of an MA(q) process is bounded in absolute value by a geometerically decreasing function.

The partial autocorrelation function of ARMA(p, q) model is a useful tool to preliminary estimate the range of the value of (p, q) together with the autocorrelation function. Their plots may appear like the following cases:

i) If the autocorrelation function plot cuts off at n^{th} lag, it is possibly a moving-average series with $q = n$.

ii) If the autocorrelation plot goes down gradually without any cut off value but the partial autocorrelation function drops sharply after n^{th} lag, then it may be an autoregressive series with $p = n$.

iii) If both of them decreases gradually, it indicates that the autoregressive and moving-average parts may both exist, or we need to check whether the series $\{\hat{X}(t)\}$ are still non-stationary.

Specifically, for an AR(1) process, the sample autocorrelation function should have an exponentially decreasing appearance. However, higher-order autoregressive processes are often a mixture of exponentially decreasing and damped sinusoidal components.

1.3 SARIMA and pARMA Models

Definition 1.3.1. *(The ARIMA(p,d,q) Process) [12]. If d is a non-negative integer, then $\{X(t)\}$ is said to be an ARIMA(p,d,q) process if $Y(t) := (1-B)^d X(t)$ is a causal ARMA(p,q) process.*

The process $\{X(t)\}$ is stationary if and only if $d = 0$.

Definition 1.3.2. *(The SARIMA$(p,d,q)\times(P,D,Q)_s$ Process) [12]. If d and D are non-negative integers, then $\{X(t)\}$ is said to be a seasonal ARIMA$(p,d,q)\times(P,D,Q)_s$ process with period s if the differenced process $Y(t) := (1-B)^d(1-B^s)^D X(t)$ is a causal ARMA process,*

$$\phi(B)\Phi(B^s)Y(t) = \theta(B)\Theta(B^s)Z(t), \quad \{Z(t)\} \sim WN(0,\sigma^2),$$

where $\phi(z) = 1-\phi_1 z - \cdots - \phi_p z^p$, $\Phi(z) = 1-\Phi_1 z - \cdots - \Phi_P z^P$, $\theta(z) = 1+\theta_1 z + \cdots + \theta_q z^q$ and $\Theta(z) = 1+\Theta_1 z + \cdots + \Theta_Q z^Q$.

Definition 1.3.3. *(Periodically Correlated Process) [22]. We shall call a random process $X(t)$, $-\infty < t < \infty$, periodically correlated with period T if its correlation function $B(s,t) = \boldsymbol{M}X(s)\overline{X(t)}$ exists, is continuous, and for any s and t satisfies the condition*

$$B(s,t) = B(s+T, t+T),$$

where T is some fixed number.

Many researchers studied periodic ARMA model based on periodically correlated process for periodic problems. The models discussed in different papers had slight difference. Here we stated below the commonly used definition for pARMA model mentioned in [4].

Definition 1.3.4. *($PARMA_d(p,q)$ Process). The periodic ARMA process $\{\tilde{X}(t)\}$ with period d has representation*

$$X(t) - \sum_{j=1}^{p}\phi_j(t)X(t-j) = Z(t) - \sum_{j=1}^{q}\theta_j(t)Z(t-j),$$

where $X(t) = \tilde{X}(t) - \mu_t$ is causal and invertible and $\{Z(t)\}$ is a sequence of random variables with mean zero and standard deviation σ_t such that $\{\sigma_t^{-1}Z(t)\}$ is i.i.d. The $\mu_t := \mathbb{E}\tilde{X}(t)$, the parameters $\phi_j(t)$, $\theta_j(t)$ and σ_t are all periodic functions of t with the same period $d \geq 1$.

1.4 Random Periodic Path

Definition 1.4.1. *Let M be a metric space, $\mathcal{B}(M)$ be is Borel σ-field. A measurable random dynamical system on the measurable space $(M, \mathcal{B}(M))$ over a metric dynamical system $(\Omega, \mathcal{F}, P, (\theta_s)_{s\in\mathbb{R}})$ is a mapping :*

$$\Phi : \mathbb{R}^+ \times \Omega \times M \to M, \quad (t, \omega, x) \mapsto \Phi(t, \omega, x),$$

with following properties:

(i) Measurability: Φ is $(\mathcal{B}(\mathbb{R}^+) \otimes \mathcal{F} \otimes \mathcal{B}(M), \mathcal{B}(M))$-measurable.

(ii) Cocycle property: for almost all $\omega \in \Omega$,

$$\Phi(0, \omega) = id_M,$$

$$\Phi(t + s, \omega) = \Phi(t, \theta_s\omega) \circ \Phi(s, \omega) \text{ for all } s, t \in \mathbb{R}^+.$$

where $\theta(t)$ is a measure preserving and measurably invertible map.

The definition of random periodic path of random dynamical system Φ is give by [20].

Definition 1.4.2. *A random periodic path of period τ of the random dynamical system Φ: $\mathbb{R}^+ \times \Omega \times M \to M$ is an $\mathcal{F}$-measurable map $Y : \mathbb{R} \times \Omega \to M$ such that for almost all $\omega \in \Omega$,*

$$\Phi(t, \theta(s)\omega)Y(s, \omega) = Y(t + s, \omega), Y(s + \tau, \omega) = Y(s, \theta(\tau)\omega), \tag{1.4.1}$$

for any $t \in \mathbb{R}^+$, $s \in \mathbb{R}$.
It is a stationary path if $Y(t, \omega) = Y(0, \theta_t\omega) =: Y_0(\theta_t\omega)$ for all $t \in \mathbb{R}^+$, i.e.

$$\Phi(t, \omega, Y_0(\omega)) = Y_0(\theta_0\omega), t \in \mathbb{R}^+ \ a.s.$$

For a statistical description, we usually do not know the exact expression of the dynamical system driving the time series. We only consider the second equation in (1.4.1) as the definition of random periodic process, while the first part is hidden in the time series evolution.

From the definition of random periodic process, we know that the stationary process can be regarded as a special kind of random periodic processes. The latter describes a mixed structure of seasonal and random patterns. Can we use random periodic process to replace stationary process in the analysis of time series to enlarge the scope of applications and to enhance the accuracy of model estimation? We shall study more properties of the random periodic process.

Define

$$\mathcal{P}(M) := \{\rho : \text{ probability measure on } (M, \mathcal{B}(M))\}.$$

The law of the random periodic paths is defined as

$$\rho_s(\Gamma) = P\{\omega : Y(s,\omega) \in \Gamma\},$$

It is a periodic measure. The definition is given as follows,

Definition 1.4.3. ***(Periodic Measure)*** *For a Markovian cocycle random dynamical system Φ, i.e. $\Phi(t,\omega)x$ is a Markov process, recall that for any $\Gamma \in \mathcal{B}(M)$, the transition probability is defined as*

$$P(t,x,\Gamma) = P\{\omega : \Phi(t,\omega)x \in \Gamma\}, t \in \mathbb{R}^+,$$

and $P_t^* : \mathcal{P}(M) \to \mathcal{P}(M)$ is defined as: for any measure ρ on $\mathcal{B}$,

$$(P_t^*\rho)(\Gamma) = \int_X P(t,x,\Gamma)\rho(dx).$$

A measure function $\{\rho_s\}_{s\in\mathbb{R}}$ in $\mathcal{P}(M)$ is a periodic measure on $(M, \mathcal{B}(M))$ if

$$\rho_{\tau+s} = \rho_s, \quad P_t^*\rho_s = \rho_{s+t}, t \in \mathbb{R}^+.$$

If $\rho_s = \rho_0$ for all s, then ρ_0 is an invariant measure, i.e.

$$P_t^*\rho_0 = \rho_0, t \in \mathbb{R}^+.$$

Note that for any fixed s, ρ_s is an invariant measure of $\{P(k\tau)\}_{k\in\mathbb{Z}}$. Set $\bar{\rho}$ to be the average of a periodic measure over one period, i.e. $\bar{\rho} = \frac{1}{\tau}\int_0^\tau \rho_s ds$, then $\bar{\rho}$ is an invariant measure with respect to transition probability $\{P(t)\}_{t\geq 0}$.

Set $L_s = supp(\rho_s)$, then $L_{\tau+s} = L_s$ and for ρ_s-almost all $x \in L_s$, $t \geq 0$,

$$P(t,x,L_{s+t}) = 1.$$

The sets $L_s \subset X,\ s \geq 0$ are called Poincaré section. Then for each L_s,

$$P(k\tau, x, L_s) = 1, \text{ for any } x \in L_s.$$

This means that starting from $x \in L_s$, $\Phi(k\tau, \omega)x$ returns to L_s with probability one. However, $\Phi(k\tau, \omega)$ does not have a fixed point on L_s. This is in consistent with many real-life time series, e.g. temperature, which inspires us to establish ARMA model based on the theory of random periodic processes.

Chapter 2

Construction of ARMA Model for Random Periodic Processes

2.1 Decomposition Model of Time Series

In this dissertation we only consider $T = \mathbb{Z}$ the discrete time case. Consider a time series $\{X(t)\}_{t \in \mathbb{Z}^+}$ with a set of observation $\{x(t)\}$.

Recall that the classical decomposition model of time series is

$$X(t, \omega) = trend(t) + s(t) + \tilde{X}(t, \omega), \tag{2.1.1}$$

where

1. $trend(t)$, trend component, is an overall slowly changing function of time t.
2. $s(t)$, seasonal component, is periodic function of time t with known period τ.
3. $\tilde{X}(t, \omega)$, random noise component, is a weak stationary process.

In some situations, there is seasonal pattern in the irregular component, in which the classical additive structure is not sufficient. We give a counterexample as follows.

Example 2.1.1.

$$\begin{cases} dX_t = (-\pi X_t + sin(\pi t))dt + (0.1 + 0.3sin(\pi t))dW_t \\ X_0 = 1 \end{cases}. \tag{2.1.2}$$

The solution of this stochastic differential equation (SDE) is (integral from zero to t)

$$X(t) = e^{-\pi t}X(0) + \int_0^t e^{-\pi(t-s)}\sin(\pi s)ds + \int_0^t \sigma(s)e^{-\pi(t-s)}dW_t,$$

where $\sigma(t) := 0.1 + 0.3\sin(\pi t)$, which is a random periodic solution with period 2. We take 20 equal-time-interval points in each period to construct a discrete data set with period 20. We produced 20000 points and set the last 18000 as our data set. The first 2000 points is to ensure the path converges to the random periodic solution. The plot of part of the data set is shown in the Figure 2.1.

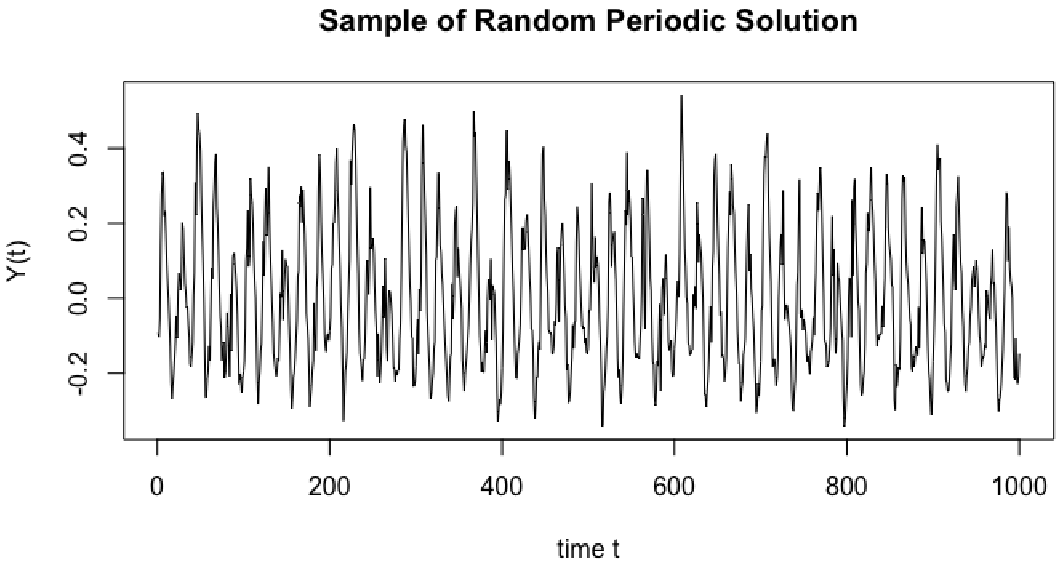


Figure 2.1: The plot of a path of a random periodic solution with period 20.

.

We emphasize that the diffusion term of the SDE is also a periodic function of time t, which implies that the volatility of the data set will change according to the time.

There are many functions in R language to help people analyze a time series. One method to analyze the additive seasonal component is to use the function *stl*(). This function assumes that the time series satisfy the additive model (2.1.1). If a time series applied by this function satisfies the additive model, then the function will decompose this times series into seasonal, trend and noise components. We applied this function to our example. The result is displayed in Figure 2.2. Next we plotted the autocorrelation and partial autocorrelation function of de-seasoned data. We observed that the autocorrelation function decreased gradually, while the partial autocorrelation function dropped sharply after the first lag, which indicated an AR(1) model for the de-seasoned data.

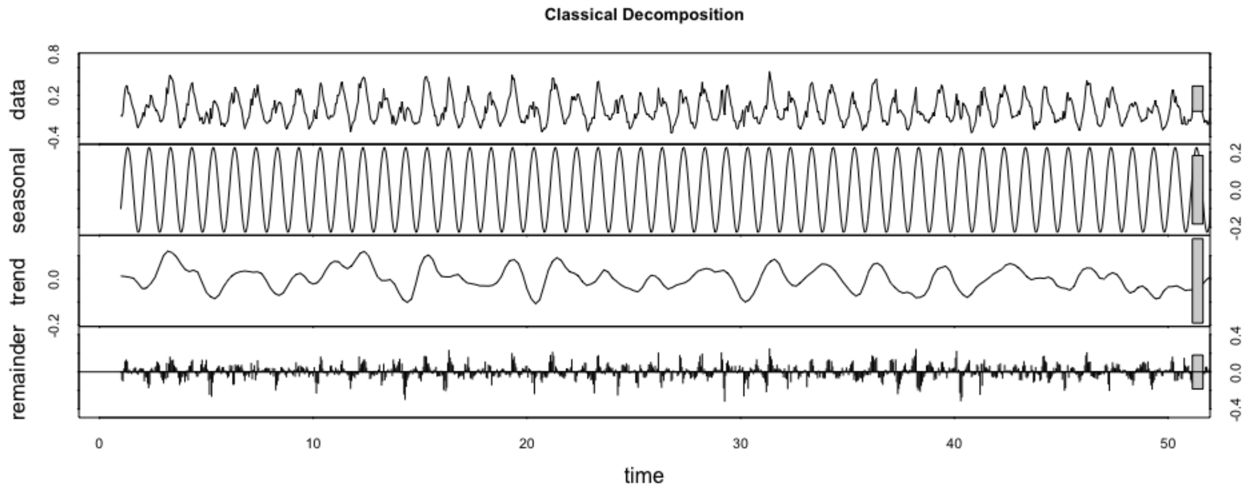


Figure 2.2: Decomposition of seasonal part

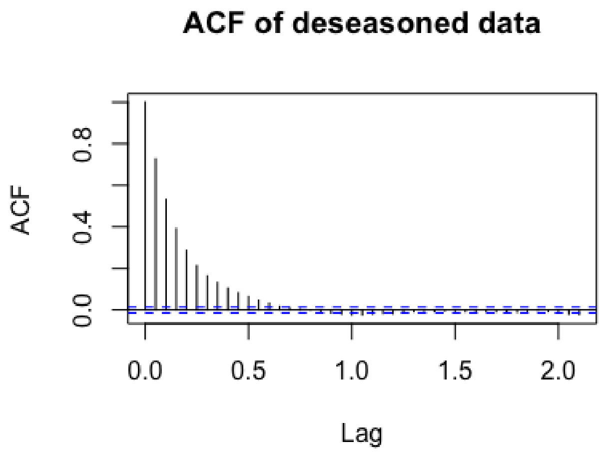


Figure 2.3: The Plot of ACF.

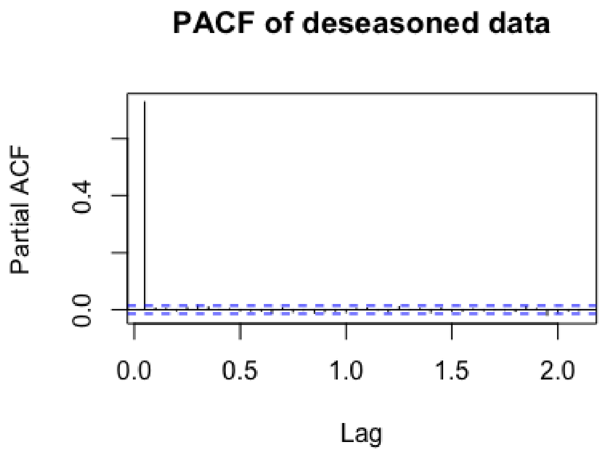


Figure 2.4: The Plot of PACF.

After we figured out the de-seasoned data may satisfy an autoregressive equation, we used Augmented Dickey-Fuller test to test if it is stationary. The procedure of Augmented Dickey-Fuller test is to detect if there is unit root in the autoregressive model for the sample which implies the non-stationarity of time series. If the value of Dickey-Fuller statistic in this test is smaller than the critical value for the Dickey-Fuller t-distribution, then it determines that no unit root is present and the sample is stationary. The corresponding function in R language is *adf.test*(). The null-hypothesis of the Augmented Dickey-Fuller test is that the time series is non-stationary. The value of Dickey-Fuller statistic for this example is -24511, which is smaller than the critical values -3.43 for 1% and -2.86 for 5%. The p-value of the null-hypothesis is smaller than 0.01, which tells us that we should reject the non-stationary hypothesis.

```
> v_decomp = stl(ts(data_v1, frequency = 20), s.window="periodic")
> deseasonal_cnt <- seasadj(v_decomp)
> adf.test(deseasonal_cnt,alternative = "stationary")

        Augmented Dickey-Fuller Test

data:  deseasonal_cnt
Dickey-Fuller = -24.511, Lag order = 26, p-value = 0.01
alternative hypothesis: stationary

Warning message:
In adf.test(deseasonal_cnt, alternative = "stationary") :
  p-value smaller than printed p-value
```

Figure 2.5: ADF test

Then we applied the function *arma(deseasonal_cnt, order = c(1,0))* to the de-seasoned data to obtain the corresponding AR(1) model.

```
> v_fit <- arma(deseasonal_cnt, order = c(1,0))
> v_fit

Call:
arma(x = deseasonal_cnt, order = c(1, 0))

Coefficient(s):
       ar1   intercept
 0.7278463  -0.0006197
```

Figure 2.6: AR(1) model for de-seasoned data.

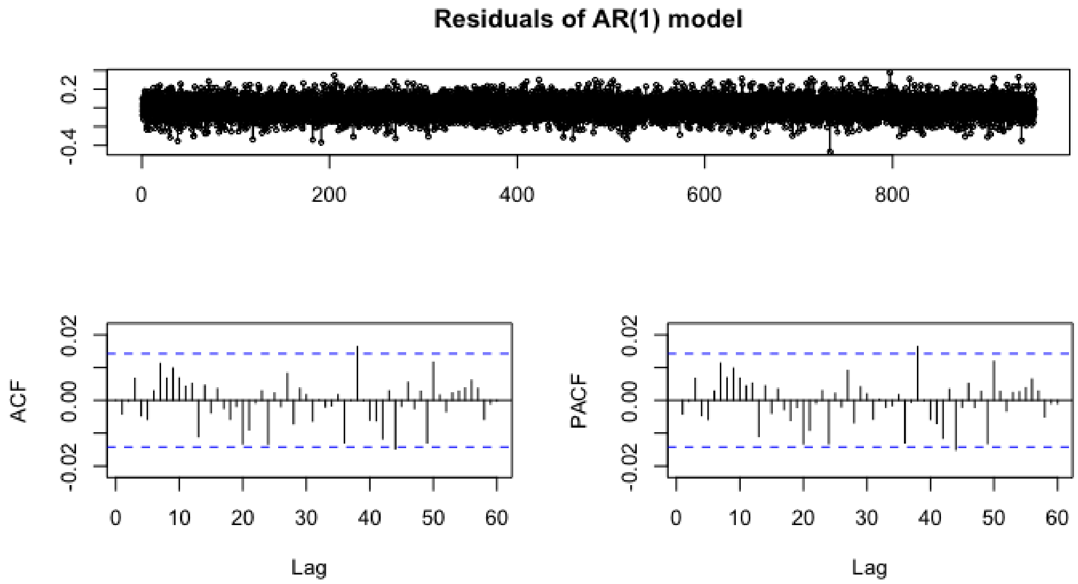


Figure 2.7: Residuals of AR(1) model.

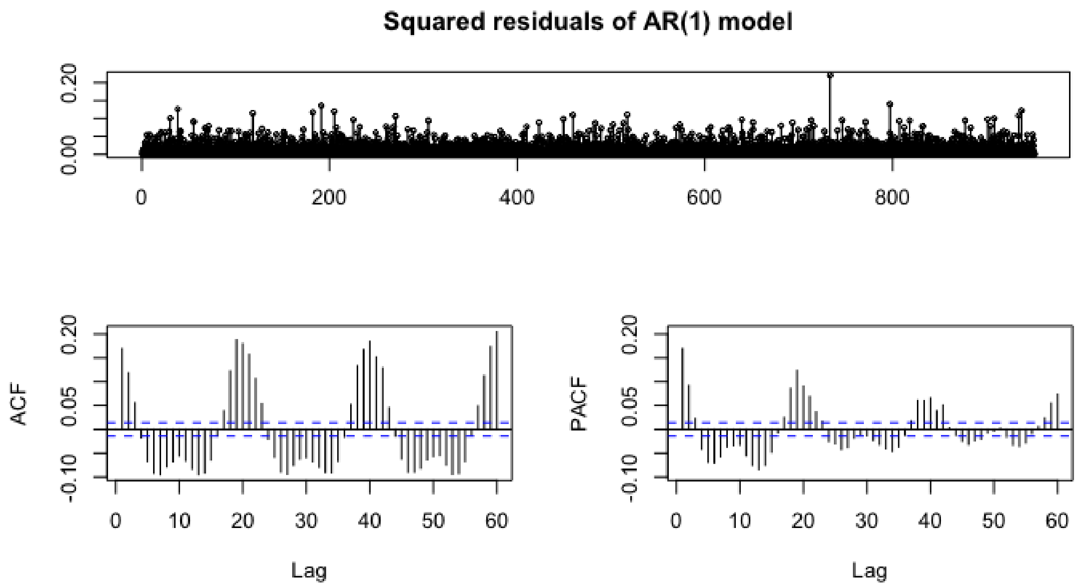


Figure 2.8: Squared residuals of AR(1) models.

We checked the residuals and squared residuals of AR(1) in Figures 2.7 and 2.8. The autocorrelations and partial autocorrelations of residuals are near zero, which indicates the independency of noise. However, the autocorrelations and partial autocorrelations of squared residuals show obvious periodic pattern, which indicates sufficient dependency of time in the volatility of noise.

We then used Shapiro-Wilk normality test to double check whether the residuals satisfy normal distribution. The hypothesis of the test is "normal distribution", while the alternative hypothesis is "not normal distribution". We observed that the Shapiro-Wilk test rejected the "normal" hypothesis with very small p-value. But if we tested on the periodic-point sequence of residuals, we observed that the Shapiro-Wilk test accepted that it satisfies a normal distribution.

```
> shapiro.test(residuals(v_fit)[1:1000])

        Shapiro-Wilk normality test

data:  residuals(v_fit)[1:1000]
W = 0.94558, p-value < 2.2e-16

> shapiro.test(residuals(v_fit)[seq(1,1000, by = 20)])

        Shapiro-Wilk normality test

data:  residuals(v_fit)[seq(1, 1000, by = 20)]
W = 0.97623, p-value = 0.4192
```

Figure 2.9: Shapiro-Wilk normality test of residuals.

This example shows that the standard procedure to analyze a time series by the classical decomposition model and ARMA model fails to obtain the correct result for a mixed season and randomness case. Although some more complicated models can combine with a periodic regression model to fit the squared residuals, it is not as parsimony as to use a random periodic process to approach the data, and the complexity of model may rise the risk of error.

The insufficiency of the classical decomposition to figure out the mixed structure of the seasonal and noise components inspires us to make modification to it.

We attempt to use random periodic process to describe the seasonal and noise components simultaneously. The modified model is as follows:

$$X(t,\omega) = trend(t) + Y(t,\omega),$$

where $\{Y(t)\}$ is a random periodic process. The properties of random periodic process are listed below.

Proposition 2.1.2. *The random periodic process $\{Y(t,\omega)\}$ of period τ satisfies:*

i) $\mathbb{E}[Y(t)^2] < \infty$.

ii) $\mathbb{E}[Y(t)] = m(t)$, *$m(t)$ is a deterministic periodic function of time t satisfying $m(t+\tau) = m(t)$.*

iii) Define $\gamma_Y(r,s) := cov(Y(r), Y(s))$. Then $\gamma_Y(r,s) = \gamma_Y(r+\tau, s+\tau)$ for any $r, s \in \mathbb{Z}$. The autocovariance function of random periodic process is periodic function of time t on both indexes.

Proof. Since $\theta(t)$ is measure-preserving,

$$\begin{aligned}
& \gamma_Y(r+\tau, s+\tau) \\
= & \ \mathbb{E}\left[(Y(r+\tau,\omega) - m(r+\tau))\,(Y(s+\tau,\omega) - m(s+\tau))\right] \\
= & \ \mathbb{E}\left[(Y(r,\theta(\tau)\omega) - m(r))\,(Y(s,\theta(\tau)\omega) - m(s))\right] \\
= & \ \mathbb{E}\left[(Y(r,\omega) - m(r))\,(Y(s,\omega) - m(s))\right] \\
= & \ \gamma_Y(r,s).
\end{aligned}$$

□

2.2 ARMA Model for Random Periodic Processes

Next we give the definition of the ARMA model for random periodic processes.

Definition 2.2.1 (ARMA Model for Random Periodic Processes)**.** *Define the random periodic process $\{Y(t)\}$ with period τ to be a ARMA(p,q) process with period τ, if for each t,*

$$Y(t) - \sum_{i=1}^{p} \phi_i(t) Y(t-i) = \sum_{i=0}^{q} \theta_i(t) Z(t-i), \tag{2.2.1}$$

where $\{Z(t)\} \sim IID(0,1)$ *is a sequence of i.i.d. random variables, and for* $i = 1, \cdots, p$, $j = 1, \cdots, q$, *the coefficients* $\phi_i(t)$ *and* $\theta_j(t)$ *are deterministic functions of time* t. *In particular, if* $q = 0$, *then it is AR(p) model for random periodic processes, i.e.*

$$Y(t) - \phi_1(t)Y(t-1) - \cdots - \phi_p(t)Y(t-p) = \theta_0(t)Z(t). \tag{2.2.2}$$

If $p = 0$, *it is MA(q) model for random periodic processes, i.e.*

$$Y(t) = \theta_0(t)Z(t) + \theta_1(t)Z(t-1) + \cdots + \theta_q(t)Z(t-q). \tag{2.2.3}$$

The periodicity of the coefficients is constructed in the following proposition.

Proposition 2.2.2. *The coefficients of AR(p) model and MA model for random periodic processes are all deterministic periodic functions of time* t.

Proof. First consider the AR(p) equation (2.2.2). Multiplying each side of (2.2.2) by $Y(k)$, $k = t-1, \cdots, t-p$ and taking expectations, as $Y(k)$ only depends on $\{Z(k), Z(k-1), Z(k-2), \cdots\}$, noting

$$\mathbb{E}[Y(k)Z(t)] = \mathbb{E}Y(k)\mathbb{E}Z(t) = 0,$$

hence one obtains

$$\Gamma_p(t)\boldsymbol{\phi}_p(t) = \boldsymbol{\gamma}_p(t), \tag{2.2.4}$$

where

$$\Gamma_p(t) = \begin{bmatrix} \gamma(t-1,t-1) & \gamma(t-1,t-2) & \cdots & \gamma(t-1,t-p) \\ \gamma(t-2,t-1) & \gamma(t-2,t-2) & \cdots & \gamma(t-2,t-p) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(t-p,t-1) & \gamma(t-p,t-2) & \cdots & \gamma(t-p,t-p) \end{bmatrix},$$

$$\boldsymbol{\phi}_p(t) = (\phi_1(t), \phi_2(t), \cdots, \phi_p(t))^T,$$

and

$$\boldsymbol{\gamma}_p(t) = (\gamma(t,t-1), \gamma(t,t-2), \cdots, \gamma(t,t-p))^T.$$

Multiplying each side of (2.2.2) by $Y(t)$ and taking expectation, one obtains

$$\begin{aligned} \theta_0^2(t) = &\ \gamma(t,t) - \phi_1(t)\gamma(t,t-1) - \cdots - \phi_p(t)\gamma(t,t-p) \\ = &\ \gamma(t,t) - \boldsymbol{\phi}_p(t) \cdot \boldsymbol{\gamma}_p(t). \end{aligned}$$

Hence the coefficients $\boldsymbol{\phi}_p(t)$ is the solution of the linear system (2.2.4), and $\theta_0(t)$ is represented by $\boldsymbol{\phi}_p(t)$ and $\boldsymbol{\gamma}_p(t)$. Similarly, consider

$$Y(t+\tau) - \phi_1(t+\tau)Y(t-1+\tau) - \cdots - \phi_p(t+\tau)Y(t-p+\tau) = \theta_0(t+\tau)Z(t+\tau),$$

the coefficients $\boldsymbol{\phi}_p(t+\tau)$ is the solution of

$$\Gamma_p(t+\tau)\boldsymbol{\phi}_p(t+\tau) = \boldsymbol{\gamma}_p(t+\tau), \tag{2.2.5}$$

where

$$\begin{aligned}\Gamma_p(t+\tau) &= \begin{bmatrix} \gamma(t+\tau-1,t+\tau-1) & \gamma(t+\tau-1,t+\tau-2) & \cdots & \gamma(t+\tau-1,t+\tau-p) \\ \gamma(t+\tau-2,t+\tau-1) & \gamma(t+\tau-2,t+\tau-2) & \cdots & \gamma(t+\tau-2,t+\tau-p) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(t+\tau-p,t+\tau-1) & \gamma(t+\tau-p,t+\tau-2) & \cdots & \gamma(t+\tau-p,t+\tau-p) \end{bmatrix} \\ &= \Gamma_p(t),\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{\gamma}_p(t+\tau) &= (\gamma(t+\tau,t+\tau-1), \gamma(t+\tau,t+\tau-2), \cdots, \gamma(t+\tau,t+\tau-p))^T \\ &= \boldsymbol{\gamma}_p(t).\end{aligned}$$

Hence

$$\phi_j(t+\tau) = \phi_j(t), j = 1, 2\cdots, p.$$

Next we consider the periodic MA(∞) equation. Suppose $Y(t)$ satisfies the following model,

$$Y(t) = \sum_{j=-\infty}^{\infty} \psi_j(t)Z(t-j), \qquad t = 0, \pm 1, \cdots. \tag{2.2.6}$$

Multiplying each side of (2.2.3) by $Z(t-j)$ and taking expectations, then

$$\psi_j(t) = \mathbb{E}[Y(t)Z(t-j)], \qquad j \in \mathbb{Z}.$$

We build an equivalent pathwise definition of $\{Z(t)\}_{t\in\mathbb{Z}}$. Define

$$\omega = \cdots Z(-2)Z(-1) : Z(0) : Z(1)Z(2)\cdots$$

as a sequence. Here

$$\omega(0) = Z(0), \qquad \omega(n) = Z(n) =: Z(n, \omega).$$

Define Ω_1 to be a set containing all possible ω above,

$$\theta\omega = \cdots Z(-1)Z(0) : Z(1) : Z(2)Z(3)\cdots ,$$

and

$$\theta^n = \underbrace{\theta\theta\cdots\theta}_{n}.$$

Then it is easy to see that for any n,

$$Z(n, \theta\omega) = (\theta\omega)(n) = Z(n+1) = Z(n+1, \omega).$$

This is also true for $Z(n+\tau, \omega) = Z(n, \theta(\tau)\omega)$. Therefore $Y(t)Z(t-j)$ is also a random periodic process with period τ, so we have $\mathbb{E}[Y(t)Z(t-j)] = \mathbb{E}[Y(t+\tau)Z(t-j+\tau)]$, hence $\psi_j(t) = \psi_j(t+\tau)$, for any $j \in \mathbb{Z}^+$. $\square$

We also proved the periodicity of coefficients for causal ARMA equation. Recall the definition of causality.

Definition 2.2.3. ***(Causality)*** *An ARMA(p, q) model for random periodic processes is said to be causal if it can be represented as*

$$Y(t) = \sum_{j=0}^{\infty} \psi_j(t)Z(t-j), \qquad t = 0, \pm 1, \cdots . \tag{2.2.7}$$

Note that the process $\{Y(t)\}$ is causal if and only if $\phi_t(z) := 1 - \phi_1(t)z - \phi_p(t)z^p$ has roots outside the unit circle by the property of Laurent series.

Proposition 2.2.4. *Assume $Y(t)$ is causal and satisfies the equation (2.2.1). The coefficients ϕ_t and θ_t are deterministic periodic functions of time t.*

Proof. Suppose $Y(t)$ is causal and satisfies the equation (2.2.1). Replacing $Y(t-i)$ in (2.2.1) by

$$Y(t-i) = \sum_{j=0}^{\infty} \psi_j(t-i)Z(t-i-j), i = 1, 2, \cdots , p,$$

and equating the coefficients of Z, we obtain

$$\begin{aligned} &\theta_0(t)Z(t) + \theta_1(t)Z(t-1) + \cdots + \theta_q(t)Z(t-q) \\ = &\sum_{j=0}^{\infty} \psi_j(t)Z(t-j) - \phi_1(t)\sum_{j=0}^{\infty} \psi_j(t-1)Z(t-1-j) - \cdots - \phi_p(t)\sum_{j=0}^{\infty} \psi_j(t-p)Z(t-p-j), \end{aligned}$$

and

$$\begin{cases} \theta_0(t) = \psi_0(t), \\ \theta_1(t) = \psi_1(t) - \phi_1(t)\psi_0(t-1), \\ \cdots, \\ \theta_q(t) = \psi_q(t) - \phi_1(t)\psi_{q-1}(t-1) - \cdots - \phi_k(t)\psi_{q-k}(t-k), \text{ where } k = max(p,q) \\ \cdots, \\ 0 = \psi_k(t) - \phi_1(t)\psi_{k-1}(t-1) - \cdots - \phi_p(t)\psi_{k-p}(t-p), \\ \cdots . \end{cases}$$

Hence the coefficients $\boldsymbol{\phi}(t)$ and $\boldsymbol{\theta}(t)$ can be calculated by

$$\begin{cases} \theta_0(t) = \psi_0(t) & \\ \theta_j(t) = \psi_j(t) - \sum_{i=1}^{min(j,p)} \phi_i(t)\psi_{j-i}(t-i) & 0 < j < max(p, q+1) \\ 0 = \psi_j(t) - \sum_{i=1}^{p} \phi_i(t)\psi_{j-i}(t-i) & j \geq max(p, q+1) \end{cases}. \tag{2.2.8}$$

If $q < p$, set $\theta_j(t) = 0$ for $j = q+1, \cdots, p$. The coefficients $\boldsymbol{\phi}_p(t)$ are calculated from $\boldsymbol{\psi}$, and $\boldsymbol{\theta}_q(t)$ are calculated from $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$, hence they are all periodic functions. □

Remark 2.2.5. *(Comparison between seasonal ARMA model, pARMA model and ARMA model for random periodic processes).*

These three models are all non-structural models which combine autoregressive function of data process and moving-average function of some noise sequence in one equation. Seasonal ARMA model admits that the seasonal component is deterministic and is additive to the random component. It says that after differencing the original data process by τ^{th} *lag, the remaining is stationary. However, the process* $\{Y(t)\}$ *considered in pARMA model and ARMA model for random periodic processes admits more complicated structure of periodicity and randomness.*

The periodicity of coefficients of pARMA model is defined by tuition, while we proved the periodicity of coefficients of ARMA model for random periodic processes in the above propositions.

A time series is a path of some stochastic process. The random periodic processes are defined pathwisely. [20] and other papers studied many important properties of random periodic process in pathwise sense recently. These researches built a fundamental theoretic basis to help improving some areas of time series analysis. For example, one can analyze a periodic time series path in the structure of random dynamical system to help understanding more characters of the corresponding stochastic process.

The periodically correlated process is defined only by its first and second moment, which show periodicity in the same way as random periodic processes. But that is not necessary to say that the periodically correlated process is a more general version than random periodic processes. In the proof of Proposition 2.2.2 we showed that one can obtain that a process is random periodic by constructing path ω and transformation θ if we know the mean and covariance functions are periodic and some other conditions.

2.3 Sample Mean and Sample Autocovariance

Recall that the original formula of sample autocovariane for stationary processes is

$$\hat{\gamma}(h) := \frac{1}{n-h-1}\sum_{i=1}^{n-h} y(i)y(i+h).$$

To avoid the sample autocovariance matrix to be singular, people usually use the following average to calculate the sample covariance in real cases ([12]):

$$\hat{\gamma}(h) := \frac{1}{n}\sum_{i=1}^{n-h} y(i)y(i+h).$$

Since the covariance function of random periodic process depends on both the time points and the distance between two time points, the original formula of sample autocovariane is not suitable any more. As the process of $\{Y(t-i\tau)\}_{i\in\mathbb{Z}}$ is stationary for fixed t, it is suggested using the average of the fixed time point in every period to represent the sample mean and sample autocovariance.

In the recent work about periodic ARMA model, people used the following average formulas to estimate the sample mean and sample autocovariance: suppose there are N cycles in the sample data set and the period is τ, then the sample mean is estimated by

$$\hat{\mu}(j) = \frac{1}{N}\sum_{i=0}^{N-1} y(j+i\tau), \quad j = 1, 2, \cdots, \tau,$$

and the sample autocovariance between j^{th} and k^{th} seasons is estimated by

$$\hat{\gamma}(j,k) = \frac{1}{N}\sum_{i=0}^{N-1}(y(j+i\tau) - \hat{\mu}(j))(y(k+i\tau) - \hat{\mu}(k))$$

where $j, k = 1, 2, \cdots, \tau$. Although this definition of sample mean and sample autocovariance functions for periodic case show periodicity of the process, but it is deterministic periodicity of approximation. That is, the same time points in different periods will have the same behaviour. This omits the small fluctuations between different periods.

For random periodic process, we do not use the above formulas. As the sample mean and sample autocovariance both depend on the time where the process is, we use a moving-average term of backward data to estimate them, which will reflect the periodicity as well as randomness hidden in the revolution of the corresponding random periodic process.

Before we define the sample mean and sample autocovariance, we first give the following assumption, which we will discuss again in the next chapter.

Assumption 1. *For any discrete random periodic process $\{Y(t)\}_{t\in\mathbb{Z}}$ with mean zero, assume that for any $\epsilon_1 > 0$ and $\epsilon_2 > 0$ there exists a n_0 such that for any $t \in \mathbb{Z}^+$ and $h \in \mathbb{Z}^+$, when $n \geq n_0$,*

$$P\left(\left|\frac{1}{n}\sum_{k=0}^{n-1} Y(t-k\tau)Y(t+h-k\tau) - \gamma(t,t+h)\right| < \epsilon_1\right) > 1-\epsilon_2.$$

Under this assumption, we define the sample mean and sample autocovariance of random periodic process.

Definition 2.3.1. ***(the Sample Mean and Sample Autocovariance)***
For a sample of data $\{y(1), \cdots, y(n)\}$, the sample mean of random periodic processes is defined as

$$\hat{m}(t) = \frac{1}{w}\sum_{i=0}^{w-1} y(t-i\tau), \quad t = 1+(w-1)\tau, 2+(w-1)\tau, \cdots, n, \tag{2.3.1}$$

where $\{y(t)\}_{t\in\mathbb{Z}^+}$ is the observed data and w is the number of cycles we choose to estimate the sample mean.

The sample autocovariance is defined as

$$\hat{\gamma}(t,s) = \frac{1}{w}\sum_{i=0}^{w-1}\left(y(t-i\tau) - \hat{m}(t-i\tau)\right)\left(y(s-i\tau) - \hat{m}(s-i\tau)\right), \tag{2.3.2}$$

where $s, t = 1+(w-1)\tau, 2+(w-1)\tau, \cdots, n$.

Example 2.3.2. *Continued with the Example 2.1.1, we calculate the sample mean and sample autocovariance and display the plots in Figures 2.10 and 2.11. We can see that the plot of sample mean is nearly a deterministic periodic function of time t. And the sample autocovariance at point $t = 15006$ tends to converge as w increases.*

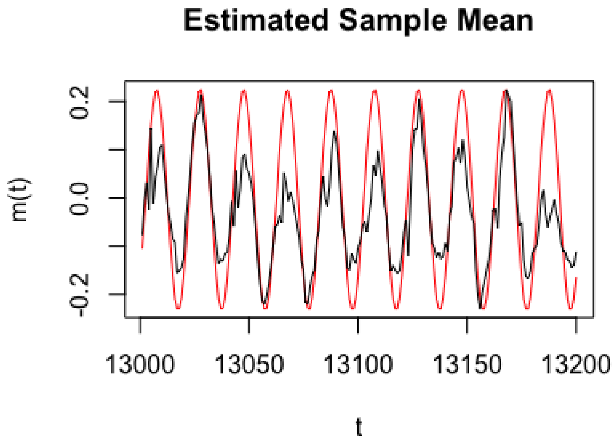


Figure 2.10: $\hat{m}(t)$ with $N = 650$.

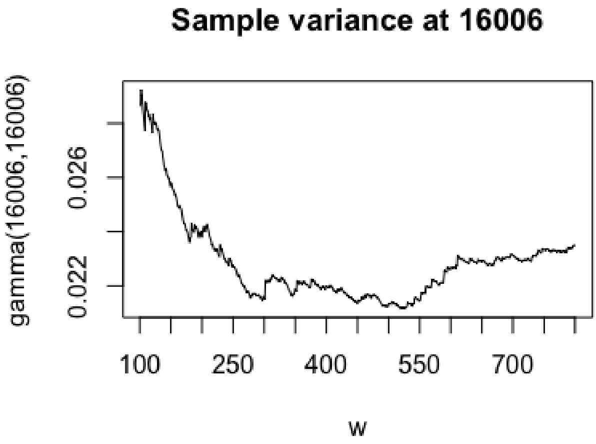


Figure 2.11: $\hat{\gamma}(15006, 15006)$.

The following 3D plot displays the evolution of $\hat{\gamma}(i, j)$ for time $i, j = 13001, \cdots, 13100$ when we take $w = 650$. We can observe that the periodicity is shown as the two indexes increase simultaneously.

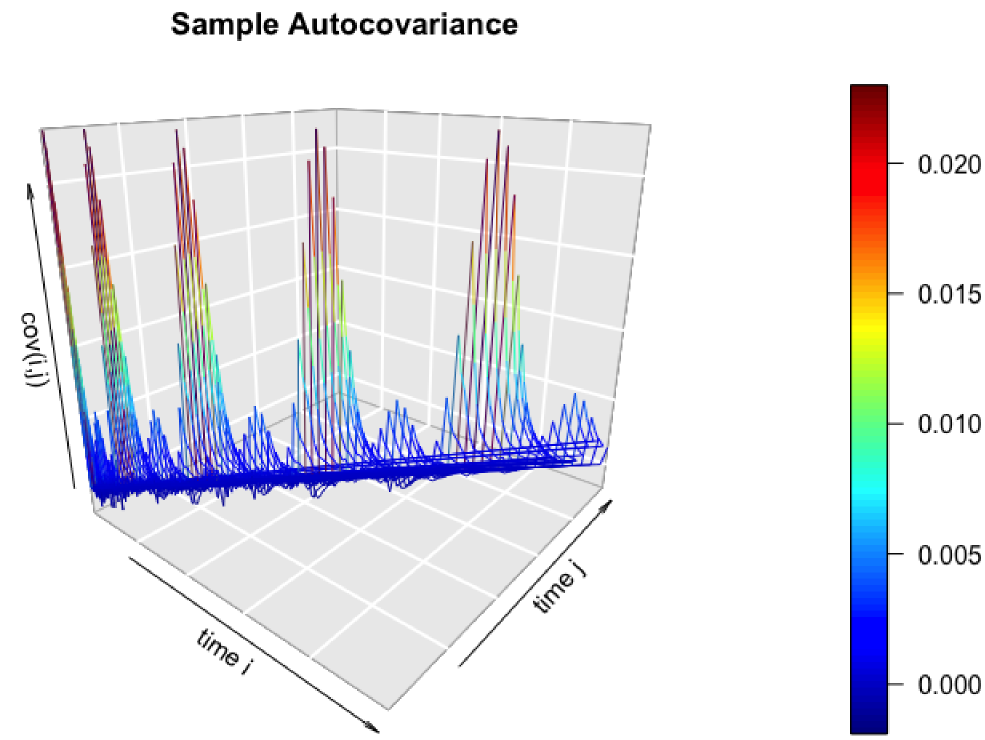


Figure 2.12: The 3D plot of the sample autocovariance.

.

2.4 Central Limit Theorem

A time series is a path of the corresponding stochastic process. Usually we cannot obtain enough paths of a stochastic process by repeating experiments to calculate the statistical qualities such as mean and variance. Under the law of large numbers and central limit theory one can use time average to approximate the state average. Therefore we have formulas for sample mean and sample autocovariance. In this section we will deduce the important fundamental theory for random periodic processes, the central limit theory.

From [20] (Feng-Zhao 2015), under certain conditions, one can construct random periodic paths from a periodic measure. They also proved the Law of Large Numbers (LLN) of the constructed random periodic process.

Condition B. *The Markovian cocycle* $\Phi : \mathbb{Z}^+ \times \Omega \times M \to M$ *has a periodic measure* $\rho : \mathbb{Z} \to \mathcal{P}(M)$ *and for any* $B \in \mathcal{B}(M)$*, as* $k \to \infty$

$$r(k) := \int_M \left| \frac{1}{m(F_0)} \sum_{s \in F_0} (P(s, y, B) - \rho_s(B)) \right| \rho_0(dy) \to 0.$$

where $F_0 := [t_1, t_2] \subset [1, \tau]$*,* $m(F_0)$ *is the Lebesgue measure of set* F_0 *and for* $k = 0, 1, 2, \cdots,$

$$F_k := [t_1 + k\tau, t_2 + k\tau], \;\; G_N := \cup_{k=0}^{N-1} F_k \quad \textit{and} \quad G_\infty := \cup_{k=0}^{\infty} F_k.$$

Theorem 2.4.1. *(WLLN) Under Condition B, the random periodic path* Y *and its law* ρ *satisfy WLLN, i.e. as* $N \to \infty$*,*

$$\frac{1}{m(G_N)} \sum_{t \in G_N} I_B(Y(t, \omega)) \xrightarrow{\mathbb{P}} \frac{1}{m(F_0)} \sum_{t \in F_0} \rho_t(B).$$

Theorem 2.4.2. *(SLLN) Under Condition B, as* $\mathbb{Z}^+ \ni T \to \infty$*,*

$$\frac{1}{m([0, T) \cap G_\infty)} \sum_{t \in [0, T) \cap G_\infty} I_B(Y(t, \omega)) \to \frac{1}{m(F_0)} \sum_{t \in F_0} \rho_t(B) \;\; a.s.$$

In particular, if Condition B holds for $F_0 = [1, \tau]$*, then as* $T \to \infty$*,*

$$\frac{1}{T} \sum_{t=1}^{T} I_B(Y(t, \omega)) \to \bar{\rho}(B) \;\; a.s.$$

where $\bar{\rho} = \frac{1}{\tau} \sum_{t=1}^{\tau} \rho_t$*.*

We introduce two more conditions to help proving the central limit theorem (CLT) of random periodic path. In the proof we will use the result of CLT for stationary process in [16] stated as follows.

Theorem 2.4.3. ***(Central Limit Theorem for Stationary Sequence)*** *Suppose* X_n*,* $n \in \mathbb{Z}$*, is an ergodic stationary sequence with* $\mathbb{E}X_n = 0$ *and* $\mathbb{E}|X_0|^{2+\delta} < \infty$*. Let*

$$\alpha(n) = \alpha(\mathcal{F}_{-n}, \sigma(X_0)) = \sup\{|P(A \bigcap B) - P(A)P(B)| : A \in \mathcal{F}_{-n}, B \in \sigma(X_0)\},$$

where $\mathcal{F}_{-n} = \sigma(X_m : m \leq -n)$, *and suppose* $\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{2(2+\delta)}} < \infty$. *Let* $S_n = X_1 + \cdots + X_n$, *then*

$$\frac{S_n}{\sqrt{n}} \to Z,$$

in distribution as $n \to \infty$, *where* $Z \sim N(0, \sigma^2)$, *and* $\sigma^2 = \mathbb{E}X_0^2 + 2\sum_{n=1}^{\infty} \mathbb{E}X_0 X_n$.

Consider a Markovian cocycle random dynamical system Φ on a filtered dynamical system $(\Omega, \mathcal{F}, P, (\theta_t)_{t\in\mathbb{Z}}, (\mathcal{F}_s^t)_{s<t})$. Let

$$P(t, x, B) = P(\{\omega : \Phi(t, \omega)x \in B\}), t \in \mathbb{Z}^+, B \in \mathcal{B}(M)$$

be the transition probability of Markovian process $\Phi(t, \omega)x$ on the Polish space M with the Borel σ-algebra $\mathcal{B}(M)$. Set $L_s \subset M, s \in \mathbb{Z}^+$ to be the invariant set of periodic measure ρ_s. It is also the Poincaré sections of the transition probability $P(t, \cdot, \cdot), t \in \mathbb{Z}^+$ defined in [20]. Assume that the random periodic path $Y(s)$ of Φ is adapted. To construct the CLT for Y, we need the following

Condition A*. *For any* $s \geq 0$, $B \subset L_{t+s}$, $t \geq 0$, *the map* $x \mapsto P(t, x, B)$ *is continuous in* $x \in L_s$.

Condition B*. *The Markovian transition probability* $P(t + n\tau, x, B)$ *converges weakly to* $\rho_t(B)$ *as* $n \to \infty$ *for any* $t \geq 0$, *any* $B \in \mathcal{B}(M)$ *and a.e.* $x \in M$. *And for any continuous and bounded function* $f \in C_b(M)$, *there exists a* $\delta > 0$ *and* $\epsilon > 0$ *such that*

$$\left| \int_B f(y)P(t + (n-1)\tau, x, dy) - \int_B f(y)\rho_t(dy) \right| < n^{-(\frac{4}{\delta}+2+\epsilon)},$$

for any $n \in \mathbb{Z}^+$ *and* $t \geq 0$, $B \subset L_t$.

Condition A^* shows the strong Feller property of Markovian semigroup. For example in [18], if the solution of an SDE satisfies weakly dissipative condition, then the solution will be continuous on the initial condition. Condition B^* shows the weak convergence of Markovian transition probability to the periodic measure. This is not very hard to achieve for ergodic random periodic processes. The properties related to these two conditions have been discussed in detail in [20].

Now we state the central limit theorem of random periodic path as follows.

Theorem 2.4.4. ***(Central Limit Theorem of Random Periodic Path)*** *Assume the semigroup transition probability $P(t,x,\cdot)$ and periodic measure $\{\rho_s\}_{s\in\mathbb{R}}$ satisfy Conditions A^* and B^*. Also assume $\mathbb{E}Y(n)=0$ for all $n\in\mathbb{Z}^+$, and $\mathbb{E}|\sum_{j=1}^{\tau}Y(j)|^{2+\delta}<\infty$. Define $S_n:=Y(1)+\cdots+Y(n)$. Then the random periodic path Y satisfies the CLT. i.e.*

$$\frac{S_n}{\sqrt{n}}\to Z(\tau),$$

in distribution as $n\to\infty$, where $Z(\tau)\sim N(0,\sigma_\tau^2)$, and

$$\sigma_\tau^2=\frac{1}{\tau}\left(\mathbb{E}\left[\sum_{j=1}^{\tau}Y(j)\right]^2+2\sum_{n=1}^{\infty}\mathbb{E}\left[\left(\sum_{j=1}^{\tau}Y(j)\right)\left(\sum_{j=1}^{\tau}Y(j+n\tau)\right)\right]\right).$$

We will first prove a lemma which is needed in the proof of the CLT.

Lemma 2.4.5. *Suppose the assumptions in Theorem 2.4.4 are satisfied. Let $0\le m_1<m_2\le\tau$, and set $X(n,\omega):=\sum_{j=m_1}^{m_2}Y(j+n\tau,\omega)$. Assume that $\mathbb{E}|\sum_{j=m_1}^{m_2}Y(j)|^{2+\delta}<\infty$. Then $\{X(n)\}_{n\in\mathbb{R}}$ satisfy Theorem 2.4.3, i.e.*

$$\frac{\sum_{i=1}^{n}X(i)}{\sqrt{n}}\to Z,$$

in distribution as $n\to\infty$, where $S_n(X)=X(1)+\cdots+X(n)$, $Z\sim N(0,\sigma^2)$, and $\sigma^2=\mathbb{E}X^2(0)+2\sum_{n=1}^{\infty}\mathbb{E}X(0)X(n)$.

Proof. For $\{X(n,\omega)\}_{n\in\mathbb{R},\omega\in\Omega}$, define

$$\alpha(n):=\sup_{A\in\mathcal{G}_{-\infty}^{-n},B\in\sigma(X(0))}\left|P(X(-n)\in A,X(0)\in B)-P(X(-n)\in A)P(X(0)\in B)\right|,$$

where $\mathcal{G}_{-\infty}^{-n}:=\sigma(X(m),m\le -n)$. First note that

$$\begin{aligned}
&P(X(0)\in B)\\
=\ &P\left(\sum_{j=m_1}^{m_2}Y(j)\in B\right)\\
=\ &\int_M P\left(\sum_{j=m_1+1}^{m_2}Y(j)\in B^{-y_{m_1}}\,|Y(m_1)=y_{m_1}\right)\rho_{m_1}(dy_{m_1})\\
=\ &\int_M\int_M P\left(\sum_{j=m_1+2}^{m_2}Y(j)\in B^{-y_{m_1}-y_{m_1+1}}\,|Y(m_1)=y_{m_1},Y(m_1+1)=y_{m_1+1}\right)\\
&\qquad P\left(1,y_{m_1},dy_{m_1+1}\right)\rho_{m_1}(dy_{m_1})
\end{aligned}$$

$$= \int_M \int_M P\left(\sum_{j=m_1+2}^{m_2} Y(j) \in B^{-y_{m_1}-y_{m_1+1}} |Y(m_1+1) = y_{m_1+1} \right)$$
$$P\left(1, y_{m_1}, dy_{m_1+1}\right) \rho_{m_1}(dy_{m_1}), \tag{2.4.1}$$

where $B^{-y} := \{x \in X : x + y \in B\}$ and ρ_i is the law of $Y(i)$ for $i = 1, \cdots, \tau$. The last equation is because of the Markov property of random periodic process Y. By repeating the total probability formula and using the transition probability, one can obtain that

$$\begin{aligned} & P(X(0) \in B) \\ = & \int_M \cdots \int_M P\left(Y(m_2) \in B^{-\sum_{j=m_1}^{m_2-1} y_j} |Y(m_2-1) = y_{m_2-1}\right) \\ & \cdot P\left(1, y_{m_2-2}, dy_{m_2-1}\right) \cdots P\left(1, y_{m_1}, dy_{m_1+1}\right) \rho_{m1}(dy_{m_1}). \\ = & \int_X \cdots \int_X P\left(1, y_{m_2-1}, B^{-\sum_{j=m_1}^{m_2-1} y_j}\right) P\left(1, y_{m_2-2}, dy_{m_2-1}\right) \cdots \\ & \cdot P\left(1, y_{m_1}, dy_{m_1+1}\right) \rho_{m1}(dy_{m_1}), \end{aligned}$$

Now notice that

$$\begin{aligned} & P(X(-n) \in A, X(0) \in B) - P(X(-n) \in A)P(X(0) \in B) \\ = \ & \mathbb{E}I_A(X(-n))I_B(X(0)) - \mathbb{E}I_A(X(-n))\mathbb{E}I_B(X(0)) \\ = \ & \mathbb{E}\left[\left(I_A(X(-n)) - \mathbb{E}I_A(X(-n))\right)\left(I_B(X(0)) - \mathbb{E}I_B(X(0))\right)\right] \\ = \ & \mathbb{E}\left[\mathbb{E}\left[\left(I_A(X(-n)) - \mathbb{E}I_A(X(-n))\right)\left(I_B(X(0)) - \mathbb{E}I_B(X(0))\right) \middle| \mathcal{F}_{-\infty}^{-(n-1)}\right]\right] \\ = \ & \mathbb{E}\left[\left(I_A(X(-n)) - \mathbb{E}I_A(X(-n))\right)\mathbb{E}\left[I_B(X(0)) - \mathbb{E}I_B(X(0)) \middle| \mathcal{F}_{-\infty}^{-(n-1)}\right]\right]. \end{aligned} \tag{2.4.2}$$

Let us consider $\mathbb{E}\left[I_B(X(0)) - \mathbb{E}I_B(X(0)) \middle| \mathcal{F}_{-\infty}^{-(n-1)}\right]$ first. Note that

$$\begin{aligned} & \mathbb{E}\left[I_B(X(0)) - \mathbb{E}I_B(X(0)) \middle| \mathcal{F}_{-\infty}^{-(n-1)}\right] \\ = \ & \mathbb{E}\left[I_B(X(0)) \middle| \mathcal{F}_{-\infty}^{-(n-1)}\right] - \mathbb{E}I_B(X(0)) \\ = \ & P(X(0) \in B \left| \mathcal{F}_{-\infty}^{-(n-1)}\right.) - P(X(0) \in B). \end{aligned}$$

By using a similar method as (2.4.1), noting that

$$Y(m_1, \omega) = \Phi(m_1 + (n-1)\tau, \theta(-(n-1)\tau)\omega)Y(-(n-1)\tau, \omega),$$

we have

$$P(X(0) \in B \left| \mathcal{F}_{-\infty}^{-(n-1)}\right.)$$

$$
\begin{aligned}
&= P\left(\sum_{j=m_1}^{m_2} Y(j) \in B \,\middle|\, \mathcal{F}_{-\infty}^{-(n-1)}\right) \\
&= \int_M P\left(\sum_{j=m_1+1}^{m_2} Y(j) \in B^{-y_{m_1}} \,\middle|\, Y(m_1) = y_{m_1}, \mathcal{F}_{-\infty}^{-(n-1)}\right) P(Y(m_1) \in dy_{m_1} \,\middle|\, \mathcal{F}_{-\infty}^{-(n-1)}) \\
&\cdots \\
&= \int_M \cdots \int_M P\left(Y(m_2) \in B^{-\sum_{j=m_1}^{m_2-1} y_j} \,|Y(m_2-1) = y_{m_2-1}\right) \\
&\quad \cdot P\left(Y(m_2-1) \in dy_{m_2-1} \,|Y(m_2-2) = y_{m_2-2}\right) \cdots \\
&\quad \cdot P\left(Y(m_1+1) \in dy_{m_1+1} \,|Y(m_1) = y_{m_1}\right) P(Y(m_1) \in dy_{m_1} \,\middle|\, \mathcal{F}_{-\infty}^{-(n-1)}) \\
&= \int_M \cdots \int_M P\left(1, y_{m_2-1}, B^{-\sum_{j=m_1}^{m_2-1} y_j}\right) P\left(1, y_{m_2-2}, dy_{m_2-1}\right) \cdots \\
&\quad \cdot P\left(1, y_{m_1}, dy_{m_1+1}\right) P(m_1 + (n-1)\tau, Y(-(n-1)\tau), dy_{m_1}),
\end{aligned}
$$

Set

$$
\begin{aligned}
f(y_{m1}) := \int_{y_{m_1+1}} \cdots \int_{y_{m_2-1}} & P\left(1, y_{m_2-1}, B^{-\sum_{j=m_1}^{m_2-1} y_j}\right) \\
& P\left(1, y_{m_2-2}, dy_{m_2-1}\right) \cdots P\left(1, y_{m_1}, dy_{m_1+1}\right).
\end{aligned}
$$

As $P(\cdot,\cdot,\cdot)$ is a probability measure, and Condition A^* holds, f is also a bounded continuous function w.r.t. y_{m1}. Then by (2.4.2) and Condition B^*,

$$
\begin{aligned}
&|P(X(-n) \in A, X(0) \in B) - P(X(-n) \in A)P(X(0) \in B)| \\
\leq\ & \mathbb{E}\left|\mathbb{E}\left[I_B(X(0)) - \mathbb{E}I_B(X(0)) \,\middle|\, \mathcal{F}_{-\infty}^{-(n-1)}\right]\right| \\
=\ & \mathbb{E}\left|\int_M f(y_{m1}) P(m_1 + (n-1)\tau, Y(-(n-1)\tau), dy_{m_1}) - \int_M f(y_{m1}) \rho_{m1}(dy_{m_1})\right| \\
=\ & \int_M \left|\int_M f(y_{m1}) P(m_1 + (n-1)\tau, y, dy_{m_1}) - \int_M f(y_{m1}) \rho_{m1}(dy_{m_1})\right| \\
& \quad \cdot P(\omega : Y(-(n-1)\tau, \omega) \in dy) \\
=\ & \int_M \left|\int_M f(y_{m1}) P(m_1 + (n-1)\tau, y, dy_{m_1}) - \int_M f(y_{m1}) \rho_{m1}(dy_{m_1})\right| \rho_0(dy) \\
<\ & n^{-(\frac{4}{\delta}+2+\epsilon)},
\end{aligned}
$$

i.e. $\alpha(n) < n^{-(\frac{4}{\delta}+2+\epsilon)}$. So $\sum_{n=1}^{\infty} \alpha(n)^{\frac{\delta}{2(2+\delta)}} < \infty$. As $\{X(u)\}_{u=0}^{\infty}$ is an ergodic stationary process by the result in [20], Theorem 2.4.3 implies that $\frac{S_n(X)}{\sqrt{n}} \to Z$ in distribution as $n \to \infty$. $\square$

Now we are ready to prove Theorem 2.4.4.

Proof. Set $m = [\frac{n}{\tau}]$ to be the number of the complete periods of the data up to time t. Let $X_i = \sum_{j=1}^{\tau} Y(j + i\tau)$, then by lemma 2.4.5,

$$\frac{1}{\sqrt{m}} \sum_{i=0}^{m-1} X_i \to Z \text{ in distribution as } m \to \infty,$$

where $Z \sim N(0, \sigma^2)$ and $\sigma^2 = \mathbb{E}X_0^2 + 2\sum_{n=1}^{\infty} \mathbb{E}X_0 X_n$. Also

$$\frac{\sum_{i=0}^{m-1} X_i}{\sqrt{n}} = \frac{\sqrt{m}}{\sqrt{n}} \frac{\sum_{i=0}^{m-1} X_i}{\sqrt{m}} \longrightarrow Z_\tau \text{ in distribution as } n \to \infty,$$

where $Z_\tau \sim N(0, \frac{\sigma^2}{\tau})$. Since

$$\frac{\sum_{i=1}^{n} Y(i)}{\sqrt{n}} = \frac{\sum_{i=0}^{m-1} X_i}{\sqrt{n}} + \frac{\sum_{i=(m-1)\tau+1}^{n} Y(i)}{\sqrt{n}},$$

and noting that $0 \leq n - (m-1)\tau - 1 < \tau$, so

$$n^{-\frac{1}{2}} \sum_{i=(m-1)\tau+1}^{n} Y(i) \to 0$$

in probability as $n \to \infty$. Hence

$$n^{-\frac{1}{2}} \sum_{i=1}^{n} Y(i) \to Z_\tau$$

in distribution as $n \to \infty$ as well. $\square$

Chapter 3

Casual Case

In this chapter, we followed the idea of stationary process case to prove the convergence of coefficients for AR(p) model. But for random periodic process case, we fixed time t and only consider the same time point in each period till $-\infty$ time. This constructed the corresponding sequence $\{Y(t-k\tau)\}_{k=0,1,\cdots}$ of t, which can be regarded as stationary. The convergence of coefficients has similar form with that in stationary case, but the distribution has periodic variance. We then derived the Durbin-Levinson Algorithm for random periodic case.

For MA(d) model, we considered innovation representation of $Y(t)$, similarly as stationary case. We followed the steps in [11] of proving the convergence of the estimated innovation coefficients $\hat{b}_j(t)$ to the real values $\frac{\psi_j(t)}{\psi_0(t)}$. That is, we first proved the convergence of theoretic innovation coefficients $b_j(t)$ to the real values, and then proved the convergence of $\hat{b}_j(t)$ to $b_j(t)$. In the second step, we learnt the idea from [4] that we took one period coefficients as a $d \times \tau$-vector $\boldsymbol{B}_t$ and proved the convergence of one-period vector rather that one-time vector in stationary case. The convergence result is similar with the one in [4]. However, some settings are confusing in the proof of [4], but our results are derived smoothly by the periodicity of coefficients proved by the property of random periodic processes in Section 2.2 and the convergence result in the first step.

By the limitation of the window length W in real problem approximation, the rank of sample covariance matrix Γ_N is no larger than W, otherwise Γ_N will be singular. To avoid the singularity, we used truncated innovation presentation of $Y(t)$ to approximate coefficients ψ. Hence we modified the innovation algorithm to truncated innovation algorithm in Section 3.4.

3.1 AR(p) Model for Random Periodic Processes

For fixed $t \in \mathbb{Z}^+$ and $h \in \mathbb{Z}$, define

$$\gamma_n^*(t, t+h) := \frac{1}{n} \sum_{k=0}^{n-1} Y(t - k\tau) Y(t + h - k\tau). \tag{3.1.1}$$

The sequence $\{Y(t - k\tau)Y(t + h - k\tau)\}_{k \in \mathbb{Z}}$ is a stationary process. We then derived the convergence of γ_n^* to γ. The proof is suitable for both causal and non-causal cases.

Assume random periodic process $\{Y(t)\}$ can be written as a moving average random periodic process, i.e.

$$Y(t) = \sum_{i=-\infty}^{+\infty} \psi_i(t) Z(t - i). \tag{3.1.2}$$

Then by a similar proof of Proposition 7.3.5 in [12], one obtains that $\gamma_n^*(t, t+h)$ converges to $\gamma(t, t+h) = \sum_{i=-\infty}^{+\infty} \psi_i(t)\psi_{i+h}(t+h)$ in probability as $n \to \infty$.

We will use the following lemmas (Proposition 6.3.5, Proposition 6.3.9 and Proposition 6.3.10 in [12]) in the proof of the convergence of γ.

Lemma 3.1.1. *Let $\boldsymbol{X}_n$, $n = 1, 2, \ldots$ be random k-vectors. If $\boldsymbol{X}_n \to \boldsymbol{b}$ in distribution as $n \to \infty$ where $\boldsymbol{b}$ is a constant k-vector, then $\boldsymbol{X}_n \to \boldsymbol{b}$ in probability as $n \to \infty$.*

Lemma 3.1.2. *Let $\boldsymbol{X}_n$, $n = 1, 2, \ldots,$ and $\mathbf{Y}_{n,j}$, $j = 1, 2, \ldots; n = 1, 2, \ldots,$ be random k-vectors such that*

i) $\mathbf{Y}_{n,j} \to \mathbf{Y}_j$ in distribution as $n \to \infty$ for each $j = 1, 2, \ldots,$

ii) $\mathbf{Y}_j \to \boldsymbol{Y}$ in distribution as $j \to \infty$, and

iii) $\lim_{j\to\infty} \limsup_{n\to\infty} P(|\boldsymbol{X}_n - \mathbf{Y}_{n,j}| > \epsilon) = 0$ for every $\epsilon > 0$.

Then $\boldsymbol{X}_n \to \boldsymbol{Y}$ in distribution as $n \to \infty$.

Lemma 3.1.3. *(WLLN for Moving Averages) Let $\{X(t)\}$ be the two-sided moving average*

$$X(t) = \sum_{i=-\infty}^{\infty} \psi_i Z(t - i),$$

where $\{Z(t)\}$ *is i.i.d. with mean* μ *and* $\sum_{i=-\infty}^{\infty}|\psi_i| < \infty$. *Then*

$$\frac{1}{n}\sum_{t=1}^{n} X(t) \to \left(\sum_{i=-\infty}^{\infty}\psi_i\right)\mu$$

in probability as $n \to \infty$.

Proposition 3.1.4. *Suppose random periodic process* $\{Y(t)\}$ *satisfy (3.1.2), where* $Z(t) \sim IID(0,1)$. *Furthermore, for any* $t = 1, 2, \ldots, \tau$, *and for any* $h \in \mathbb{Z}$, *suppose* $\sum_{i=-\infty}^{+\infty}|\psi_i(t)\psi_{i+h}(t)| < \infty$. *Then for any* $h \in \mathbb{Z}$,

$$\gamma_n^*(t, t+h) \to \gamma(t, t+h)$$

in probability as $n \to \infty$.

Proof. Note that

$$\begin{aligned}
&\gamma(t, t+h)\\
=\ & \mathbb{E}Y(t)Y(t+h)\\
=\ & \mathbb{E}\left[\left(\sum_{i=-\infty}^{+\infty}\psi_i(t)Z(t-i)\right)\left(\sum_{j=-\infty}^{+\infty}\psi_j(t+h)Z(t+h-j)\right)\right]\\
=\ & \mathbb{E}\left[\left(\sum_{i=-\infty}^{+\infty}\psi_i(t)Z(t-i)\right)\left(\sum_{j=-\infty}^{+\infty}\psi_{j+h}(t+h)Z(t-j)\right)\right]\\
=\ & \mathbb{E}\left[\sum_{i,j=-\infty}^{+\infty}\psi_i(t)\psi_{j+h}(t+h)Z(t-i)Z(t-j)\right]\\
=\ & \sum_{i=-\infty}^{+\infty}\psi_i(t)\psi_{i+h}(t+h),
\end{aligned}$$

since for $i \neq j$, $\mathbb{E}[Z(t-i)Z(t-j)] = 0$. For fixed t and h,

$$\begin{aligned}
&\gamma_n^*(t, t+h)\\
=\ & \frac{1}{n}\sum_{k=0}^{n-1} Y(t-k\tau)Y(t+h-k\tau)\\
=\ & \frac{1}{n}\sum_{k=0}^{n-1}\left(\sum_{i=-\infty}^{+\infty}\psi_i(t-k\tau)Z(t-k\tau-i)\right)\left(\sum_{j=-\infty}^{+\infty}\psi_j(t+h-k\tau)Z(t+h-k\tau-j)\right)\\
=\ & \frac{1}{n}\sum_{k=0}^{n-1}\left(\sum_{i=-\infty}^{+\infty}\psi_i(t)Z(t-k\tau-i)\right)\left(\sum_{j=-\infty}^{+\infty}\psi_j(t+h)Z(t+h-k\tau-j)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{k=0}^{n-1}\left(\sum_{i=-\infty}^{+\infty}\psi_i(t)Z(t-k\tau-i)\right)\left(\sum_{j=-\infty}^{+\infty}\psi_{j+h}(t+h)Z(t-k\tau-j)\right)\\
&= \frac{1}{n}\sum_{k=0}^{n-1}\left(\sum_{i=-\infty}^{+\infty}\psi_i(t)\psi_{i+h}(t+h)Z(t-k\tau-i)^2\right)+\epsilon_n,
\end{aligned}
$$

where

$$
\begin{aligned}
\epsilon_n &:= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{i=-\infty}^{+\infty}\sum_{j\neq i}\psi_i(t)\psi_{j+h}(t+h)Z(t-k\tau-i)Z(t-k\tau-j)\\
&= \sum_{i=-\infty}^{+\infty}\sum_{j\neq i}\psi_i(t)\psi_{j+h}(t+h)\left(\frac{1}{n}\sum_{k=0}^{n-1}Z(t-k\tau-i)Z(t-k\tau-j)\right).
\end{aligned}
$$

As $\{Z(t)^2\}$ are i.i.d. with mean 1 and $\sum_{i=-\infty}^{+\infty}|\psi_i(t)\psi_{i+h}(t+h)|<\infty$, by lemma 3.1.3,

$$
\frac{1}{n}\sum_{k=0}^{n-1}\left(\sum_{i=-\infty}^{+\infty}\psi_i(t)\psi_{i+h}(t+h)Z(t-k\tau-i)^2\right)\longrightarrow\sum_{i=-\infty}^{+\infty}\psi_i(t)\psi_{i+h}(t+h)
$$

in probability as $n\to\infty$.

Moreover, note that when $i\neq j$,

$$
Cov(Z(t-i)Z(t-j),Z(t+h-i)Z(t+h-j))=\begin{cases}1, & h=0,\\ 0, & h\neq 0.\end{cases}\tag{3.1.3}
$$

Hence

$$
\begin{aligned}
&Var\left(\frac{1}{n}\sum_{k=0}^{n-1}Z(t-k\tau-i)Z(t-k\tau-j)\right)\\
&= \frac{1}{n^2}\sum_{k=0}^{n-1}Var\left(Z(t-k\tau-i)Z(t-k\tau-j)\right)\\
&\quad+\sum_{k,h=0,k\neq h}^{n-1}Cov(Z(t-k\tau-i)Z(t-k\tau-j),Z(t-h\tau-i)Z(t-h\tau-j))\\
&= \frac{1}{n^2}\sum_{k=0}^{n-1}1+0\\
&= \frac{1}{n}\to 0,
\end{aligned}
$$

as $n\to\infty$. Define for each positive integer m,

$$
\epsilon_n^m := \sum_{|i|\leq m}\sum_{|j|\leq m,j\neq i}\psi_i(t)\psi_{j+h}(t+h)\left(\frac{1}{n}\sum_{k=0}^{n-1}Z(t-k\tau-i)Z(t-k\tau-j)\right).
$$

By Chebyshev's inequality, for any $\sigma > 0$,

$$\begin{aligned}
& P(|\epsilon_n^m| \geq \sigma) \\
\leq\ & \frac{1}{\sigma^2} Var(\epsilon_n^m) \\
\leq\ & \frac{1}{\sigma^2} \sum_{|i|\leq m} \sum_{|j|\leq m, j\neq i} |\psi_i(t)\psi_{j+h}(t+h)|^2 Var\left(\frac{1}{n}\sum_{k=0}^{n-1} Z(t-k\tau-i)Z(t-k\tau-j)\right).
\end{aligned}$$

Hence $\epsilon_n^m \to 0$ in probability as $n \to \infty$. Furthermore,

$$\begin{aligned}
& \lim_{m\to\infty} \limsup_{n\to\infty} \mathbb{E}|\epsilon_n - \epsilon_n^m| \\
\leq\ & \lim_{m\to\infty} \limsup_{n\to\infty} \left(\sum_{|i|>m} \sum_{|j|>m, j\neq i} + \sum_{|i|\leq m} \sum_{|j|>m} + \sum_{|i|>m} \sum_{|j|\leq m} \right) |\psi_i(t)\psi_{j+h}(t+h)| \\
& \qquad \mathbb{E}\left| \frac{1}{n} \sum_{k=0}^{n-1} Z(t-k\tau-1)Z(t-k\tau-2) \right| \\
=\ & 0.
\end{aligned}$$

Thus by lemma 3.1.1 and 3.1.2, $\epsilon_n \to 0$ in probability as $n \to \infty$. Together with the first part of the proof , we can obtain the final convergence of $\gamma_n^*(t, t+h)$. □

Now let us consider AR(p) model. As $\phi_i(t) = \phi_i(t+\tau)$ for every $i = 1, 2, \cdots, p$, write the AR(p) equation (2.2.2) in the form of

$$\boldsymbol{Y}_{t,n} = \boldsymbol{X}_{t,n}\boldsymbol{\phi}_p(t) + \theta_0^t \boldsymbol{Z}_{t,n},$$

where

$$\boldsymbol{Y}_{t,n} := (Y(t), Y(t-\tau), \cdots, Y(t-(n-1)\tau))^T,$$

$$\boldsymbol{X}_{t,n} := \begin{bmatrix} Y(t-1) & Y(t-2) & \cdots & Y(t-p) \\ Y(t-1-\tau) & Y(t-2-\tau) & \cdots & Y(t-p-\tau) \\ \vdots & \vdots & \ddots & \vdots \\ Y(t-1-(n-1)\tau) & Y(t-2-(n-1)\tau) & \cdots & Y(t-p-(n-1)\tau) \end{bmatrix},$$

$$\boldsymbol{\phi}_p(t) := (\phi_1(t), \phi_2(t), \cdots, \phi_p(t))^T,$$

and

$$\boldsymbol{Z}_{t,n} := (Z(t), Z(t-\tau), \cdots, Z(t-(n-1)\tau))^T.$$

Considering that the respective components of $Y(t-1), Y(t-2), \cdots, Y(t-(n-1)\tau)$ in $\boldsymbol{X}_{t,n}$ are independent of the respective components of $\boldsymbol{Z}_{t,n}$, it is suggested that the linear regression estimate $\boldsymbol{\phi}_p^n(t)$ of $\boldsymbol{\phi}_p(t)$ is defined by

$$\boldsymbol{\phi}_p^n(t) := (\boldsymbol{X}_{t,n}^T \boldsymbol{X}_{t,n})^{-1} \boldsymbol{X}_{t,n}^T \boldsymbol{Y}_{t,n}. \tag{3.1.4}$$

The $(i,j)^{th}$-component of $\frac{1}{n}\boldsymbol{X}_{t,n}^T \boldsymbol{X}_{t,n}$ is

$$\frac{1}{n}\sum_{k=0}^{n-1} Y(t-i-k\tau)Y(t-j-k\tau),$$

and the i^{th} component of $\frac{1}{n}\boldsymbol{X}_{t,n}^T \boldsymbol{Y}_{t,n}$ is equal to

$$\frac{1}{n}\sum_{k=0}^{n-1} Y(t-k\tau)Y(t-i-k\tau).$$

Then by Proposition 3.1.4,

$$\frac{1}{n}\boldsymbol{X}_{t,n}^T \boldsymbol{X}_{t,n} \to \Gamma_p(t) \tag{3.1.5}$$

in probability and

$$\frac{1}{n}\boldsymbol{X}_{t,n}^T \boldsymbol{Y}_{t,n} \to \boldsymbol{\gamma}_p(t) \tag{3.1.6}$$

in probability as $n \to \infty$, where

$$\Gamma_p(t) := (\gamma(t-i, t-j))_{i,j=1,\cdots,p},$$

and

$$\boldsymbol{\gamma}_p(t) := (\gamma(t,t-1), \gamma(t,t-2), \cdots, \gamma(t,t-p))^T.$$

By using a similar method of stationary process shown in [12], in the next proposition we can prove that for fixed t, $\sqrt{n}\left(\boldsymbol{\phi}_p^n(t) - \boldsymbol{\phi}_p\right) \to V_t$ in distribution as $n \to \infty$, where $V_t \sim N(\boldsymbol{0}, \theta_0(t)^2\Gamma_p^{-1}(t))$. We will use the following two lemmas in [12] in the proof.

Lemma 3.1.5. *If $\{\boldsymbol{X}_n\}$ and $\{\boldsymbol{Y}_n\}$ are sequences of random k-vectors such that $\boldsymbol{X}_n \to \boldsymbol{X}$ in probability and $\boldsymbol{Y}_n \to \boldsymbol{b}$ in distribution as $n \to \infty$, where $\boldsymbol{b}$ is constant, then*

$$\boldsymbol{X}_n + \boldsymbol{Y}_n \to \boldsymbol{X} + \boldsymbol{b}$$

and

$$\boldsymbol{Y}_n^T \boldsymbol{X}_n \to \boldsymbol{b}^T \boldsymbol{X}$$

in distribution as $n \to \infty$.

Lemma 3.1.6. *Let $\{\boldsymbol{X}_n\}$ be a sequence of random k-vector and Σ_n be the covariance matrix of $\boldsymbol{X}_n$. As Σ_n is symmetric, suppose it can be decomposed as $\Sigma_n = Q_n\Lambda_n\Lambda_n Q_n^T$, where Λ_n is a diagonal matrix with square root of eigenvalues of Σ_n on the diagonal. If $(Q_n\Lambda_n)^{-1}(\boldsymbol{X}_n - \boldsymbol{\mu}) \to \boldsymbol{Z}$ in distribution as $n \to \infty$ where $\boldsymbol{Z} \sim N(\boldsymbol{0}, I)$ with I to be the identity matrix, and B is any non-zero $m \times k$ matrix such that the matrices $B\Sigma_n B^T$, $n = 1, 2, \cdots$, have no zero diagonal elements, then*

$$(BQ_n\Lambda_n)^{-1}(B\boldsymbol{X}_n - B\boldsymbol{\mu}) \to \boldsymbol{Z}$$

in distribution as $n \to \infty$.

Proposition 3.1.7. *Assume that random periodic process $\{Y(t)\}_{t\in\mathbb{Z}}$ satisfies AR(p) equation (2.2.2), where $Z(t) \sim IID(0,1)$. For fixed t, define $\boldsymbol{\phi}_p^n(t)$ as (3.1.4), then*

$$\sqrt{n}\left(\boldsymbol{\phi}_p^n(t) - \boldsymbol{\phi}_p(t)\right) \to V_t$$

in distribution as $n \to \infty$, where $V_t \sim N(\boldsymbol{0}, \theta_0(t)^2\Gamma_p^{-1}(t))$.

Proof. From the definition of $\boldsymbol{\phi}_p^n(t)$,

$$\begin{aligned}
&\sqrt{n}(\boldsymbol{\phi}_p^n(t) - \boldsymbol{\phi}_p(t)) \\
=\ &\sqrt{n}\left[(\boldsymbol{X}_t^{n\prime}\boldsymbol{X}_t^n)^{-1}\boldsymbol{X}_{t,n}^T\boldsymbol{Y}_{t,n} - \boldsymbol{\phi}_p(t)\right] \\
=\ &\sqrt{n}\left[(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1}\boldsymbol{X}_{t,n}^T(\boldsymbol{X}_{t,n}\boldsymbol{\phi}_p(t) + \theta_0(t)\boldsymbol{Z}_{t,n}) - \boldsymbol{\phi}_p(t)\right] \\
=\ &n\theta_0(t)(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1}(\frac{1}{\sqrt{n}}\boldsymbol{X}_{t,n}^T\boldsymbol{Z}_{t,n}).
\end{aligned}$$

Set $\boldsymbol{U}_t := (Y(t-1)Z(t), Y(t-2)Z(t), \cdots, Y(t-p)Z(t))^T$, then

$$\frac{1}{\sqrt{n}}\boldsymbol{X}_{t,n}^T\boldsymbol{Z}_{t,n} = \frac{1}{\sqrt{n}}\sum_{j=0}^{n-1}\boldsymbol{U}_{t-j\tau}.$$

By the assumption of causality, $Y(t)$ can be represented by $Y(t) = \sum_{j=0}^{+\infty}\psi_j(t)Z(t-j)$. Hence $Z(t)$ is independent of $(Y(t-1), Y(t-2), \cdots, Y(t-p))^T$. Then $\{\boldsymbol{U}_{t-j\tau}\}_{j\in\mathbb{Z}}$ is a sequence of stationary process with $\mathbb{E}\boldsymbol{U}_t = \boldsymbol{0}$ and

$$\mathbb{E}\boldsymbol{U}_t\boldsymbol{U}_{t+h}^T = \begin{cases} \Gamma_p(t), & h = 0, \\ \boldsymbol{0}_{p\times p}, & h \neq 0. \end{cases}$$

For fixed t, define $\mathcal{F}_n^t := \sigma(\boldsymbol{U}_{t+m\tau}, m \leq n)$. Then for any $n \geq 1$,

$$\begin{aligned}
&\mathbb{E}[\boldsymbol{U}_t|\mathcal{F}_{-n}^t] = \mathbb{E}[\boldsymbol{U}_t|\boldsymbol{U}_{t-n\tau}] \\
=\ &\mathbb{E}\left[(Y(t-1),\ldots,Y(t-p))^T Z(t) \mid (Y(t-1-n\tau),\cdots,Y(t-p-n\tau))^T Z(t-n\tau)\right] \\
=\ &\mathbb{E}[Z(t)]\mathbb{E}\left[(Y(t-1),\cdots,Y(t-p))^T \mid (Y(t-1-n\tau),\cdots,Y(t-p-n\tau))^T Z(t-n\tau)\right] \\
=\ &0.
\end{aligned}$$

By the central limit theorem of stationary processes,

$$\frac{1}{\sqrt{n}}\boldsymbol{X}_{t,n}^T \boldsymbol{Z}_{t,n} \to S_t$$

in distribution as $n \to \infty$, where $S_t \sim N(\boldsymbol{0}, \Gamma_p(t))$. Besides, by (3.1.5), $n(\boldsymbol{X}_{t,n}^T \boldsymbol{X}_{t,n})^{-1} \to \Gamma_p^{-1}(t)$ in probability. Hence by Lemma 3.1.5 and 3.1.6, one can obtain the result that

$$\sqrt{n}\left(\boldsymbol{\phi}_p^n(t) - \boldsymbol{\phi}_p(t)\right) \to V_t,$$

in distribution as $n \to \infty$, where $V_t \sim N(\boldsymbol{0}, \theta_0(t)^2 \Gamma_p^{-1}(t))$. □

Remark 3.1.8. *If we use the weak law of large number of stationary process instead of the central limit theorem in the proof above, we obtain that $\boldsymbol{\phi}_t^n - \boldsymbol{\phi}_t \to 0$ in probability. The details are as follows.*

$$\boldsymbol{\phi}_p^n(t) - \boldsymbol{\phi}_p(t) = n\theta_0(t)(\boldsymbol{X}_{t,n}^T \boldsymbol{X}_{t,n})^{-1}(\frac{1}{n}\boldsymbol{X}_{t,n}^T \boldsymbol{Z}_{t,n}).$$

By the weal law of large number of stationary process,

$$\frac{1}{n}\boldsymbol{X}_{t,n}^T \boldsymbol{Z}_{t,n} = \frac{1}{n}\sum_{j=0}^{n-1} \boldsymbol{U}_{t-j\tau} \to \mathbb{E}\boldsymbol{U}_t = \boldsymbol{0}$$

in probability as $n \to \infty$. Hence $\boldsymbol{\phi}_t^n \to \boldsymbol{\phi}_t$ in probability as $n \to \infty$.

Remark 3.1.9. *In real life calculations, in order to avoid taking values at negative time, we let n go to infinity along with t tending to infinity. That is to say, we can regard $n \to \infty$ as $t \to \infty$.*

With the concern of the previous remark, it is required that we use finite periods of data to estimate the covariance function within certain error permission in real life calculations. That is the reason why we assume Assumption 1.

To ensure the estimation of $\phi(t)$ by the finite-periods estimation of covariance has the same asymptotic normality as the infinite one, i.e. $\phi^n(t)$, we assume the following assumption:

Assumption 2. *For any small positive ϵ_1 and ϵ_2, there exists an m_0 such that, for $m_1 > m_0$,*

$$P\left(m_1^{-\frac{1}{2}}\left|\frac{1}{m_1}\sum_{k=0}^{m_1-1}Y(t-k\tau)Y(s-k\tau)-\frac{1}{m_0}\sum_{k=0}^{m_0-1}Y(t-k\tau)Y(s-k\tau)\right|>\epsilon_1\right)<1-\epsilon_2, \tag{3.1.7}$$

for any $t,s\in\mathbb{Z}^+$.

This assumption shows the possibility that under certain error permission, one can use the average of a finite-length window of $\{Y(t-k\tau)Y(s-k\tau)\}_{k\in\mathbb{Z}}$ to approximate the average of the whole sequence. Most time series in real problems have finite second moment. Also, this assumption seems quite reasonable and common in approximation.

Then we can set $w = max(n_0, m_0)$ to estimate the sample autocovriance function $\hat{\gamma}(\cdot,\cdot)$ in the real world problem. Denote $\hat{\boldsymbol{\phi}}$ as the Yule-Walkers estimator obtained by the sample autocovariance function $\hat{\gamma}$, i.e. $\hat{\boldsymbol{\phi}}_p(t) = (\hat{\Gamma}_p^t)^{-1}\hat{\boldsymbol{\gamma}}_p(t)$. Then

$$\begin{aligned}
& n^{-\frac{1}{2}}\left(\hat{\boldsymbol{\phi}}_p(t)-\boldsymbol{\phi}_p^n(t)\right)\\
= {} & n^{-\frac{1}{2}}(\hat{\Gamma}_p^t)^{-1}\hat{\boldsymbol{\gamma}}_p(t)-n^{-\frac{1}{2}}(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1}\boldsymbol{X}_{t,n}^T\boldsymbol{Y}_{t,n}\\
= {} & (\hat{\Gamma}_p^t)^{-1}n^{-\frac{1}{2}}(\hat{\boldsymbol{\gamma}}_p(t)-\frac{1}{n}\boldsymbol{X}_{t,n}^T\boldsymbol{Y}_{t,n})+n^{-\frac{1}{2}}\left[(\hat{\Gamma}_p^t)^{-1}-n(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1}\right]\frac{1}{n}\boldsymbol{X}_{t,n}^T\boldsymbol{Y}_{t,n}.
\end{aligned}$$

From Assumption 1 we have $\hat{\boldsymbol{\gamma}}_p(t)\to\boldsymbol{\gamma}_p(t)$ and $\hat{\Gamma}_p^t\to\Gamma_p(t)$ in probability. Assumption 2 implies the convergence of $n^{-\frac{1}{2}}(\hat{\boldsymbol{\gamma}}_p(t)-\frac{1}{n}\boldsymbol{X}_{t,n}^T\boldsymbol{Y}_{t,n})$ to zero in probability as $n\to\infty$. Set $|\cdot|$ to be the Euclidean norm.

$$\begin{aligned}
& n^{-\frac{1}{2}}\left|(\hat{\Gamma}_p^t)^{-1}-n(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1}\right|\\
= {} & n^{-\frac{1}{2}}\left|(\hat{\Gamma}_p^t)^{-1}\left(\frac{1}{n}\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n}-\hat{\Gamma}_p^t\right)n(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1}\right|\\
\le {} & n^{-\frac{1}{2}}\left|(\hat{\Gamma}_p^t)^{-1}\right|n^{-\frac{1}{2}}\left|\frac{1}{n}\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n}-\hat{\Gamma}_p^t\right|\left|n(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1}\right|.
\end{aligned}$$

Assumption 2 implies $n^{-\frac{1}{2}}\left|\frac{1}{n}\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n}-\hat{\Gamma}_p^t\right|\to 0$ in probability as $n\to\infty$. Then by $\hat{\Gamma}_p^t\to\Gamma_p(t)$ and (3.1.5), we have $n^{-\frac{1}{2}}\left|(\hat{\Gamma}_p^t)^{-1}-n(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1}\right|\to 0$ in probability as

$n \to \infty$. As we also have (3.1.6), we conclude that $n^{-\frac{1}{2}} \left(\hat{\boldsymbol{\phi}}_p(t) - \boldsymbol{\phi}_p^n(t) \right) \to \mathbf{0}$ in probability as $n \to \infty$. Thus $\sqrt{n} \left(\hat{\boldsymbol{\phi}}_p(t) - \boldsymbol{\phi}_p(t) \right) \to V_t$ in distribution as well.

3.2 Durbin-Levinson Algorithm

For the casual AR(p) model (2.2.2), if we replace the covariance function $\gamma(t,s)$ in (2.2.5) by the sample covariance function of $\{Y(t)\}$ and solve the corresponding linear system, then the solution $\hat{\boldsymbol{\phi}}_p(t) = (\hat{\phi}_1(t), \hat{\phi}_2(t), \cdots, \hat{\phi}_p(t))^T$ is called the Yule-Walker estimator of $\boldsymbol{\phi}_p(t)$. From the above section, we have obtained the asymptotic behaviour of $\hat{\boldsymbol{\phi}}_p(t)$. Moreover, by [12], the Durbin-Levinson algorithm is used to solve the linear equation system $\Gamma_n \boldsymbol{\phi}_n = \boldsymbol{\gamma}_n$ recursively without solving the inverse of coefficients matrix. Hence this algorithm can be used to obtain $\hat{\boldsymbol{\phi}}_p(t)$.

Suppose we know the history data $\{y(1), y(2), \cdots, y(n)\}$. We would like to estimate $y(n+1)$ from them.

For spanned linear subspace of $\{1, Y(1), Y(2), \cdots, Y(n)\}$, we know that

$$\begin{aligned} & span(1, Y(1), Y(2), \cdots, Y(n)) \\ = \; & span(1) \oplus span\left((Y(1) - \mathbb{E}[Y(1)|span(1)]), (Y(2) - \mathbb{E}[Y(2)|span(1)]), \cdots, \right. \\ & \quad \left. (Y(n) - \mathbb{E}[Y(n)|span(1)]) \right), \end{aligned}$$

where $\mathbb{E}[Y_s|span(1)] = \mathbb{E}[Y(s)] =: \mu_s$. Without loss of generality, assume for any t, $\mathbb{E}Y(t) = 0$. Define $\mathcal{H}_1^n := \sigma\{Y(1), Y(2), \cdots, Y(n)\}$, we take $\mathbb{E}[Y(n+1)|\mathcal{H}_1^n]$ as the one-step best linear estimator of $Y(n+1)$ with respect to the history data $\{Y(1), Y(2), \cdots, Y(n)\}$. In the Durbin-Levinson Algorithm for stationary process, suppose for any n, the one-step linear estimator of $Y(n+1)$ with respect to $\{Y(1), Y(2), \cdots, Y(n)\}$ satisfies

$$\hat{Y}(n+1) = \mathbb{E}[Y(n+1)|\mathcal{H}_1^n] = \sum_{j=1}^{n} a_j(n+1) Y(n+1-j),$$

where the coefficients $\{a_j(n+1)\}$ are obtained by solving

$$\mathbb{E}[\hat{Y}(n+1)Y(j)] = \mathbb{E}[Y(n+1)Y(j)]$$

for all $j = 1, 2, \cdots, n$. In the Durbin-Levinson Algorithm for stationary process, $\{a_j(n+1)\}$ are calculated by iteration without calculating the inverse of matrix. We would like to find the corresponding algorithm for random periodic processes.

For $n > \tau$, define

$$\mathcal{H}_{\tau+1}^n := \sigma(Y(\tau+1), Y(\tau+2), \cdots, Y(n)),$$

and for $k = 1, 2, \cdots, \tau$, define

$$\begin{cases} \mathcal{H}_1 := \sigma(Y(1)), \\ \mathcal{H}_k := \sigma\left(Y(k) - \sum_{j=1}^{k-1} \mathbb{E}[Y(k)|\mathcal{H}_j] - \mathbb{E}[Y(k)|\mathcal{H}_{\tau+1}^n]\right), k = 2, \cdots, \tau. \end{cases}$$

Then

$$\hat{Y}(n+1) = \mathbb{E}[Y(n+1)|\mathcal{H}_1^n] = \mathbb{E}[Y(n+1)|\mathcal{H}_1] + \cdots + \mathbb{E}[Y(n+1)|\mathcal{H}_\tau] + \mathbb{E}[Y(n+1)|\mathcal{H}_{\tau+1}^n]. \tag{3.2.1}$$

Suppose

$$\hat{Y}(n+1) = \sum_{j=1}^{n} a_j(n+1) Y(n+1-j). \tag{3.2.2}$$

The coefficients are obtained by the property of conditional expectation: for any $i = 1, 2, \cdots, n$,

$$\mathbb{E}[\hat{Y}(n+1)Y(i)] = \mathbb{E}[Y(n+1)Y(i)]. \tag{3.2.3}$$

Suppose $\mathbb{E}[Y(n+1)|\mathcal{H}_{\tau+1}^n] = \sum_{j=1}^{n-\tau} \alpha_j^{n+1} Y_{n+1-j}$. By the equation above, the coefficient

$$\boldsymbol{\alpha}_{n-\tau}(n+1) := (\alpha_{n-\tau}(n+1), \alpha_{n-\tau-1}(n+1), \cdots, \alpha_1(n+1))^T$$

is the solution of the following linear system

$$\boldsymbol{\gamma}_{n-\tau}(n+1) = \Gamma_{n-\tau}(n+1)\boldsymbol{\alpha}_{n-\tau}(n+1),$$

where

$$\boldsymbol{\gamma}_{n-\tau}(n+1) := (\gamma(n+1, 1+\tau), \gamma(n+1, 2+\tau), \cdots, \gamma(n+1, n))^T,$$

and

$$\Gamma_{n-\tau}(n+1) := \begin{bmatrix} \gamma(1+\tau, 1+\tau) & \gamma(1+\tau, 2+\tau) & \cdots & \gamma(1+\tau, n) \\ \gamma(2+\tau, 1+\tau) & \gamma(2+\tau, 2+\tau) & \cdots & \gamma(2+\tau, n) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n, 1+\tau) & \gamma(n, 2+\tau) & \cdots & \gamma(n, n) \end{bmatrix}.$$

For $\hat{Y}(n+1-\tau) = \sum_{j=1}^{n-\tau} a_j(n+1-\tau)Y(n+1-\tau-j)$, the coefficients

$$\boldsymbol{a}_{n-\tau}(n+1-\tau) := (a_{n-\tau}(n+1-\tau), \cdots, a_1(n+1-\tau))^T$$

is the solution of the linear system

$$\boldsymbol{\gamma}_{n-\tau}(n+1-\tau) = \Gamma_{n-\tau}(n+1-\tau)\boldsymbol{a}_{n-\tau}(n+1-\tau),$$

where

$$\boldsymbol{\gamma}_{n-\tau}(n+1-\tau) := (\gamma(n+1-\tau,1), \gamma(n+1-\tau,2), \cdots, \gamma(n+1-\tau,n-\tau))^T,$$

and

$$\Gamma_{n-\tau}(n+1-\tau) := \begin{bmatrix} \gamma(1,1) & \gamma(1,2) & \cdots & \gamma(1,n-\tau) \\ \gamma(2,1) & \gamma(2,2) & \cdots & \gamma(2,n-\tau) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-\tau,1) & \gamma(n-\tau,2) & \cdots & \gamma(n-\tau,n-\tau) \end{bmatrix}.$$

As the covariance function of random periodic process has the periodic property

$$\gamma(t,s) = \gamma(t+\tau, s+\tau),$$

we obtain that $\boldsymbol{\alpha}_{n-\tau}(n+1) = \boldsymbol{a}_{n-\tau}(n+1-\tau)$. Hence

$$\mathbb{E}[Y(n+1)|\mathcal{H}_{\tau+1}^n] = \sum_{j=1}^{n-\tau} a_j(n+1-\tau)Y(n+1-j).$$

For each $k = 2, \cdots, \tau$, suppose

$$\mathbb{E}[Y_{n+1}|\mathcal{H}_k] = \alpha_{n+1-k}(n+1)\left(Y(k) - \sum_{j=1}^{k-1}\mathbb{E}[Y(k)|\mathcal{H}_j] - \mathbb{E}[Y(k)|\mathcal{H}_{\tau+1}^n]\right),$$

where

$$\alpha_{n+1-k}(n+1) = \frac{\mathbb{E}\left[Y(n+1)\left(Y(k) - \sum_{j=1}^{k-1}\mathbb{E}[Y(k)|\mathcal{H}_j] - \mathbb{E}[Y(k)|\mathcal{H}_{\tau+1}^n]\right)\right]}{\mathbb{E}\left[Y(k) - \sum_{j=1}^{\tau}\mathbb{E}[Y(k)|\mathcal{H}_j] - \mathbb{E}[Y(k)|\mathcal{H}_{\tau+1}^n]\right]^2}.$$

For $k = 1, 2, \cdots, \tau$, assume $\mathbb{E}[Y(k)|\mathcal{H}_{\tau+1}^n] = \sum_{j=1}^{n-\tau} b_{0,j}(n+1,k)Y(n+1-j)$, then by the property of conditional expectation, one can list that for $i = 1+\tau, \cdots, n$,

$$\mathbb{E}[Y(k)Y(i)] = \mathbb{E}\left[\left(\sum_{j=1}^{n-\tau} b_{0,j}(n+1,k)Y(n+1-j)\right)Y(i)\right].$$

Replace it by the covariance function and rewrite in the matrix form, we have

$$\boldsymbol{\gamma}_{n-\tau}(k) = \Gamma_{n-\tau}(n+1)\boldsymbol{b}^0_{n-\tau}(n+1,k),$$

where

$$\boldsymbol{\gamma}_{n-\tau}(k) := (\gamma(k,1+\tau), \gamma(k,2+\tau), \cdots, \gamma(k,n))^T$$

and

$$\boldsymbol{b}^0_{n-\tau}(n+1,k) := (b_{0,1+\tau}(n+1,k), b_{0,2+\tau}(n+1,k), \cdots, b_{0,n}(n+1,k))^T.$$

By solving this linear system one can obtain the coefficients for the conditional expectation of $Y(k)$ with respect to $\mathcal{H}^n_{\tau+1}$. For $\mathcal{H}_i$, $i = 1, 2, \cdots, \tau$, assume for $k = i+1, i+2, \cdots, \tau$,

$$\mathbb{E}[Y(k)|\mathcal{H}_i] = b_i(n+1,k)\left(Y(i) - \sum_{j=0}^{i-1}\mathbb{E}[Y(i)|\mathcal{H}_j]\right),$$

where

$$b_i(n+1,k) = \frac{\mathbb{E}\left[Y(k)\left(Y(i) - \sum_{j=1}^{i-1}\mathbb{E}[Y(i)|\mathcal{H}_j]\right)\right]}{Var\left(Y(i) - \sum_{j=1}^{i-1}\mathbb{E}[Y(i)|\mathcal{H}_j]\right)}.$$

Then by comparing the coefficients of the two equations (3.2.1) and (3.2.2), we can find out the formula for the coefficients $\boldsymbol{a}_{n+1}$ in the form of $\boldsymbol{\alpha}_{n+1}$, $\boldsymbol{\beta}_{n+1}$ and $\boldsymbol{a}_{n+1-\tau}$.

For $n+1=2$, $\hat{Y}(2) = \mathbb{E}[Y(2)|Y(1)] = a_1(2)Y(1)$, where $a_1(2) = \frac{\gamma(1,2)}{\gamma(1,1)}$.

For $0 < n < 1+\tau$,

$$\hat{Y}(n+1) = \mathbb{E}[Y(n+1)|\mathcal{H}^n_1] = \alpha_n(n+1)Y(1) + \sum_{i=2}^{n}\alpha_{n+1-i}(n+1)\left(Y_i - \sum_{j=1}^{i-1}\mathbb{E}[Y(i)|\mathcal{H}_j]\right), \tag{3.2.4}$$

where

$$\begin{cases} \mathcal{H}_1 := \sigma(Y(1)), \\ \mathcal{H}_k := \sigma\left(Y_i - \sum_{j=1}^{i-1}\mathbb{E}[Y(i)|\mathcal{H}_j]\right), \quad k = 2, \cdots, n. \end{cases}$$

By (3.2.3), $\alpha_n(n+1) = \frac{\gamma(n+1,1)}{\gamma(1,1)}$ and

$$\alpha_{n+1-i}(n+1) = \frac{\mathbb{E}\left[Y(n+1)\left(Y(i) - \sum_{j=1}^{i-1}\mathbb{E}[Y(i)|\mathcal{H}_j]\right)\right]}{Var\left(Y(i) - \sum_{j=1}^{i-1}\mathbb{E}[Y(i)|\mathcal{H}_j]\right)}, \qquad i = 2, \cdots, n.$$

For each i, consider $j = 1, \cdots, i-1$, if $j = 1$,

$$\mathbb{E}[Y(i)|\mathcal{H}_j] = \beta_n(n+1, n+1-i)Y(1),$$

where $\beta_n(n+1, n+1-i) = \frac{\gamma(i,1)}{\gamma(1,1)}$; if $j = 2, \cdots, i-1$, assume

$$\mathbb{E}[Y(i)|\mathcal{H}_j] = \beta_{n+1-j}(n+1, n+1-i)\left(Y(j) - \sum_{k=1}^{j-1}\mathbb{E}[Y(j)|\mathcal{H}_k]\right),$$

then

$$\beta_{n+1-j}(n+1, n+1-i) = \frac{\mathbb{E}\left[Y(i)\left(Y(j) - \sum_{k=1}^{j-1}\mathbb{E}[Y(j)|\mathcal{H}_k]\right)\right]}{Var\left(Y(j) - \sum_{k=1}^{j-1}\mathbb{E}[Y(j)|\mathcal{H}_k]\right)}.$$

Hence one could compare the two equations of (3.2.2) and (3.2.4) and find out the formula of coefficients $\boldsymbol{a}_{n+1}$ in the form of $\boldsymbol{\alpha}_{n+1}$ and $\boldsymbol{\beta}_{n+1}$ as well.

3.3 MA Model for Random Periodic Processes

Recall the definition of invertible ARMA process.

Definition 3.3.1. *An ARMA process for random periodic processes (2.2.1) is said to be invertible if there exists a sequence of functions $\psi_0(t)$ and $\{\pi_j(t)\}$ such that $\sum_{j=0}^{\infty}|\pi_j(t)| < \infty$ and*

$$\psi_0(t)Z(t) = \sum_{j=0}^{\infty}\pi_j(t)Y(t-j), \quad t \in \mathbb{Z}^+. \tag{3.3.1}$$

Now let us consider another simple model, moving average model, i.e.

$$Y(t) = \sum_{j=0}^{\infty}\psi_j(t)Z(t-j), \quad t \in \mathbb{Z}^+, \tag{3.3.2}$$

where $\{Z(t)\}$ is an i.i.d. sequence of random variables with mean 0 and variance 1, and the coefficients $\{\psi_j(t)\}_{j=0.1,\cdots}$ are periodic functions of t with the period τ as proved in Proposition 2.2.2. Suppose that $\psi_t(z) := \sum_{j=0}^{\infty}\psi_j(t)z^j$ is non-zero for all $z \in \mathbb{C}$ such that $|z| \leq 1$ and $\sum_{j=0}^{\infty}|\psi_j(t)| < \infty$. This means that the MA process is invertible, i.e. it has the representation (3.3.1), where $\pi_0(t) = 1$ and $\pi_t(z) := \sum_{j=0}^{t}\pi_j(t)z^j = \frac{1}{\psi_t(z)}$, $|z| \leq 1$ ([12]).

The detail of the relation between the coefficients ψ and π is as follows. Define the shift operator B such that $B^jY(t) = Y(t-j)$ and $B^j\psi_i(t) = \psi_i(t-j)$ for $j = 0, \pm1, \pm2, \cdots$.

Hence the equations (3.3.2) and (3.3.1) can be represented as

$$\psi_0(t)Z(t) = \sum_{j=0}^{\infty} \pi_j(t)B^j(Y(t)),$$

and

$$\begin{aligned} Y(t) = & \sum_{j=0}^{\infty} \psi_j(t)B^j(Z(t)) \\ = & \sum_{j=0}^{\infty} \psi_j(t)B^j\left(\frac{1}{\psi_0(t)}\sum_{i=0}^{\infty}\pi_i(t)Y(t-i)\right) \\ = & \sum_{j=0}^{\infty} \frac{\psi_j(t)}{\psi_0(t-j)}\sum_{i=0}^{\infty}\pi_i(t-j)Y(t-i-j), \end{aligned}$$

where $\pi_0(t) = 1$. By comparing the coefficients of two sides, we obtain that

$$\begin{cases} \frac{\psi_0(t)}{\psi_0(t)}\pi_0(t) = 1 \\ \frac{\psi_0(t)}{\psi_0(t)}\pi_1(t) + \frac{\psi_1(t)}{\psi_0(t-1)}\pi_0(t-1) = 0 \\ \frac{\psi_0(t)}{\psi_0(t)}\pi_2(t) + \frac{\psi_1(t)}{\psi_0(t-1)}\pi_1(t-1) + \frac{\psi_2(t)}{\psi_0(t-2)}\pi_0(t-2) = 0 \\ \frac{\psi_0(t)}{\psi_0(t)}\pi_3(t) + \frac{\psi_1(t)}{\psi_0(t-1)}\pi_2(t-1) + \frac{\psi_2(t)}{\psi_0(t-2)}\pi_1(t-2) + \frac{\psi_3(t)}{\psi_0(t-3)}\pi_0(t-3) = 0 \\ \cdots \end{cases} . \tag{3.3.3}$$

From the relation above, one can easily conclude that $\{\pi_i(t)\}_{i=1,2,\cdots}$ and $\theta_0(t)$ are periodic functions of t with period τ as well.

To estimate the coefficients ψ, first we recall the innovation representation of $Y(t)$. Suppose

$$Y(t) = \sum_{j=0}^{t-1} b_j(t)\left(Y(t-j) - \mathbb{E}[Y(t-j)|\mathcal{H}_1^{t-j-1}]\right), \tag{3.3.4}$$

where $b_0(t) = 1$, $\mathbb{E}[Y(1)|\mathcal{H}_1^0] = 0$ and $\mathcal{H}_1^{t-1} = \sigma(Y(s) : 0 < s < t)$, the filtration based on the past history. The coefficients $b_j(t), 0 < j < t, t = 1, 2, \cdots$ can be calculated recursively from the innovation algorithm stated in [12] (Proposition 5.2.2).

Proposition 3.3.2. *If $\{Y(t)\}$ has zero mean and $\mathbb{E}[Y(i)Y(j)] = \kappa(i, j)$, where the matrix $[\kappa(i, j)]_{i,j=1}^n$ is non-singular for each $n = 1, 2, \cdots$, then the one-step predictors $\hat{Y}(n +$*

1), $n \geq 0$, *and their mean squared errors* v_{n+1}, $n \geq 1$, *are given by*

$$\hat{Y}(n+1) = \begin{cases} 0 & n = 0 \\ \sum_{k=1}^{n} b_k(n+1)(Y(n+1-k) - \hat{Y}(n+1-k)) & n \geq 1 \end{cases}, \tag{3.3.5}$$

and

$$\begin{cases} v_1 = \kappa(1,1) \\ b_n(n+1) = v_1^{-1}\kappa(n+1,1) \\ b_{n-k}(n+1) = v_{k+1}^{-1}\left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} b_{k-j}(k+1)b_{n-j}(n+1)v_{j+1}\right) & k = 1, \cdots, n-1 \\ v_{n+1} = \kappa(n+1,n+1) - \sum_{j=0}^{n-1} b_{n-j}(n+1)^2 v_{j+1} \end{cases} \cdot \tag{3.3.6}$$

If we replace $\kappa(i,j)$ with the sample covariance function $\hat{\gamma}(i,j)$, we can obtain the estimated coefficients $\hat{b}$ by the innovation algorithm. Next we need to show that $\hat{b}$ converges to ψ.

The first step is to figure out the relation between ψ and b. We follow the idea in [11] for stationary case to prove the convergence of coefficients b in the following lemma for the random periodic case.

Lemma 3.3.3. *Assume that for any $t > 1$, the coefficients of (3.3.1) satisfy,*

$$\sqrt{t} \sum_{i>t-1} \sum_{j>t-1} |\pi_i(t)\pi_j(t)| < \infty,$$

then the coefficients of innovation representation (3.3.4) satisfy that, for any $k \in \mathbb{Z}^+$,

$$\sqrt{t}\left(b_k(t) - \frac{\psi_k(t)}{\psi_0(t-k)}\right) \to 0 \text{ as } t \to \infty.$$

Proof. Define the mean-square error

$$v_t := \mathbb{E}\left[Y(t) - \mathbb{E}[Y(t)|\mathcal{H}_1^{t-1}]\right]^2.$$

As $\mathbb{E}[Y(t)|\mathcal{H}_1^{t-1}]$ is the best linear estimation of $Y(t)$ based on the history

$$\{Y(1), Y(2), \cdots, Y(t-1)\},$$

and by the representation (3.3.1) we have

$$Y(t) = -\sum_{j=1}^{\infty} \pi_j(t)Y(t-j) + \psi_0(t)Z(t).$$

Then we know that $-\sum_{j=1}^{\infty} \pi_j(t)Y(t-j)$ is the best linear estimation of $Y(t)$ based on all the history $\{Y(t-1), Y(t-2), Y(t-3), \cdots, Y(-\infty)\}$, hence the mean-squared error between $Y(t)$ and $-\sum_{j=1}^{\infty} \pi_j(t)Y(t-j)$ will be less than v_t, i.e.

$$\begin{aligned} \psi_0(t)^2 &= Var\left(\psi_0(t)Z(t)\right) \\ &= \mathbb{E}\left[Y(t) + \sum_{j=1}^{\infty} \pi_j(t)Y(t-j)\right]^2 \\ &\leq v_t \end{aligned} \tag{3.3.7}$$

On the other hand, the mean-squared error between $Y(t)$ and $-\sum_{j=1}^{t-1} \pi_j(t)Y(t-j)$ will be greater than v_t, hence

$$\begin{aligned} v_t &\leq \mathbb{E}\left[Y(t) + \sum_{j=1}^{t-1} \pi_j(t)Y(t-j)\right]^2 \\ &= \mathbb{E}\left[\psi_0(t)Z(t) - \sum_{j>t-1} \pi_j(t)Y(t-j)\right]^2 \\ &= \mathbb{E}[\psi_0(t)Z(t)]^2 + \mathbb{E}\left[\sum_{j>t-1} \pi_j(t)Y(t-j)\right]^2. \end{aligned} \tag{3.3.8}$$

And

$$\begin{aligned} &\mathbb{E}\left[\sum_{j>t-1} \pi_j(t)Y(t-j)\right]^2 \\ &= \mathbb{E}[\sum_{i>t-1}\sum_{j>t-1} \pi_i(t)\pi_j(t)Y(t-i)Y(t-j)] \\ &= \sum_{i>t-1}\sum_{j>t-1} \pi_i(t)\pi_j(t)\gamma(t-i, t-j). \end{aligned}$$

For each k, multiplying on both sides of (3.3.2) by $Z(t-k)$ and taking expectations, we have

$$\psi_k(t) = \mathbb{E}[Y(t)Z(t-k)].$$

Also, multiplying by $\left(Y(t-k) - \mathbb{E}[Y(t-k)|\mathcal{H}_1^{t-k-1}]\right)$ on both sides of (3.3.4) and taking expectations, we have

$$b_k(t)v_{t-k} = \mathbb{E}\left[Y(t)\left[Y(t-k) - \mathbb{E}[Y(t-k)|\mathcal{H}_1^{t-k-1}]\right]\right].$$

Hence

$$
\begin{aligned}
&\left|b_k(t) - \frac{\psi_k(t)}{\psi_0(t-k)}\right|^2 \\
= &\left|\frac{\mathbb{E}\left[Y(t)\left[Y(t-k) - \mathbb{E}[Y(t-k)|\mathcal{H}_1^{t-k-1}]\right]\right]}{v_{t-k}} - \frac{\mathbb{E}[Y(t)Z(t-k)]}{\psi_0(t-k)}\right|^2 \\
\leq &\ \mathbb{E}[Y(t)]^2 \mathbb{E}\left[\frac{Y(t-k) - \mathbb{E}[Y(t-k)|\mathcal{H}_1^{t-k-1}]}{v_{t-k}} - \frac{Z(t-k)}{\psi_0(t-k)}\right]^2 \\
= &\ \gamma(t,t)\left(\frac{\mathbb{E}\left[Y(t-k) - \mathbb{E}[Y(t-k)|\mathcal{H}_1^{t-k-1}]\right]^2}{v_{t-k}^2}\right. \\
&\left. - \frac{2\mathbb{E}\left[Z(t-k)\left[Y(t-k) - \mathbb{E}[Y(t-k)|\mathcal{H}_1^{t-k-1}]\right]\right]}{\psi_0(t-k)v(t-k)} + \frac{\mathbb{E}[Z(t-k)]^2}{\psi_0(t-k)^2}\right) \\
= &\ \gamma(t,t)\left(\frac{v_{t-k}}{v_{t-k}^2} - \frac{2\psi_0(t-k)}{\psi_0(t-k)v_{t-k}} + \frac{1}{\psi_0(t-k)^2}\right) \\
= &\ \gamma(t,t)\left(\psi_0(t-k)^{-2} - (v_{t-k})^{-1}\right) \\
= &\ \gamma(t,t)\frac{v_{t-k} - \psi_0(t-k)^2}{\psi_0(t-k)^2 v_{t-k}},
\end{aligned}
$$

where the first inequality is obtained by the Cauchy-Schwartz inequality. By equations (3.3.7) and (3.3.8), we obtain

$$
\left|b_k(t) - \frac{\psi_k(t)}{\psi_0(t-k)}\right|^2 \leq \gamma(t,t)\psi_0(t-k)^{-4}\mathbb{E}\left[\sum_{j>t-k-1} \pi_j(t)Y(t-j)\right]^2.
$$

By the assumption that for any t,

$$
\sqrt{t}\sum_{i>t-1}\sum_{j>t-1} |\pi_i(t)\pi_j(t)| < \infty,
$$

one obtain the convergence of

$$
t|b_k(t) - \frac{\psi_k(t)}{\psi_0(t-k)}|^2 \to 0 \text{ as } t \to \infty,
$$

i.e.

$$
\sqrt{t}\left(b_k(t) - \frac{\psi_k(t)}{\psi_0(t-k)}\right) = o(1).
$$

□

In the next proposition, we prove the asymptotic behaviour of the estimated coefficients $\hat{b}$. This idea is used in [11] and [4].

Proposition 3.3.4. *Let $\{Y(t)\}$ be the process defined by (3.3.2). Assume $\sqrt{m+1}\sum_{i>m}\sum_{j>m}|\pi_i(m+1)\pi_j(m+1)| < \infty$. Then for any $1 < d < m+2-\tau$,*

$$\sqrt{m+1}\left(\hat{\boldsymbol{B}}_{m+1} - \boldsymbol{\psi}_{m+1}\right) \to N(\boldsymbol{0}, \boldsymbol{V_1})$$

in distribution as $m \to \infty$, where

$$\hat{\boldsymbol{B}}_{m+1} := (\hat{\boldsymbol{b}}_{m+1}^T, \hat{\boldsymbol{b}}_m^T, \cdots, \hat{\boldsymbol{b}}_{m+2-\tau}^T)^T, \quad \hat{\boldsymbol{b}}_{m+1} = (\hat{b}_1(m+1), \cdots, \hat{b}_d(m+1))^T,$$

$$\boldsymbol{\psi}_{m+1} := \left(\frac{\psi_1(m+1)}{\psi_0(m)}, \cdots, \frac{\psi_d(m+1)}{\psi_0(m+1-d)}, \cdots, \frac{\psi_1(m+2-\tau)}{\psi_0(m+1-\tau)}, \cdots, \frac{\psi_d(m+2-\tau)}{\psi_0(m+2-\tau-d)}\right)^T$$

and

$$\boldsymbol{V_1} := (\boldsymbol{I} - C)^{-1} diag\left(\Psi_\tau, \Psi_{\tau-1}, \cdots, \Psi_1\right) V diag\left(\Psi_\tau, \Psi_{\tau-1}, \cdots, \Psi_1\right)^T \left((\boldsymbol{I} - C)^{-1}\right)^T.$$

Proof. As $\mathbb{E}[Y(m+1)|\mathcal{H}_{m+1-d}^m]$ has two representations,

$$\mathbb{E}[Y(m+1)|\mathcal{H}_{m+1-d}^m] = \sum_{j=1}^{d} a_j(m+1)Y(m+1-j),$$

and

$$\begin{aligned}
&\mathbb{E}[Y(m+1)|\mathcal{H}_{m+1-d}^m] \\
= &\sum_{j=1}^{d} b_j(m+1)(Y(m+1-j) - \hat{Y}(m+1-j)) \\
= &\sum_{j=1}^{d} b_j(m+1)\left(Y(m+1-j) - \sum_{i=1}^{min(m-j,d)} a_i(m+1-j)Y(m+1-j-i)\right) \\
= &\sum_{j=1}^{d} a_j(m+1-j)Y(m+1-j)
\end{aligned}$$

for any integer $1 < d < m+2-\tau$, the coefficients a and b has the relation as

$$\boldsymbol{a}_{m+1} = Q_m \boldsymbol{b}_{m+1}, \tag{3.3.9}$$

where

$$\boldsymbol{a}_{m+1} = (a_1(m+1), a_2(m+1), \cdots, a_d(m+1))^T,$$

$$\boldsymbol{b}_{m+1} = (b_1(m+1), b_2(m+1), \cdots, b_d(m+1))^T,$$

and

$$Q_m := \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ -a_1(m) & 1 & \cdots & 0 & 0 \\ -a_2(m) & -a_1(m-1) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ -a_{d-2}(m) & -a_{d-3}(m-1) & \cdots & 1 & 0 \\ -a_{d-1}(m) & -a_{d-2}(m-1) & \cdots & -a_1(m-d+2) & 1 \end{bmatrix}.$$

Moreover,

$$\begin{aligned} & \mathbb{E}[Y(m+1)|\mathcal{H}_{m+1-d}^m] \\ = & \sum_{j=1}^{d} a_j(m+1-j)Y(m+1-j) \\ = & \sum_{j=1}^{d} a_j(m+1-j)\left(\sum_{i=1}^{d} b_i(m+1-j)(Y(m+1-j-i)-\hat{Y}(m+1-j-i))\right) \\ = & \sum_{j=1}^{d} b_j(m+1-j)(Y(m+1-j)-\hat{Y}(m+1-j)), \end{aligned}$$

the coefficients a and b also satisfy

$$\boldsymbol{b}_{m+1} = R_m \boldsymbol{a}_{m+1}, \tag{3.3.10}$$

where

$$R_m := \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ b_1(m) & 1 & \cdots & 0 & 0 \\ b_2(m) & b_1(m-1) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{d-2}(m) & b_{d-3}(m-1) & \cdots & 1 & 0 \\ b_{d-1}(m) & b_{d-2}(m-1) & \cdots & b_1(m-d+2) & 1 \end{bmatrix}.$$

By the construction of the estimators $\hat{a}$ and $\hat{b}$, there are also $\hat{\boldsymbol{a}}_{m+1} = \hat{Q}_m \hat{\boldsymbol{b}}_{m+1}$ and $\hat{\boldsymbol{b}}_{m+1} = \hat{R}_m \hat{\boldsymbol{a}}_{m+1}$.

Besides, set

$$\Pi_m = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ \pi_1(m) & 1 & \cdots & 0 & 0 \\ \pi_2(m) & \pi_1(m-1) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \pi_{d-2}(m) & \pi_{d-3}(m-1) & \cdots & 1 & 0 \\ \pi_{d-1}(m) & \pi_{d-2}(m-1) & \cdots & \pi_1(m-d+2) & 1 \end{bmatrix}.$$

Define

$$\Psi_m := \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ \frac{\psi_1(m)}{\psi_0(m-1)} & 1 & \cdots & 0 & 0 \\ \frac{\psi_2(m)}{\psi_0(m-2)} & \frac{\psi_1(m-1)}{\psi_0(m-2)} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{\psi_{d-2}(m)}{\psi_0(m-d+2)} & \frac{\psi_{d-3}(m-1)}{\psi_0(m-d+2)} & \cdots & 1 & 0 \\ \frac{\psi_{d-1}(m)}{\psi_0(m-d+1)} & \frac{\psi_{d-2}(m-1)}{\psi_0(m-d+1)} & \cdots & \frac{\psi_1(m-d+2)}{\psi_0(m-d+1)} & 1 \end{bmatrix},$$

then by relation (3.3.3), $Q_m\Psi_m = I_m$, where I_m is the $d \times d$ identity matrix.

Next we consider parameters in one period simultaneously. Set $(d \times \tau)$ vectors as

$$\boldsymbol{B}_{m+1} := (\boldsymbol{b}_{m+1}^T, \boldsymbol{b}_m^T, \cdots, \boldsymbol{b}_{m+2-\tau}^T)^T,$$

and

$$\boldsymbol{A}_{m+1} := (\boldsymbol{a}_{m+1}^T, \boldsymbol{a}_m^T, \cdots, \boldsymbol{a}_{m+2-\tau}^T)^T.$$

Then we have

$$\begin{aligned} &\sqrt{m+1}\left(\hat{\boldsymbol{B}}_{m+1} - \boldsymbol{B}_{m+1}\right) && (3.3.11) \\ = \ &\sqrt{m+1}diag\left(\hat{R}_{m+1}, \hat{R}_m, \cdots, \hat{R}_{m+2-\tau}\right)\hat{\boldsymbol{A}}_{m+1} - \sqrt{m+1}diag\left(R_{m+1}, R_m, \cdots, R_{m+2-\tau}\right)\boldsymbol{A}_{m+1} \\ = \ &diag\left(\hat{R}_{m+1}, \hat{R}_m, \cdots, \hat{R}_{m+2-\tau}\right)\sqrt{m+1}\left(\hat{\boldsymbol{A}}_{m+1} - \boldsymbol{A}_{m+1}\right) \\ &+ \sqrt{m+1}\left(diag\left(\hat{R}_{m+1}, \hat{R}_m, \cdots, \hat{R}_{m+2-\tau}\right) - diag\left(R_{m+1}, R_m, \cdots, R_{m+2-\tau}\right)\right)\boldsymbol{A}_{m+1}. \\ & && (3.3.12) \end{aligned}$$

Without loss of generality, we assume $m+1$ is multiple times of period. From Section 3.1 we have

$$\sqrt{m+1}\left(\hat{\boldsymbol{A}}_{m+1} - \boldsymbol{A}_{m+1}\right) \to N(\boldsymbol{0}, \boldsymbol{V}),$$

in distribution as $m \to \infty$, where $\boldsymbol{V} := diag\left(\theta_0^2(\tau)\Gamma_d^{-1}(\tau), \cdots, \theta_0^2(1)\Gamma_d^{-1}(1)\right)$. Next we will show that

$$\sqrt{m+1}\left[\left(diag\left(\hat{R}_{m+1}, \hat{R}_m, \cdots, \hat{R}_{m+2-\tau}\right) - diag\left(R_{m+1}, R_m, \cdots, R_{m+2-\tau}\right)\right)\boldsymbol{A}_{m+1}\right.$$
$$\left. - C_{m+1}\left(\hat{\boldsymbol{B}}_{m+1} - \boldsymbol{B}_{m+1}\right)\right] \to 0,$$

in probability as $m \to \infty$, where $C_{m+1} = \sum_{i=1}^{d-1} D_{i,m+1} P^{i(d-1)}$ with

$$D_{i,m+1} = diag\Bigg(\overbrace{0, \cdots, 0}^{i}, \overbrace{a_i(m+1), \cdots, a_i(m+1)}^{d-i}, \overbrace{0, \cdots, 0}^{i}, \overbrace{a_i(m), \cdots, a_i(m)}^{d-i},$$
$$\cdots, \underbrace{0, \cdots, 0}_{i}, \underbrace{a_i(m+2-\tau), \cdots, a_i(m+2-\tau))}_{d-i}\Bigg),$$

and P to be the orthogonal $d\tau \times d\tau$ cyclic permutation matrix

$$P := \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

To view this structure more straightly, we consider an example: for $\tau = 2$ and $d = 3$,

$$C_{m+1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & a_1(m+1) & 0 & 0 \\ a_2(m+1) & 0 & 0 & a_1(m+1) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ a_1(m) & 0 & 0 & 0 & 0 & 0 \\ 0 & a_1(m) & 0 & a_2(m) & 0 & 0 \end{bmatrix}.$$

Then

$$\left(diag\left(\hat{R}_{m+1}, \hat{R}_m\right) - diag\left(R_{m+1}, R_m\right)\right)\boldsymbol{A}_{m+1}$$

$$
= \begin{bmatrix} 0 \\ \left(\hat{b}_1(m) - b_1(m)\right) a_1(m+1) \\ \left(\hat{b}_2(m) - b_2(m)\right) a_1(m+1) + \left(\hat{b}_1(m-1) - b_1(m-1)\right) a_2(m+1) \\ 0 \\ \left(\hat{b}_1(m-1) - b_1(m-1)\right) a_1(m) \\ \left(\hat{b}_2(m-1) - b_2(m-1)\right) a_1(m) + \left(\hat{b}_1(m-2) - b_1(m-2)\right) a_2(m) \end{bmatrix}
$$

and

$$
\begin{aligned}
& C_{m+1}\left(\hat{\boldsymbol{B}}_{m+1} - \boldsymbol{B}_{m+1}\right) \\
= & \begin{bmatrix} 0 \\ \left(\hat{b}_1(m) - b_1(m)\right) a_1(m+1) \\ \left(\hat{b}_2(m) - b_2(m)\right) a_1(m+1) + \left(\hat{b}_1(m+1) - b_1(m+1)\right) a_2(m+1) \\ 0 \\ \left(\hat{b}_1(m+1) - b_1(m+1)\right) a_1(m) \\ \left(\hat{b}_2(m+1) - b_2(m+1)\right) a_1(m) + \left(\hat{b}_1(m) - b_1(m)\right) a_2(m) \end{bmatrix}.
\end{aligned}
$$

That is to say, it suffices to show that

$$
\sqrt{t}\left(b_i(t) - b_i(t-\tau)\right) \to 0 \tag{3.3.13}
$$

and

$$
\sqrt{t}\left(\hat{b}_i(t) - \hat{b}_i(t-\tau)\right) \to 0 \tag{3.3.14}
$$

in probability as $t \to \infty$.

Lemma 3.3.3 and the periodicity of $\frac{\psi_i(t-\tau)}{\psi_0(t-\tau-i)} = \frac{\psi_i(t)}{\psi_0(t-i)}$ lead to (3.3.13). For (3.3.14), it suffices to prove $\sqrt{t}\left(\hat{a}_i(t+\tau) - \hat{a}_i(t)\right) \to 0$ in probability by (3.3.9). By (3.1.4),

$$
\begin{aligned}
& \sqrt{t}\left(\hat{\boldsymbol{a}}_d(t) - \hat{\boldsymbol{a}}_d(t-\tau)\right) \\
= \ & \sqrt{t}(\boldsymbol{X}_{t,n}^T \boldsymbol{X}_{t,n})^{-1} \boldsymbol{X}_{t,n}^T \boldsymbol{Y}_{t,n} - \sqrt{t}(\boldsymbol{X}_{t-\tau,n-1}^T \boldsymbol{X}_{t-\tau,n-1})^{-1} \boldsymbol{X}_{t-\tau,n-1}^T \boldsymbol{Y}_{t,n} \\
= \ & \sqrt{t}\left[(\boldsymbol{X}_{t,n}^T \boldsymbol{X}_{t,n})^{-1} - (\boldsymbol{X}_{t-\tau,n-1}^T \boldsymbol{X}_{t-\tau,n-1})^{-1}\right] \boldsymbol{X}_{t,n}^T \boldsymbol{Y}_{t,n} \\
& + (\boldsymbol{X}_{t-\tau,n-1}^T \boldsymbol{X}_{t-\tau,n-1})^{-1} \sqrt{t}\left(\boldsymbol{X}_{t,n}^T \boldsymbol{Y}_{t,n} - \boldsymbol{X}_{t-\tau,n-1}^T \boldsymbol{Y}_{t,n}\right),
\end{aligned}
$$

where $n = \lceil \frac{t}{\tau} \rceil$ with $\lceil \cdot \rceil$ to be the greatest integer function. The i^{th} component of $\sqrt{t}\left(\boldsymbol{X}_{t,n}^T \boldsymbol{Y}_{t,n} - \boldsymbol{X}_{t-\tau,n-1}^T \boldsymbol{Y}_{t,n}\right)$ is

$$\begin{aligned}
&\frac{\sqrt{t}}{n}\sum_{k=0}^{n-1} Y(t-i-k\tau)Y(t-k\tau) - \frac{\sqrt{t}}{n-1}\sum_{k=0}^{n-2} Y(t-\tau-i-k\tau)Y(t-\tau-k\tau)\\
=&\frac{\sqrt{t}(n-1)}{n(n-1)}\sum_{k=0}^{n-1} Y(t-i-k\tau)Y(t-k\tau) - \frac{\sqrt{t}n}{n(n-1)}\sum_{k=1}^{n-1} Y(t-i-k\tau)Y(t-k\tau)\\
=&\frac{\sqrt{t}}{n}Y(t-i)Y(t) - \frac{\sqrt{t}}{n(n-1)}\sum_{k=1}^{n-1} Y(t-i-k\tau)Y(t-k\tau)\\
\to& 0
\end{aligned}$$

in probability as $t \to \infty$ by Proposition 3.1.4. Besides,

$$\begin{aligned}
&\sqrt{t}\left|(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1} - (\boldsymbol{X}_{t-\tau,n-1}^T\boldsymbol{X}_{t-\tau,n-1})^{-1}\right|\\
=&\sqrt{t}\left|(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1}\left(\boldsymbol{X}_{t-\tau,n-1}^T\boldsymbol{X}_{t-\tau,n-1} - \boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n}\right)(\boldsymbol{X}_{t-\tau,n-1}^T\boldsymbol{X}_{t-\tau,n-1})^{-1}\right|\\
\le&\left|(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1}\right|\sqrt{t}\left|\boldsymbol{X}_{t-\tau,n-1}^T\boldsymbol{X}_{t-\tau,n-1} - \boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n}\right|\left|(\boldsymbol{X}_{t-\tau,n-1}^T\boldsymbol{X}_{t-\tau,n-1})^{-1}\right|.
\end{aligned}$$

By similar calculation, we can prove that each component of $\sqrt{t}\left(\boldsymbol{X}_{t-\tau,n-1}^T\boldsymbol{X}_{t-\tau,n-1} - \boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n}\right)$ converges to zero in probability as $t \to \infty$, hence

$$\sqrt{t}\left|(\boldsymbol{X}_{t,n}^T\boldsymbol{X}_{t,n})^{-1} - (\boldsymbol{X}_{t-\tau,n-1}^T\boldsymbol{X}_{t-\tau,n-1})^{-1}\right| \to 0$$

in probability. Therefore,

$$\sqrt{t}\left(\hat{\boldsymbol{a}}_d(t) - \hat{\boldsymbol{a}}_d(t-\tau)\right) \to \boldsymbol{0}$$

in probability as $t \to \infty$.

Combing above results, (3.3.11) is equivalent to

$$\begin{aligned}
&\sqrt{m+1}(\boldsymbol{I} - C_{m+1})(\hat{\boldsymbol{B}}_{m+1} - \boldsymbol{B}_{m+1})\\
=&\sqrt{m+1}diag\left(\hat{R}_{m+1}, \hat{R}_m, \cdots, \hat{R}_{m+2-\tau}\right)\left(\hat{\boldsymbol{A}}_{m+1} - \boldsymbol{A}_{m+1}\right) + 0_p(1).
\end{aligned}$$

By Remark 3.1.8, we have

$$(a_1(m+1), \cdots, a_d(m+1))^T \to (-\pi_1(m+1), \cdots, -\pi_d(m+1))^T$$

in probability as $m \to \infty$. Then $C_{m+1} \to C := \sum_{i=1}^{d-1}\tilde{\Pi}_{i,\tau}P^{i(d-1)}$ in probability as $m \to \infty$, where

$$\tilde{\Pi}_{i,m+1} = diag\left(\overbrace{0,\cdots,0}^{i}, \overbrace{-\pi_i(\tau),\cdots,-\pi_i(\tau)}^{d-i}, \overbrace{0,\cdots,0}^{i}, \overbrace{-\pi_i(\tau-1),\cdots,-\pi_i(\tau-1)}^{d-i},\right.$$

$$\cdots, \underbrace{0, \cdots, 0}_{i}, \underbrace{-\pi_i(1), \cdots, -\pi_i(1))}_{d-i}\Bigg) .$$

And by the relation between coefficients a and b in (3.3.9),

$$diag\left(\hat{R}_{m+1}, \hat{R}_m, \cdots, \hat{R}_{m+2-\tau}\right) \to diag\left(\Psi_\tau, \Psi_{\tau-1}, \cdots, \Psi_1\right)$$

in probability as $m \to \infty$. Consequently, combining with Lemma 3.3.3, we have

$$\sqrt{m+1}\left(\hat{\boldsymbol{B}}_{m+1} - \boldsymbol{\psi}_{m+1}\right) \to N(\mathbf{0}, \boldsymbol{V}_{\mathbf{1}})$$

in distribution as $m \to \infty$, where

$$\boldsymbol{\psi}_{m+1} := \left(\frac{\psi_1(m+1)}{\psi_0(m)}, \cdots, \frac{\psi_d(m+1)}{\psi_0(m+1-d)}, \cdots, \frac{\psi_1(m+2-\tau)}{\psi_0(m+1-\tau)}, \cdots, \frac{\psi_d(m+2-\tau)}{\psi_0(m+2-\tau-d)}\right)^T$$

and

$$\boldsymbol{V}_{\mathbf{1}} := (\boldsymbol{I} - C)^{-1} diag\left(\Psi_\tau, \Psi_{\tau-1}, \cdots, \Psi_1\right) V diag\left(\Psi_\tau, \Psi_{\tau-1}, \cdots, \Psi_1\right)^T \left((\boldsymbol{I} - C)^{-1}\right)^T .$$

□

3.4 Truncated Innovation Algorithm

From the algorithm shown in the previous section, the coefficients b only depend on the covariance matrix of Y. But as in the real life we usually cannot choose w in the sample autocovariance function to be large enough, so in the next lemma we will see that the sample covariance matrix of random periodic process has rank w, which is the number of previous cycles we use to estimate the sample autocovariance function. If the size of the sample covariance matrix is greater than w, the matrix will be singular, which will cause the estimated v_t drop to zero. Hence we will use the truncated innovation algorithm to estimate the coefficients $\hat{b}$.

Lemma 3.4.1. *The rank of the sample covariance matrix of random periodic process is no greater than w.*

Proof. For $N > w$, let

$$\hat{\Gamma}_N := \begin{bmatrix} \hat{\gamma}(1,1) & \hat{\gamma}(1,2) & \cdots & \hat{\gamma}(1,N) \\ \hat{\gamma}(2,1) & \hat{\gamma}(2,2) & \cdots & \hat{\gamma}(2,N) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(N,1) & \hat{\gamma}(N,2) & \cdots & \hat{\gamma}(N,N) \end{bmatrix},$$

where $\hat{\gamma}(i,j) = \frac{1}{w}\sum_{k=0}^{w-1} y(i-k\tau)y(j-k\tau)$.

Set

$$P_N := \begin{bmatrix} y(1) & y(1-\tau) & \cdots & y(1-(w-1)\tau) \\ y(2) & y(2-\tau) & \cdots & y(2-(w-1)\tau) \\ \vdots & \vdots & \ddots & \vdots \\ y(N) & y(N-\tau) & \cdots & y(N-(w-1)\tau) \end{bmatrix}.$$

Then $\hat{\Gamma}_N = \frac{1}{w}P_N P_N^T$. As $Rank(P_N) \le w$,

$$Rank(\hat{\Gamma}_N) \le Rank(P_N) \le w.$$

□

In order to use the innovation algorithm to calculate the coefficients b, we use the truncated innovation representation of Y, in which we only take several periods of history data to estimate Y instead of using all past information. Then the modified innovation algorithm is shown as follows.

Proposition 3.4.2. *(Truncated Innovation Algorithm) Set $\hat{Y}(n+1) := \mathbb{E}[Y(n+1)|\mathcal{H}_{n+1}]$, for some positive integer K with $K\tau < w$, if*

$$\hat{Y}(n+1) = \begin{cases} 0 & n = 0 \\ \sum_{k=1}^{n} b_k(n+1)(Y(n+1-k) - \hat{Y}(n+1-k)) & 1 \le n \le K\tau \\ \sum_{k=1}^{K\tau} b_k(n+1)(Y(n+1-k) - \hat{Y}(n+1-k)) & n > K\tau \end{cases}, \quad (3.4.1)$$

then when $n > K\tau$*, for* $k = n - K\tau, \cdots, n-1$*,*

$$\begin{cases} b_{n-k}(n+1) = v_{k+1}^{-1}\left(\gamma(n+1,k+1) - \sum_{j=k-K\tau}^{k-1} b_{k-j}(k+1)b_{n-j}(n+1)v_{j+1}\right) & k > K\tau \\ b_{n-k}(n+1) = v_{k+1}^{-1}\left(\gamma(n+1,k+1) - \sum_{j=0}^{k-1} b_{k-j}(k+1)b_{n-j}(n+1)v_{j+1}\right) & k \leq K\tau \\ v_{n+1} = \kappa(n+1,n+1) - \sum_{k=n-K\tau}^{n-1} b_{n-k}(n+1)^2 v_{k+1} \end{cases}, \tag{3.4.2}$$

Proof. When $n \leq K\tau$, we use the original algorithm.

When $n > K\tau$, for each k satisfying $n - K\tau \leq k \leq n-1$, multiplying $\hat{Y}(n+1)$ with $Y(k+1) - \hat{Y}(k+1)$ and taking expectations, and by (3.2.3), we have

$$\begin{aligned} & \mathbb{E}\left[Y(n+1)\left(Y(k+1) - \hat{Y}(k+1)\right)\right] \\ = \; & \mathbb{E}\left[\hat{Y}(n+1)\left(Y(k+1) - \hat{Y}(k+1)\right)\right] \\ = \; & b_{n-k}(n+1)v_{k+1}. \end{aligned}$$

Hence $b_{n-k}(n+1)$, $k = n - K\tau, \ldots, k \leq n-1$ are given by

$$b_{n-k}(n+1) = v_{k+1}^{-1}\mathbb{E}\left[Y(n+1)\left(Y(k+1) - \hat{Y}(k+1)\right)\right].$$

When $k > K\tau$,

$$\begin{aligned} & b_{n-k}(n+1) \\ = \; & v_{k+1}^{-1}\left(\mathbb{E}[Y(n+1)Y(k+1)] - \mathbb{E}[Y(n+1)\hat{Y}(k+1)]\right) \\ = \; & v_{k+1}^{-1}\left\{\kappa(n+1,k+1) - \mathbb{E}\left[Y(n+1)\left(\sum_{j=1}^{K\tau} b_j(k+1)(Y(k+1-j) - \hat{Y}(k+1-j))\right)\right]\right\} \\ = \; & v_{k+1}^{-1}\left\{\kappa(n+1,k+1) - \sum_{j=1}^{K\tau} b_j(k+1)\mathbb{E}\left[Y(n+1)\left(Y(k+1-j) - \hat{Y}(k+1-j)\right)\right]\right\} \\ = \; & v_{k+1}^{-1}\left\{\kappa(n+1,k+1) - \sum_{j=k-K\tau}^{k-1} b_{k-j}(k+1)\mathbb{E}\left[Y(n+1)\left(Y(j+1) - \hat{Y}(j+1)\right)\right]\right\} \\ = \; & v_(k+1)^{-1}\left(\kappa(n+1,k+1) - \sum_{j=k-K\tau}^{k-1} b_{k-j}(k+1)b_{n-j}(n+1)v_{j+1}\right). \end{aligned}$$

When $k \leq K\tau$,

$$b_{n-k}(n+1) = v_{k+1}^{-1}\left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} b_{k-j}(k+1)b_{n-j}(n+1)v_{j+1}\right).$$

By (3.2.3), $\mathbb{E}[\hat{Y}(t)]^2 = \mathbb{B}[Y(t)\hat{Y}(t)]$, hence

$$\begin{aligned}
& \mathbb{E}[Y(n+1)]^2 \\
= \ & \mathbb{E}[\hat{Y}(n+1) + Y(n+1) - \hat{Y}(n+1)]^2 \\
= \ & \mathbb{E}[\hat{Y}(n+1)]^2 + 2\mathbb{E}\left[\hat{Y}(n+1)\left(Y(n+1) - \hat{Y}(n+1)\right)\right] + \mathbb{E}[Y(n+1) - \hat{Y}(n+1)]^2 \\
= \ & \mathbb{E}[\hat{Y}(n+1)]^2 + \mathbb{E}[Y(n+1) - \hat{Y}(n+1)]^2.
\end{aligned}$$

Besides,

$$\begin{aligned}
\mathbb{E}[\hat{Y}(n+1)]^2 &= \mathbb{E}\left[\sum_{j=1}^{K\tau} b_j(n+1)(Y(n+1-j) - \hat{Y}(n+1-j))\right]^2 \\
&= \mathbb{E}\left[\sum_{k=n-K\tau}^{n-1} b_{n-k}(n+1)(Y(k+1) - \hat{Y}(k+1))\right]^2 \\
&= \sum_{k=n-K\tau}^{n-1} b_{n-k}(n+1)^2 v_{k+1}.
\end{aligned}$$

Thus,

$$\begin{aligned}
v_{n+1} &= \mathbb{E}[Y(n+1) - \hat{Y}(n+1)]^2 \\
&= \mathbb{E}[Y(n+1)]^2 - \mathbb{E}[\hat{Y}(n+1)]^2 \\
&= \kappa(n+1,n+1) - \sum_{k=n-K\tau}^{n-1} b_{n-k}(n+1)^2 v_{k+1}.
\end{aligned}$$

□

We will see in examples that the K frequently takes the value 1.

Chapter 4

Non-causal Case

4.1 Background

Recall that a causal autoregressive model of data means that the data can be fully expressed as a series of history noise. It requires the characteristic function of autoregressive parameters $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$ has no root in the unit circle. On the other hand, if $\phi(z)$'s roots are all in the unit circle, the model is called purely non-causal autoregressive model, then the data fully depends on future noise. If $\phi(z)$ has roots both in and out of the unit circle, then the data depends not only on history noise, but also the future's. This is called mixed non-causal autoregressive model.

Brockwell and Davis stated in [12] that non-casual autoregressive model can be re-expressed as a causal or purely non-causal autoregressive model driven by a new noise sequence $\{\tilde{Z}(t)\}$. For non-Gaussian case $\{\tilde{Z}(t)\}$ is uncorrelated, but not independent with each other, which will make estimation much harder. However, the autocovariance of data remains unchanged. Hence the Yule-Walker and innovation algorithms are unable to distinguish among causal and non-causal cases.

Breidt *et al.* studied the mixed non-causal case in [10]. Suppose in the following model

$$\phi(B)X_t = Z_t, \tag{4.1.1}$$

the autoregressive polynomial $\phi(z)$ has roots both in and out of the unit circle. $\phi(z)$ can be factorized as $\phi(z) = \phi^+(z)\phi^*(z)$, where ϕ^+ only has roots out of the unit circle and $\phi^*(z)$ only has roots in the unit circle. That is to say, if we define $U_t = \phi^*(B)X_t$ and

$V_t = \phi^+(B)X_t$, then U_t is a causal autoregressive process and V_t is a purely non-causal one. Next they approximated the likelihood function of the parameters and concluded the asymptotic normality of the maximum likelihood estimation.

Lanne and Saikkonen ([23]) re-expressed the mixed non-causal autoregressive as

$$\varphi(B^{-1})\phi(B)X_t = Z_t, \tag{4.1.2}$$

where $\varphi(z^{-1})$ represents the purely non-causal polynomial and $\phi(z)$ represents the causal one. They stated that if the last parameter of $\varphi(z^{-1})$, i.e. φ_s, is not zero, then there is a one-to-one correspondence between (4.1.2) and (4.1.1). Besides, they stated that (4.1.2) containes the overfitting condition of ϕ and φ. The asymptotic normality of the maximum likelihood estimation of parameters in (4.1.2) were deduced by similar idea as [10].

In the following sections, we construct a very simple version of the mixed non-causal model for random periodic processes based on (4.1.2). For this we assume that the coefficients are all constants and the number of coefficients is independent with time t. In the future we attempt to release this assumption to general case. The asymptotic normality of the maximum likelihood estimation of parameters will be deduced according to the idea in [10], but under the central limit theorem and ergodic theorem for random periodic processes.

4.2 Mixed Non-causal Autoregressive Model for Random Periodic Processes

Let $\{Y(t)\}_{t\in\mathbb{Z}}$ be a random periodic process satisfying

$$\varphi(B^{-1})\phi(B)Y(t) = Z(t), \tag{4.2.1}$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_r B^r$, $\varphi(B^{-1}) = 1 - \varphi_1 B^{-1} - \cdots - \varphi_s B^{-s}$ and $Z(t)$ is a sequence of i.i.d. random variables with mean 0 and variance σ_t^2. Assume that σ_t is deterministic periodic function of time t. Also assume that the polynomials $\phi(z)$, $\varphi(z)$ have their zeros outside the unit circle, so that $\phi(z) \neq 0$ for $|z| \leq 1$ and $\varphi(z) \neq 0$ for $|z| \leq 1$ and $\phi_r \neq 0$, $\varphi_s \neq 0$. Moreover, assume that the probability density function of

$Z(t)$ is $f_t(x) := \frac{1}{\sigma_t} f(\frac{x}{\sigma_t})$, where $f(x)$ satisfies the following assumptions ([10]):

$$
\begin{aligned}
A1 :& f(x) > 0 \ for\ all\ x. \\
A2 :& f \in C^2(\mathbb{R}). \\
A3 :& f' \in L^1(\mathbb{R})\ with\ \int f'(x)dx = 0. \\
A4 :& \int xf'(x)dx = -1. \\
A5 :& \int f''(x)dx = 0. \\
A6 :& \int xf''(x)dx = 0. \\
A7 :& \int x^2 f''(x)dx = 2. \\
A8 :& \int (1+x^2)\frac{(f'(x))^2}{f(x)}dx < \infty.
\end{aligned}
$$

Define $U(t) = \varphi(B^{-1})Y(t)$ and $V(t) = \phi(B)Y(t)$. From (4.2.1) and $\varphi(B^{-1})\phi(B) = \phi(B)\varphi(B^{-1})$, we have $\phi(B)U(t) = Z(t)$. Thus $U(t)$ is causal, and so $U(t)$ can be expressed as

$$U(t) = \sum_{i=0}^{\infty} \alpha_i Z(t-i), \tag{4.2.2}$$

with $\alpha_0 = 1$ and α_i decay to zero at exponential rate as $i \to \infty$. From $\varphi(B^{-1})V(t) = Z(t)$, $V(t)$ is purely non-causal, then $V(t)$ can be expresses as

$$V(t) = \sum_{j=0}^{\infty} \beta_j Z(t+j), \tag{4.2.3}$$

with $\beta_0 = 1$ and β_i decay to zero at exponential rate as $i \to \infty$. The process $Y(t)$ itself has the two-sided moving average representation

$$Y(t) = \sum_{j=-\infty}^{\infty} \psi_j Z(t), \tag{4.2.4}$$

where $\psi(z) := \phi(z)^{-1}\varphi(z^{-1})^{-1}$.

Next we are going to express the joint density function of $\{Y(t)\}$. First we have

$$\begin{bmatrix} U(1) \\ \vdots \\ U(n-s) \\ V(n-s+1) \\ \vdots \\ V(n) \end{bmatrix} = \begin{bmatrix} Y(1) - \sum_{j=1}^{s} \varphi_j Y(1+j) \\ \vdots \\ Y(n-s) - \sum_{j=1}^{s} \varphi_j Y(n-s+j) \\ Y(n-s+1) - \sum_{i=1}^{r} \phi_i Y(n-s+1-i) \\ \vdots \\ Y(n) - \sum_{i=1}^{r} \phi_i Y(n-i) \end{bmatrix} = A_n \begin{bmatrix} Y(1) \\ \vdots \\ Y(n-s) \\ Y(n-s+1) \\ \vdots \\ Y(n) \end{bmatrix},$$

where

$$A_n = \begin{bmatrix} 1 & -\varphi_1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\varphi_1 & -\varphi_2 & \cdots & -\varphi_s & 0 \\ 0 & 0 & \cdots & 0 & 1 & -\varphi_1 & \cdots & -\varphi_{s-1} & -\varphi_s \\ 0 & 0 & \cdots & -\phi_2 & -\phi_1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1,n-s-1} & a_{n-1,n-s} & a_{n-1,n-s+1} & \cdots & 1 & 0 \\ 0 & 0 & \cdots & a_{n,n-s-1} & a_{n,n-s} & a_{n,n-s+1} & \cdots & -\phi_1 & 1 \end{bmatrix}.$$

The values of a_{ij} in the last two rows depend on the values of s and j.

Similarly,

$$\begin{bmatrix} U(1) \\ \vdots \\ U(r) \\ Z(r+1) \\ \vdots \\ Z(n-s) \\ V(n-s+1) \\ \vdots \\ V(n) \end{bmatrix} = \begin{bmatrix} U(1) \\ \vdots \\ U(r) \\ U(r+1) - \sum_{i=1}^{r} \phi_i U(r+1-i) \\ \vdots \\ U(n-s) - \sum_{i=1}^{r} \phi_i U(n-s-i) \\ V(n-s+1) \\ \cdots \\ V(n) \end{bmatrix} = C_n \begin{bmatrix} U(1) \\ \vdots \\ U(r) \\ U(r+1) \\ \vdots \\ U(n-s) \\ V(n-s+1) \\ \vdots \\ V(n) \end{bmatrix},$$

where

$$C_n = \begin{bmatrix} 1 & \cdots & 0 & & & & & & & & \\ \vdots & \ddots & \vdots & & & & & & & & \\ 0 & \cdots & 1 & & & & & & & & \\ -\phi_r & \cdots & -\phi_1 & 1 & 0 & \cdots & 0 & 0 & & & \\ 0 & \cdots & -\phi_2 & -\phi_1 & 1 & \cdots & 0 & 0 & & & \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & & & \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 & 0 & & & \\ 0 & \cdots & 0 & 0 & 0 & \cdots & -\phi_1 & 1 & & & \\ & & & & & & & & 1 & \cdots & 0 \\ & & & & & & & & \vdots & \ddots & \vdots \\ & & & & & & & & 0 & \cdots & 1 \end{bmatrix}.$$

We can see that C_n are non-singular and $det(C_n) = 1$. Re-represent A_n as

$$A_n = \begin{bmatrix} A_1 & A_2 \\ A_3 & A_4 \end{bmatrix},$$

where A_1 is a $(n-s)\times(n-s)$ upper triangular matrix with $det(A_1) = 1$, and A_4 is a $s \times s$ lower triangular matrix with $det(A_4) = 1$. As $det(A_n) = det(A_1)det(A_4 - A_3A_1^{-1}A_2) = 1$, A_n is also non-sigular. Hence

$$\begin{bmatrix} Y(1) \\ \vdots \\ Y(r) \\ Y(r+1) \\ \vdots \\ Y(n-s) \\ Y(n-s+1) \\ \vdots \\ Y(n) \end{bmatrix} = (C_nA_n)^{-1} \begin{bmatrix} U(1) \\ \vdots \\ U(r) \\ Z(r+1) \\ \vdots \\ Z(n-s) \\ V(n-s+1) \\ \vdots \\ V(n) \end{bmatrix}. \tag{4.2.5}$$

From (4.2.2) and (4.2.3), we know that $U(t)$ is independent of $V(t+s)$, hence $(U(1), \cdots, U(r))^T$, $(Z(r+1), \cdots, Z(n-s))^T$ and $(V(n-s+1), \cdots, V(n))^T$ are independent. The joint density function of $(U(1), \cdots, U(r), Z(r+1), \cdots, Z(n-s), V(n-s+1), \cdots, V(n))^T$ can be expressed as

$$h_U(U(1),\cdots,U(r))\left(\prod_{t=r+1}^{n-s} f_t(Z(t))\right)h_V(V(n-s+1),\cdots,V(n)),$$

where h_U and h_V signify the joint density functions of U and V respectively. We can see that the non-stochastic matrices A_n and C_n are non-singular and $det(C_n)=1$. Then the joint density function of $(Y(1),\cdots,Y(n))^T$ is expressed as

$$h_U(\varphi(B^{-1})Y(1),\cdots,\varphi(B^{-1})Y(r))\left(\prod_{t=r+1}^{n-s} f_t(\varphi(B^{-1})\phi(B)Y(t))\right)$$
$$\times\, h_V(\phi(B)Y(n-s+1),\cdots,\phi(B)Y(n))det(A_n)det(C_n).$$

As $det(A_n)$ is independent of sample size n, we approximate the log-likelihood function of parameters $\boldsymbol{\theta}_t := (\phi_1,\cdots,\phi_r,\varphi_1,\cdots,\varphi_s,\sigma_t))^T$ as

$$L_n(\boldsymbol{\theta}) := \sum_{t=r+1}^{n-s} g_t(\boldsymbol{\theta}_t), \tag{4.2.6}$$

where

$$\begin{aligned} g_t(\boldsymbol{\theta}_t) &:= logf_t(Z(t)) \\ &= logf_t(U(t)-\phi_1 U(t-1)-\cdots-\phi_r U(t-r)) \\ &= logf_t(V(t)-\varphi_1 V(t+1)-\cdots-\varphi_s V(t+r)). \end{aligned}$$

Evaluating the partial derivatives of g_t at the true values of parameters and using the logogram f_t for $f_t(Z(t))$, we obtain

$$\frac{\partial g_t}{\partial \phi_i} = -U(t-i)\frac{f_t'}{f_t}, \qquad i=1,\cdots,r,$$

$$\frac{\partial g_t}{\partial \varphi_i} = -V(t+j)\frac{f_t'}{f_t}, \qquad j=1,\cdots,s,$$

and

$$\frac{\partial g_t}{\partial \sigma_t} = -\frac{1}{\sigma_t}\left(Z(t)\frac{f_t'(Z(t))}{f_t(Z(t))}+1\right).$$

The assumption $A4$ on $f(x)$ implies

$$\mathbb{E}\left[Z(t)\frac{f_t'}{f_t}\right] = \int x\frac{f_t'(x)}{f_t(x)}f_t(x)dx = \int x\left[\frac{1}{\sigma_t}f\left(\frac{x}{\sigma_t}\right)\right]' dx = \int \frac{x}{\sigma_t^2}f'\left(\frac{x}{\sigma_t}\right)dx = -1.$$

Then

$$\mathbb{E}\left[Z(s)\frac{f_t'}{f_t}\right] = \begin{cases} 0, & s \neq t, \\ -1, & s = t. \end{cases}$$

Hence for $i = 1, \cdots, r$ and $j = 1, \cdots, s$,

$$\mathbb{E}\left[\frac{\partial g_t}{\partial \phi_i}\right] = 0, \qquad \mathbb{E}\left[\frac{\partial g_t}{\partial \varphi_j}\right] = 0, \qquad \mathbb{E}\left[\frac{\partial g_t}{\partial \sigma_t}\right] = 0.$$

Next we determine the limiting covariance matrix of $\left(\sum_{t=r+1}^{n-s} \frac{\partial g_t}{\partial \theta_1(t)}, \cdots, \sum_{t=r+1}^{n-s} \frac{\partial g_t}{\partial \theta_{r+s+1}(t)}\right)^T$. Define $\tilde{I} := \int_{\mathbb{R}} \frac{(f'(x))^2}{f(x)} dx$, $\tilde{J} := \int_{\mathbb{R}} x^2 \frac{(f'(x))^2}{f(x)} dx - 1$, then

$$Cov(Z(t-i)\frac{f_t'}{f_t}, Z(k-j)\frac{f_k'}{f_k}) = \begin{cases} \tilde{J}, & t = k, i = j = 0, \\ \sigma_{t-i}^2 \sigma_t^{-2} \tilde{I}, & t = k, i = j \neq 0, \\ 1, & t \neq k, i = t-k, j = k-t, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\gamma_U(\cdot,\cdot)$ and $\gamma_V(\cdot,\cdot)$ denote the autocovariance functions of $\{U(t)\}$ and $\{V(t)\}$ respectively. Then from the representations of (4.2.2) and (4.2.3), we obtain

$$Cov(U(t-i)\frac{f_t'}{f_t}, U(k-j)\frac{f_k'}{f_k}) = \begin{cases} \gamma_U(t-i, t-j)\sigma_t^{-2}\tilde{I}, & t = k, i, j = 1, \cdots, r, \\ 0, & t \neq k, i, j = 1, \cdots, r. \end{cases}$$

$$Cov(V(t+i)\frac{f_t'}{f_t}, V(k+j)\frac{f_k'}{f_k}) = \begin{cases} \gamma_V(t+i, t+j)\sigma_t^{-2}\tilde{I}, & t = k, i, j = 1, \cdots, s, \\ 0, & t \neq k, i, j = 1, \cdots, s. \end{cases}$$

$$\begin{aligned} & Cov(U(t-i)\frac{f_t'}{f_t}, V(k+j)\frac{f_k'}{f_k}) \\ = \ & Cov(\sum_{a=0}^{\infty} \alpha_a Z(t-i-a)\frac{f_t'}{f_t}, \sum_{b=0}^{\infty} \beta_b Z(k+j+b)\frac{f_k'}{f_k}) \\ = \ & \begin{cases} \alpha_{t-i-k}\beta_{t-k-j}, & t-k \geq m_0, i, j = 1, \cdots, r, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

where $m_0 := max(i, j)$, and for any $i = 1, \cdots, r$ and $j = 1, \cdots, s$,

$$Cov(\frac{\partial g_t}{\partial \sigma_t}, \frac{\partial g_k}{\partial \phi_i}) = 0, \qquad Cov(\frac{\partial g_t}{\partial \sigma_t}, \frac{\partial g_k}{\partial \varphi_j}) = 0, \qquad \text{for any } t, k,$$

and

$$Cov(\frac{\partial g_t}{\partial \sigma_t}, \frac{\partial g_k}{\partial \sigma_k}) = \begin{cases} \sigma_t^{-2}\tilde{J}, & t = k, \\ 0, & t \neq k. \end{cases}$$

Also, for any $i = 1, \cdots, r$ and $j = 1, \cdots, s$,

$$\begin{aligned} & \frac{1}{n-r-s} Cov(\sum_{t=r+1}^{n-s} \frac{\partial g_t}{\partial \phi_i}, \sum_{k=r+1}^{n-s} \frac{\partial g_k}{\partial \varphi_j}) \\ = & \frac{1}{n-r-s} \sum_{t=r+1}^{n-s} \sum_{k=r+1}^{n-s} Cov(U(t-i)\frac{f_t'}{f_t}, V(k+j)\frac{f_k'}{f_k}) \\ = & \frac{1}{n-r-s} \sum_{k=r+1}^{n-m_0-s} \sum_{t=k+m_0}^{n-s} \alpha_{t-i-k}\beta_{t-k-j} \\ = & \frac{1}{n-r-s} \sum_{k=r+1}^{n-m_0-s} \sum_{t=m_0-i}^{n-s-i-k} \alpha_t \beta_{t+i-j} \end{aligned}$$

As $\{\alpha_i\}_{i\in\mathbb{Z}^+}$ and $\{\beta_j\}_{j\in\mathbb{Z}^+}$ decay at exponential rate, hence, as $n \to \infty$,

$$\frac{1}{n-r-s} Cov(\sum_{t=r+1}^{n-s} \frac{\partial g_t}{\partial \phi_i}, \sum_{k=r+1}^{n-s} \frac{\partial g_k}{\partial \varphi_j}) \to \sum_{t=m_0-i}^{\infty} \alpha_t \beta_{t+i-j}.$$

Similarly, for $i, j = 1, \cdots, r$, by the periodicity of σ_t,

$$\begin{aligned} & \frac{1}{n-r-s} Cov(\sum_{t=r+1}^{n-s} \frac{\partial g_t}{\partial \phi_i}, \sum_{k=r+1}^{n-s} \frac{\partial g_k}{\partial \phi_j}) \\ = & \frac{1}{n-r-s} \sum_{t=r+1}^{n-s} \sum_{k=r+1}^{n-s} Cov(U(t-i)\frac{f_t'}{f_t}, U(k-j)\frac{f_k'}{f_k}) \\ \to & \frac{1}{\tau} \sum_{t=1}^{\tau} \gamma_U(t-i, t-j)\sigma_t^{-2}\tilde{I}, \end{aligned}$$

as $n \to \infty$. For $i, j = 1, \cdots, s$,

$$\begin{aligned} & \frac{1}{n-r-s} Cov(\sum_{t=r+1}^{n-s} \frac{\partial g_t}{\partial \varphi_i}, \sum_{k=r+1}^{n-s} \frac{\partial g_k}{\partial \varphi_j}) \\ = & \frac{1}{n-r-s} \sum_{t=r+1}^{n-s} \sum_{k=r+1}^{n-s} Cov(V(t+i)\frac{f_t'}{f_t}, V(k+j)\frac{f_k'}{f_k}) \end{aligned}$$

$$\to \frac{1}{\tau}\sum_{t=1}^{\tau}\gamma_V(t+i,t+j)\sigma_t^{-2}\tilde{I},$$

as $n\to\infty$. And

$$\frac{1}{n-r-s}Cov(\sum_{t=r+1}^{n-s}\frac{\partial g_t}{\partial\sigma_t},\sum_{k=r+1}^{n-s}\frac{\partial g_k}{\partial\sigma_k})=\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sigma_t^{-2}\tilde{J}\to\frac{1}{\tau}\sum_{t=1}^{\tau}\sigma_t^{-2}\tilde{J}.$$

Combining the preceding results, we conclude that

$$\frac{1}{n-r-s}Cov\left(\sum_{t=r+1}^{n-s}\frac{\partial g_t}{\partial\boldsymbol{\theta}_t},\left(\sum_{t=r+1}^{n-s}\frac{\partial g_t}{\partial\boldsymbol{\theta}_t}\right)^T\right)\to\Sigma_0,$$

where

$$\Sigma_0=\begin{bmatrix}A_0 & B_0\\ B_0^T & D_0\end{bmatrix}. \tag{4.2.7}$$

A_0 is a $r\times r$ symmetric matrix with $(i,j)^{th}$-element $\sigma_{ij}=\frac{1}{\tau}\sum_{t=1}^{\tau}\gamma_U(t-i,t-j)\sigma_t^{-2}\tilde{I}$ for $i,j=1,\cdots,r$. B_0 is a $r\times(s+1)$ matrix with $(i,j)^{th}$-element $\sigma_{ij}=\sum_{t=m_0-i}^{\infty}\alpha_t\beta_{t+i-j}$ for $i=1,\cdots,r$, $j=1,\cdots,s$ and zero otherwise. D_0 is a $(s+1)\times(s+1)$ symmetric matrix with element $\sigma_{ij}=\frac{1}{\tau}\sum_{t=1}^{\tau}\gamma_V(t+i,t+j)\sigma_t^{-2}\tilde{I}$ for $i,j=1,\cdots,s$, $\sigma_{s+1,s+1}=\frac{1}{\tau}\sum_{t=1}^{\tau}\sigma_t^{-2}\tilde{J}$ and zero otherwise.

Next we prove the asymptotic behaviour of $\frac{1}{\sqrt{n-r-s}}\sum_{t=r+1}^{n-s}\frac{\partial g_t}{\partial\boldsymbol{\theta}_t}$ in the following proposition.

Proposition 4.2.1. *If the probability density function $f(x)$ satisfies assumptions A1-A8, and the parameters $\{\alpha(t)\}$ and $\{\beta(k)\}$ are exponential decay, then*

$$\frac{1}{\sqrt{n-r-s}}\sum_{t=r+1}^{n-s}\frac{\partial g_t}{\partial\boldsymbol{\theta}_t}\to N(\mathbf{0},\Sigma_0), \tag{4.2.8}$$

in distribution, where Σ_0 is given in (4.2.7)

Proof. By the Cramér-Wold device, it suffices to prove that for any $\boldsymbol{a}\in\mathbb{R}^{r+s+1}$,

$$\frac{1}{\sqrt{n-r-s}}\sum_{t=r+1}^{n-s}\boldsymbol{a}^T\frac{\partial g_t}{\partial\boldsymbol{\theta}_t}\to N(0,\boldsymbol{a}^T\Sigma_0\boldsymbol{a}),$$

in distribution as $n\to\infty$.

For a large positive integer m, define $U_m(t) := \sum_{a=0}^{m} \alpha_a Z(t-a)$, and $V_m(t) := \sum_{b=0}^{m} \beta_b Z(t+b)$. Define

$$\boldsymbol{Y}_m(t) := \Bigg(-U_m(t-1)\frac{f_t^{'}}{f_t}, \cdots, -U_m(t-r)\frac{f_t^{'}}{f_t}, -V_m(t+1)\frac{f_t^{'}}{f_t}, \cdots, \\ -V_m(t+s)\frac{f_t^{'}}{f_t}, -\frac{1}{\sigma_t}\left(Z(t)\frac{f_t^{'}(Z(t))}{f_t(Z(t))} + 1 \right) \Bigg)^T.$$

Then for fixed m, $U_m(t)$ and $V_m(t)$ are still random periodic processes, so as $\boldsymbol{Y}_m(t)$.

Define $\Sigma_{0,m} := \frac{1}{n-r-s} Cov\left(\sum_{t=r+1}^{n-s} \mathbf{Y}_m(t), \left(\sum_{t=r+1}^{n-s} \mathbf{Y}_m(t)\right)^T \right)$. Then $\Sigma_{0,m} \to \Sigma_0$ as $m \to \infty$. By the central limit theorem for random periodic processes, under the conditions that $\{\alpha_i\}_{i\in\mathbb{Z}^+}$ and $\{\beta_j\}_{j\in\mathbb{Z}^+}$ are exponential decay and there exists a positive δ such that $\mathbb{E}[(Z(t))^{2+\delta}] < \infty$, there is

$$\frac{1}{\sqrt{n-r-s}} \sum_{t=r+1}^{n-s} \boldsymbol{a}^T \boldsymbol{Y}_m(t) \to N(0, \boldsymbol{a}^T \Sigma_{0,m} \boldsymbol{a})$$

in distribution as $n \to \infty$.

Since

$$\lim_{m\to\infty} \lim_{n\to\infty} Var\left(\frac{1}{\sqrt{n-r-s}} \sum_{t=r+1}^{n-s} \left(\boldsymbol{a}^T \boldsymbol{Y}_m(t) - \boldsymbol{a}^T \frac{\partial g_t}{\partial \boldsymbol{\theta}_t} \right) \right) = 0.$$

The convergence in (4.2.8) is immediate from Proposition 3.1.2. □

4.3 Asymptotic Normality

In this section, we will follow the idea in [10] to prove that there exists a sequence of solutions, $\hat{\boldsymbol{\theta}}^n$, to the likelihood equations,

$$\frac{\partial L_n(\boldsymbol{\theta})}{\partial \theta_j(t)} = 0, \qquad t = r+1, \cdots, n-s, j = 1, \cdots, r+s+1,$$

where L_n is given in (4.2.6), which is consistent with the true parameter value $\boldsymbol{\theta}^*$ in the sense of distribution. We represent this result in the following theorem.

Theorem 4.3.1. *For the non-causal autoregressive model (4.2.1) for a random periodic process $\{Y(t)\}$, there exists a sequence of solutions, $\hat{\boldsymbol{\theta}}^n$, to the likelihood equations (4.2.6) which satisfy*

$$\frac{\sqrt{n}}{\tau} \sum_{t=1}^{\tau} \Sigma_t (\hat{\boldsymbol{\theta}}_t^n - \boldsymbol{\theta}_t^*) \to N(\mathbf{0}, \Sigma_0),$$

in distribution as $n \to \infty$, where $\boldsymbol{\theta}_t^* = (\theta_1^*(t), \cdots, \theta_{r+s+1}^*(t))^T$ is the true values of parameters at time t and the weight matrix Σ_t is given in (4.3.2).

Proof. Define $h(x) := \frac{f'(x)}{f(x)}$. As in [10], we assume that $h'(x) = h_1(x) - h_2(x)$ where h_1 and h_2 are non-decreasing functions with $h_i(x) = O(|x|^\delta)$ as $|x| \to \infty$, where we have $\mathbb{E}|Z(t)|^{2+\delta} < \infty$. This implies $\mathbb{E}|Z(t)|^j |h'(Z(t))| < \infty$ for $j = 0, 1, 2$. By calculation, we have

$$h'(x) = \frac{d}{dx}\left(\frac{f'(x)}{f(x)}\right) = \frac{f''(x)f(x) - \left(f'(x)\right)^2}{\left(f(x)\right)^2},$$

$\frac{f_t'(x)}{f_t(x)} = \sigma_t^{-1} h(\frac{x}{\sigma_t})$, $\frac{d}{dx}\left(\frac{f_t'(x)}{f_t(x)}\right) = \sigma_t^{-2} h'(\frac{x}{\sigma_t})$ and $\mathbb{E}\left[h'\left(\frac{Z(t)}{\sigma_t}\right)\right] = -\tilde{I}$.

Expanding $L_n(\boldsymbol{\theta})$ in a neighbourhood of $\boldsymbol{\theta}^*$, we have

$$\begin{aligned}
& \frac{1}{n-r-s}(L_n(\boldsymbol{\theta}) - L_n(\boldsymbol{\theta}^*)) \\
= & \frac{1}{n-r-s} \sum_{i=1}^{r+s+1} \sum_{t=r+1}^{n-s} \frac{\partial L_n(\boldsymbol{\theta}^*)}{\partial \theta_i(t)} (\theta_i(t) - \theta_i^*(t)) \\
& + \frac{1}{2(n-r-s)} \sum_{i=1}^{r+s+1} \sum_{j=1}^{r+s+1} \sum_{t=r+1}^{n-s} \sum_{k=r+1}^{n-s} \frac{\partial^2 L_n(\boldsymbol{\theta}^*)}{\partial \theta_i(t) \partial \theta_j(k)} (\theta_i(t) - \theta_i^*(t))(\theta_j(k) - \theta_j^*(k)) \\
& + \frac{1}{2(n-r-s)} \sum_{i=1}^{r+s+1} \sum_{j=1}^{r+s+1} \sum_{t=r+1}^{n-s} \sum_{k=r+1}^{n-s} \left(\frac{\partial^2 L_n(\tilde{\boldsymbol{\theta}})}{\partial \theta_i(t) \partial \theta_j(k)} - \frac{\partial^2 L_n(\boldsymbol{\theta}^*)}{\partial \theta_i(t) \partial \theta_j(k)}\right) \\
& \times (\theta_i(t) - \theta_i^*(t))(\theta_j(k) - \theta_j^*(k)) \\
= & P_1 + P_2 + P_3,
\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}$ is between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$. By the ergodic theorem for random periodic processes in [20],

$$\begin{aligned}
P_1 = & \frac{1}{n-r-s} \sum_{i=1}^{r+s+1} \sum_{t=r+1}^{n-s} \frac{\partial g_t(\boldsymbol{\theta}^*)}{\partial \theta_i(t)} (\theta_i(t) - \theta_i^*(t)) \\
\to & \sum_{i=1}^{r+s+1} \frac{1}{\tau} \sum_{t=1}^{\tau} \mathbb{E}\left[\frac{\partial g_t(\boldsymbol{\theta}^*)}{\partial \theta_i(t)}\right] (\theta_i(t) - \theta_i^*(t)) = 0 \qquad \text{a.s.}
\end{aligned}$$

as $n \to \infty$.

Next we consider the second partial derivative term. For $i, j = 1, \cdots, r$,

$$\frac{1}{n-r-s} \sum_{t=r+1}^{n-s} \sum_{k=r+1}^{n-s} \frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial \theta_i(t) \partial \theta_j(k)}$$

$$
\begin{aligned}
&= \frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\phi_i\partial\phi_j} \\
&= \frac{1}{n-r-s}\sum_{t=r+1}^{n-s} -U(t-i)\frac{\partial}{\partial\phi_j}\left(\frac{f_t^{'}(Z(t))}{f_t(Z(t)}\right) \\
&= \frac{1}{n-r-s}\sum_{t=r+1}^{n-s} U(t-i)U(t-j)\sigma_t^{-2}h^{'}\left(\frac{Z(t)}{\sigma_t}\right).
\end{aligned}
$$

By ergodic theorem of random periodic processes, at the true value of parameter $\boldsymbol{\theta}^*$,

$$
\begin{aligned}
&\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta}^*)}{\partial\theta_i(t)\partial\theta_j(k)} \\
\to\ &\frac{1}{\tau}\sum_{t=1}^{\tau}\mathbb{E}\left[U^*(t-i)U^*(t-j)(\sigma_t^*)^{-2}h^{'}\left(\frac{Z(t)}{\sigma_t^*}\right)\right] \\
=\ &-\frac{1}{\tau}\sum_{t=1}^{\tau}\gamma_U(t-i,t-j)(\sigma_t^*)^{-2}\tilde{I} \qquad \text{a.s.}
\end{aligned}
$$

as $n\to\infty$. Similarly, for $i,j=1,\cdots,s$, we have

$$
\begin{aligned}
&\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial\theta_{i+r}(t)\partial\theta_{j+r}(k)} \\
=\ &\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\varphi_i\partial\varphi_j} \\
=\ &\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}V(t+i)V(t+j)\sigma_t^{-2}h^{'}\left(\frac{Z(t)}{\sigma_t}\right) \\
\to\ &-\frac{1}{\tau}\sum_{t=1}^{\tau}\gamma_V(t+i,t+j)(\sigma_t^*)^{-2}\tilde{I} \qquad \text{a.s.}
\end{aligned}
$$

as $n\to\infty$.

For $i=1,\cdots,r$ and $j=1,\cdots,s$,

$$
\begin{aligned}
&\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial\theta_i(t)\partial\theta_{j+r}(k)} \\
=\ &\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\phi_i\partial\varphi_j} \\
=\ &\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\left[-U(t-i)\frac{\partial}{\partial\varphi_j}\left(\frac{f_t^{'}(Z(t))}{f_t(Z(t))}\right)+\sum_{t=r+1+i}^{n-s}-\frac{f_t^{'}(Z(t))}{f_t(Z(t))}\frac{\partial}{\partial\varphi_j}(U(t-i))\right] \\
=\ &\frac{1}{n-r-s}\left[\sum_{t=r+1}^{n-s}U(t-i)V(t+j)\sigma_t^{-2}h^{'}\left(\frac{Z(t)}{\sigma_t}\right)+\sum_{t=r+1+i}^{n-s}Y(t-i+j)\sigma_t^{-1}h\left(\frac{Z(t)}{\sigma_t}\right)\right].
\end{aligned}
$$

As $\mathbb{E}\left[U(t-i)V(t+j)h'\left(\frac{Z(t)}{\sigma_t}\right)\right]=0$ and

$$\begin{aligned}&\mathbb{E}\left[Y(t-i+j)\sigma_t^{-1}h\left(\frac{Z(t)}{\sigma_t}\right)\right]\\ =\ &\mathbb{E}\left[\sum_{a=-\infty}^{\infty}\psi_a Z(t-i+j-a)\frac{f_t'(Z(t))}{f_t(Z(t))}\right]\\ =\ &-\psi_{j-i},\end{aligned}$$

by ergodic theorem, at point $\boldsymbol{\theta}^*$,

$$\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial\theta_i(t)\partial\theta_{l+r}(k)}\to-\frac{1}{\tau}\sum_{t=1}^{\tau}\psi_{j-i}\qquad\text{a.s.}$$

as $n\to\infty$. Similarly, as $n\to\infty$, for $i=1,\cdots,s$ and $j=1,\cdots,r$,

$$\begin{aligned}&\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial\theta_{i+r}(t)\partial\theta_j(k)}\\ =\ &\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\varphi_i\partial\phi_j}\\ =\ &\frac{1}{n-r-s}\left[\sum_{t=r+1}^{n-s}Y(t+i-j)\sigma_t^{-1}h\left(\frac{Z(t)}{\sigma_t}\right)+\sum_{t=r+1+i}^{n-s}V(t+i)U(t-j)\sigma_t^{-2}h'\left(\frac{Z(t)}{\sigma_t}\right)\right]\\ \to\ &-\frac{1}{\tau}\sum_{t=1}^{\tau}\psi_{i-j}\qquad\text{a.s.}\end{aligned}$$

Besides, by calculations, as $n\to\infty$, for $i=1,\cdots,r$,

$$\begin{aligned}&\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial\theta_i(t)\partial\sigma_k}=\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial\sigma_t\partial\theta_i(k)}\\ =\ &\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\phi_i\partial\sigma_t}=\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\sigma_t\partial\phi_i}\\ =\ &\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sigma_t^{-2}U(t-i)\left[h\left(\frac{Z(t)}{\sigma_t}\right)+\frac{Z(t)}{\sigma_t}h'\left(\frac{Z(t)}{\sigma_t}\right)\right]\\ \to\ &\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sigma_t^{-2}\mathbb{E}\left\{U(t-i)\left[h\left(\frac{Z(t)}{\sigma_t}\right)+\frac{Z(t)}{\sigma_t}h'\left(\frac{Z(t)}{\sigma_t}\right)\right]\right\}=0\qquad\text{a.s.}\end{aligned}$$

For $i=1,\cdots,s$,

$$\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial\theta_{i+r}(t)\partial\sigma_k}=\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial\sigma_t\partial\theta_{i+r}(k)}$$

$$
\begin{aligned}
&= \frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\varphi_i\partial\sigma_t} = \frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\frac{\partial^2 g_t(\boldsymbol{\theta})}{\partial\sigma_t\partial\varphi_i}\\
&= \frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sigma_t^{-2}V(t+i)\left[h\left(\frac{Z(t)}{\sigma_t}\right)+\frac{Z(t)}{\sigma_t}h'\left(\frac{Z(t)}{\sigma_t}\right)\right]\\
&\to 0 \qquad\qquad \text{a.s.}
\end{aligned}
$$

And

$$
\begin{aligned}
&\frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial\sigma_t\partial\sigma_k}\\
&= \frac{1}{n-r-s}\sum_{t=r+1}^{n-s}\sigma_t^{-2}\left[2\frac{Z(t)}{\sigma_t}h\left(\frac{Z(t)}{\sigma_t}\right)+\left(\frac{Z(t)}{\sigma_t}\right)^2 h'\left(\frac{Z(t)}{\sigma_t}\right)+1\right]\\
&\to \frac{1}{\tau}\sum_{t=1}^{\tau}\sigma_t^{-2}\tilde{J} \qquad\qquad \text{a.s.}
\end{aligned}
$$

Combining previous results, we have at the true parameter values $\boldsymbol{\theta}^*$

$$
\frac{1}{n-r-s}\frac{\partial^2 L_n(\boldsymbol{\theta}^*)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}} \to -\frac{1}{\tau}\sum_{t=1}^{\tau}\Sigma_t. \tag{4.3.1}
$$

For each $t = 1, \cdots, \tau$,

$$
\Sigma_t = \begin{bmatrix} A_t & B_t \\ B_t^T & D_t \end{bmatrix}. \tag{4.3.2}
$$

A_t is a $r\times r$ symmetric matrix with $(i,j)^{th}$-element $\sigma_{ij} = \gamma_U(t-i, t-j)(\sigma_t^*)^{-2}\tilde{I}$ for $i, j = 1, \cdots, r$. B_t is a $r\times(s+1)$ matrix with $(i,j)^{th}$-element $\sigma_{ij} = \psi_{j-i}$ for $i = 1, \cdots, r$, $j = 1, \cdots, s$. D_0 is a $(s+1)\times(s+1)$ symmetric matrix with element $\sigma_{ij} = \gamma_V(t+i, t+j)(\sigma_t^*)^{-2}\tilde{I}$ for $i, j = 1, \cdots, s$, $\sigma_{s+1,s+1} = (\sigma_t^*)^{-2}\tilde{J}$ and zero otherwise. Therefore,

$$
P_2 \to -\frac{1}{2\tau}\sum_{t=1}^{\tau}(\boldsymbol{\theta}_t-\boldsymbol{\theta}_t^*)'\,\Sigma_t\,(\boldsymbol{\theta}_t-\boldsymbol{\theta}_t^*) \qquad \text{a.s.}
$$

By the same idea in [10] and ergodic theorem for random periodic processes, we can prove that

$$
\limsup_{n\to\infty}\sup_{\boldsymbol{\theta}\in Q_\epsilon}\frac{1}{n-r-s}\left|\sum_{t=r+1}^{n-s}\sum_{k=r+1}^{n-s}\frac{\partial^2 L_n(\boldsymbol{\theta})}{\partial\theta_i(t)\partial\theta_j(k)}-\frac{\partial^2 L_n(\boldsymbol{\theta}*)}{\partial\theta_i(t)\partial\theta_j(k)}\right| \to 0 \qquad \text{a.s.} \tag{4.3.3}
$$

as the radius of the neighbourhood of $\boldsymbol{\theta}^*$ satisfies $\epsilon\to\infty$ for each $i, j = 1, \cdots, r+s+1$.

Therefore we conclude that for ϵ small,

$$\sup_{\boldsymbol{\theta} \in \partial Q_\epsilon} (P_1 + P_2 + P_3) < 0 \qquad \text{a.s.}$$

as $n \to \infty$. Hence for large n, $L_n(\boldsymbol{\theta}) < L_n(\boldsymbol{\theta}^*)$ a.s. There exists a sequence of local maximum $\hat{\boldsymbol{\theta}}^n$ to L_n which converge to $\boldsymbol{\theta}^*$ a.s.

To explore the asymptotic behaviour of $\hat{\boldsymbol{\theta}}^n$, we expand $\frac{\partial L_n(\hat{\boldsymbol{\theta}}^n)}{\partial \boldsymbol{\theta}}$ at $\boldsymbol{\theta}^*$ as

$$0 = \frac{1}{\sqrt{n}} \frac{\partial L_n(\hat{\boldsymbol{\theta}}^n)}{\partial \boldsymbol{\theta}} = \frac{1}{\sqrt{n}} \sum_{t=r+1}^{n-s} \frac{\partial g_t(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}_t} + \frac{1}{\sqrt{n}} \frac{\partial^2 L_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \sqrt{n}(\hat{\boldsymbol{\theta}}^n - \boldsymbol{\theta}^*).$$

Since $\tilde{\boldsymbol{\theta}} \to \boldsymbol{\theta}^*$ a.s. By equation (4.3.3) and ergodic theorem for random periodic processes,

$$\frac{1}{\sqrt{n}} \frac{\partial^2 L_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} = \frac{1}{\sqrt{n}} \frac{\partial^2 L_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} + \frac{1}{\sqrt{n}} \left(\frac{\partial^2 L_n(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} - \frac{1}{\sqrt{n}} \frac{\partial^2 L_n(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}} \right) \to -\frac{1}{\tau} \sum_{t=1}^{\tau} \Sigma_t \text{ a.s.}$$

Hence by Proposition 4.2.1,

$$\frac{\sqrt{n}}{\tau} \sum_{t=1}^{\tau} \Sigma_t (\hat{\boldsymbol{\theta}}_t^n - \boldsymbol{\theta}_t^*) \to N(\mathbf{0}, \Sigma_0),$$

in distribution as $n \to \infty$. □

Chapter 5

Simulation Results

For an observed time series, to fit it with ARMA model for random periodic processes, our estimation procedure is listed below,

1. Assume $Y(t)$ is causal. Consider the corresponding MA($K\tau$) model for random periodic processes with $K\tau < w$,

$$Y_t = \sum_{j=0}^{K\tau} \psi_j(t) Z(t-j), \qquad t = 1, \cdots . \tag{5.0.1}$$

2. Use truncated innovation algorithm to estimate the coefficients of ψ.

3. By (3.3.3) calculate the coefficients of $\psi_0(t)Z_t = \sum_{j=0}^{m} \pi_j(t)Y(t-j)$, then use history data to estimate history noise.

4. For each pair of (p, q) satisfying $p + q \leq K\tau$, calculate the coefficients of the corresponding ARMA(p, q) model by (2.2.8).

5. Determine the order of the most suitable model by model criteria.

In the following sections, we will introduce some model criteria to help us determine the most suitable model.

5.1 One-step Prediction of ARMA(p,q) Model for Random Periodic Processes

For the ARMA(p, q) process (2.2.1), recall the definition of the transformed process (cf. [5]),

$$W(t) = \begin{cases} \theta_0^{-1}(t)Y(t), & t = 1, \cdots, m, \\ \theta_0^{-1}(t)\left(Y(t) - \sum_{i=1}^{p}\phi_i(t)Y(t-i)\right) & t > m, \end{cases} \tag{5.1.1}$$

where $m = \max(p, q)$. We will use this transformed process to find the one-step prediction of the corresponding ARMA process. The autocovariance functions γ_W of $W(t)$ are calculated as follows: for $t \geq s$,

(i) when $t \leq m$, $\gamma_W(t,s) = \mathbb{E}[\theta_0^{-1}(t)Y(t)\theta_0^{-1}(s)Y(s)] = \theta_0^{-1}(t)\theta_0^{-1}(s)\gamma_Y(t,s)$;

(ii) when $t > m$ and $s \leq m$,

$$\begin{aligned} \gamma_W(t,s) &= \mathbb{E}\left[\theta_0^{-1}(t)\left(Y(t) - \sum_{i=1}^{p}\phi_i(t)Y(t-i)\right)\theta_0^{-1}(s)Y(s)\right] \\ &= \theta_0^{-1}(t)\theta_0^{-1}(s)\left(\gamma_Y(t,s) - \sum_{i=1}^{p}\phi_i(t)\gamma_Y(t-i,s)\right); \end{aligned}$$

(iii) when $s > m$ and $t - q \leq s$,

$$\begin{aligned} \gamma_W(t,s) &= \mathbb{E}\left[\theta_0^{-1}(t)\sum_{i=0}^{q}\theta_i(t)Z(t-i)\theta_0^{-1}(s)\sum_{j=0}^{q}\theta_j(s)Z(s-j)\right] \\ &= \theta_0^{-1}(t)\theta_0^{-1}(s)\sum_{i=0}^{q}\theta_i(t)\sum_{j=0}^{q}\theta_j(s)\mathbb{E}[Z(t-i)Z(s-j)] \\ &= \theta_0^{-1}(t)\theta_0^{-1}(s)\sum_{j=0}^{s-t+q}\theta_{t-s+j}(t)\theta_j(s); \end{aligned}$$

(iv) when $s > m$ and $t - q > s$, $\gamma_W(t,s) = 0$.

Applying the innovation algorithm in Proposition 3.3.2 to $W(t)$, one obtains the one-step predictor $\hat{W}(n+1) := \mathbb{E}[W(n+1)|\sigma(W(1), \cdots, W(n))]$ as follows,

$$\begin{cases} \hat{W}(n+1) = \sum_{j=1}^{n} b_j(n+1)(W(n+1-j) - \hat{W}(n+1-j)) & 1 \leq n < m, \\ \hat{W}(n+1) = \sum_{j=1}^{q} b_j(n+1)(W(n+1-j) - \hat{W}(n+1-j)) & n \geq m, \end{cases} \tag{5.1.2}$$

where the coefficients $b_j(t)$ and $r_t := \mathbb{E}[W(t) - \hat{W}(t)]$ are calculated by using γ_W.

When $n \geq m$, $\hat{W}(n+1)$ only depends on the previous q of $W(n+1-j) - \hat{W}(n+1-j)$, the coefficients $b_j(t) = 0$ for $j > q$. This is because when $t - s > q$, $\gamma_W(t,s) = 0$. For example,

$$\begin{cases} b_{q+2}(q+3) = \dfrac{\gamma_W(q+2,1)}{v_1} = 0 \\ b_{q+1}(q+3) = \dfrac{1}{v_2}\left(\gamma_W(q+3,2) - b_0(2)b_{q+2}(q+3)v_1\right) = 0 \\ b_q(q+3) = \dfrac{1}{v_3}\left(\gamma_W(q+3,3) - b_1(3)b_{q+2}(q+3)v_1 - b_0(3)b_{q+1}(q+3)v_2\right) = \dfrac{\gamma_W(q+3,3)}{v_3} \\ \cdots\cdots \end{cases}.$$

From the construction of $W(t)$, we have

$$\mathcal{H}_1^n := \sigma(Y(1), \cdots, Y(n)) = \sigma(W(1), \cdots, W(n)),$$

hence

$$\hat{W}(t) = \mathbb{E}[W(t)|\mathcal{H}_1^{t-1}] = \begin{cases} \theta_0^{-1}(t)\hat{Y}(t), & 1 \leq t \leq m \\ \theta_0^{-1}(t)\left(\hat{Y}(t) - \sum_{i=1}^{p} \phi_i(t)Y(t-i)\right) & t > m \end{cases}, \tag{5.1.3}$$

Moreover, for $t > m$, we have

$$\begin{aligned} & W(t) - \hat{W}(t) \\ = \;& \theta_0^{-1}(t)\left(Y(t) - \sum_{i=1}^{p} \phi_i(t)Y(t-i)\right) - \theta_0^{-1}(t)\left(\hat{Y}(t) - \sum_{i=1}^{p} \phi_i(t)Y(t-i)\right) \\ = \;& \theta_0^{-1}(t)(Y(t) - \hat{Y}(t)), \end{aligned}$$

together with $W(t)-\hat{W}(t)=\theta_0^{-1}(t)(Y(t)-\hat{Y}(t))$ for $t\le m$. Therefore, substituting $\hat{W}(t)$ into (5.1.2), we obtain the one-step prediction $\hat{Y}(n+1)$ as follows,

$$\begin{cases} \hat{Y}(n+1)=\sum_{j=1}^{n} b_j(n+1)(Y(n+1-j)-\hat{Y}(n+1-j)) & 1\le n<m, \\ \hat{Y}(n+1)=\sum_{i=1}^{p}\phi_i(n+1)Y(n+1-i) & \\ \qquad +\sum_{j=1}^{q} b_j(n+1)(Y(n+1-j)-\hat{Y}(n+1-j)) & n\ge m, \end{cases} \tag{5.1.4}$$

and

$$\mathbb{E}[Y(n+1)-\hat{Y}(n+1)]^2=\theta_0^2(n+1)\mathbb{E}[W(n+1)-\hat{W}(n+1)]^2=\theta_0^2(n+1)r_{n+1}. \tag{5.1.5}$$

5.2 Order Selection Criteria

In this section we recall the standard AIC, AICc and BIC criteria for model selection. First we introduce the computation of the likelihood of ARMA(p,q) process. We will use them to verify our model selection. Examples will be given in the next section.

According to the idea in [12], we assume $\{Y(t)\}$ is Gaussian process. For $\boldsymbol{Y}_n := (Y(1),\cdots,Y(n))^T$, denote $\Gamma_n:=\mathbb{E}[\boldsymbol{Y}_n^T\boldsymbol{Y}_n]$, the likelihood is

$$L=(2\pi)^{-\frac{n}{2}}(\det\Gamma_n)^{-\frac{1}{2}}\exp(-\frac{1}{2}\boldsymbol{Y}_n^T\Gamma_n^{-1}\boldsymbol{Y}_n).$$

The determinant and inverse matrix of Γ_n can be represented by the parameters b and v calculated by the innovation algorithm. Set $C_n:=[b_{i-j}(i)]_{i,j=1}^n$, where define $b_0(i)=1$ and $b_{i-j}(i)=0$ for $i<j$. Also set $D_n:=diag(v_1,v_2,\cdots,v_n)$. Then the innovation representation of $Y(t)$ can be represented as

$$\hat{\boldsymbol{Y}}_n=(C_n-I_n)(\boldsymbol{Y}_n-\hat{\boldsymbol{Y}}_n)=C_n(\boldsymbol{Y}_n-\hat{\boldsymbol{Y}}_n)-\boldsymbol{Y}_n+\hat{\boldsymbol{Y}}_n.$$

Then we have $\boldsymbol{Y}_n=C_n(\boldsymbol{Y}_n-\hat{\boldsymbol{Y}}_n)$. Multiplying $\boldsymbol{Y}_n$ and taking expectation of both sides, we have

$$\Gamma_n=\mathbb{E}[\boldsymbol{Y}_n\boldsymbol{Y}_n^T]=C_n\mathbb{E}[(\boldsymbol{Y}_n-\hat{\boldsymbol{Y}}_n)(\boldsymbol{Y}_n-\hat{\boldsymbol{Y}}_n)^T]C_n^T=C_nD_nC_n^T.$$

Hence

$$\det\Gamma_n=(\det C_n)^2(\det D_n)=v_1v_2\cdots v_n,$$

and

$$\begin{aligned}
&\boldsymbol{Y}_n^T \Gamma_n^{-1} \boldsymbol{Y}_n \\
=\ & (\boldsymbol{Y}_n - \hat{\boldsymbol{Y}}_n)^T C_n^T (C_n D_n C_n^T)^{-1} C_n (\boldsymbol{Y}_n - \hat{\boldsymbol{Y}}_n) \\
=\ & (\boldsymbol{Y}_n - \hat{\boldsymbol{Y}}_n)^T D_n^{-1} (\boldsymbol{Y}_n - \hat{\boldsymbol{Y}}_n) \\
=\ & \sum_{i=1}^{n} \frac{(Y(i) - \hat{Y}(i))^2}{v_i}.
\end{aligned}$$

Therefore the likelihood function is rewritten as

$$L = (2\pi)^{-\frac{n}{2}} (v_1 v_2 \cdots v_n)^{-\frac{1}{2}} \exp(-\frac{\sum_{i=1}^{n} (Y(i) - \hat{Y}(i))^2}{2v_i}).$$

For a pair of (p, q), we calculated the corresponding parameters $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$, then we can use the innovation algorithm and (5.1.4) to calculate $\hat{Y}$. Therefore the likelihood L is regarded as a function of $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$,

$$L(\boldsymbol{\phi}, \boldsymbol{\theta}) = (2\pi)^{-\frac{n}{2}} (\theta_0(1) \cdots \theta_0(n))^{-1} (r_1 r_2 \cdots r_n)^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(Y(i) - \hat{Y}(i))^2}{\theta_0^2(i) r_i} \right). \tag{5.2.1}$$

The log-likelihood function is

$$lnL(\boldsymbol{\phi}, \boldsymbol{\theta}) = -\frac{n}{2} ln(2\pi) - \sum_{i=1}^{n} ln(\theta_0(i)) - \frac{1}{2} \sum_{i=1}^{n} r_i - \frac{1}{2} \sum_{i=1}^{n} \frac{(Y(i) - \hat{Y}(i))^2}{\theta_0^2(i) r_i}.$$

When computing the likelihood of $\{Y(t)\}$ by computer, we usually use the log-likelihood form in order to avoid the divisors being too small to be recognized as zero.

Although $\{Y(t)\}$ are not i.i.d. and not Gaussian, we can also use the likelihood function (5.2.1) as a measure of choosing the parameters by maximizing it.

The AIC (Akaike Information Criterion) is first introduced by statistician Hirotugu Akaike in [1]. It is using the likelihood of model with different parameters to estimate the relative information loss by such model. The model with the smallest AIC value is chosen. However, when the sample size is small, there is a tendency that AIC will overfit the models, that is, it will prefer model with larger parameters. The AICc was developed with a more strict penalty term for large number of parameters for small sample sizes. We will see that as the sample size $n \to \infty$, AICc will tend to AIC. The BIC (Bayesian Information Criterion) is another criterion with larger penalty term. The formulas of these three criteria are as follows,

$$AIC(\boldsymbol{\phi}, \boldsymbol{\theta}) = 2k - 2\ln(L),$$

$$AICc(\boldsymbol{\phi}, \boldsymbol{\theta}) = AIC + \frac{2k^2 + 2k}{n - k - 1},$$

and

$$BIC(\boldsymbol{\phi}, \boldsymbol{\theta}) = \ln(n)k - 2\ln(L),$$

where k is the number of parameters in the model.

5.3 Simulation Procedure

For the ARMA model for random periodic processes, we analyze the asymptotic behaviours of the coefficients and give the corresponding optimisation algorithm in the previous sections. We aim to write it as a computer programme and realize inputting new data one by one and outputting prediction results as soon as possible, i.e. machine learning. In the following sections we give two examples of simulation and in the appendix we give the main functions written in R language which we used in the simulation.

To do simulation with the ARMA model for random periodic processes, inspired by the idea of machine learning, we set the procedure of estimation in three stages. The first stage is to estimate w in the sample autocovariance function. This stage will end if the autocovariance function with respect to w shows convergence tendency. In the second stage we estimate the coefficients of MA($K\tau$) model for random periodic processes based on the truncated innovation algorithm, we will see that K is chosen to be 1 in most of the cases. We will use model fit criteria to compare each ARMA(p, q) model with $p + q \leq K\tau$ and choose the one with smallest value of model fit criteria. Recall that we will use MA($K\tau$) model to obtain the corresponding AR(m) model and use it to estimate the history noise by data we have already known. We will test which m is suitable in the third stage. The procedure is to predict by ARMA(p, q) model we choose in the second stage with the history noise estimated by AR(m) model and compare the result with the real data. We choose the m which minimizes the mean-square error between predicted data and the real one.

5.4 Example of Temperature

Many business activities and people's livelihood are influenced greatly by weather, for example, energy production and consumption, agricultural commodities production, airline passengers and among others. Even accident deaths may be influenced by extreme weather conditions. Nowadays, people are establishing new type of security called weather derivatives to help hedging their risks against weather-driven poor performance of business activities. The payoffs of these instruments may be linked to various weather-related variables, including heating degree days, cooling degree days, maximum temperature, minimum temperature, humidity, sunshine and precipitation (rainfall, snow-fall) etc. (Campbell *et al.* [14]). The market of weather derivatives grows rapidly in America, and has great potential in Europe. Weather forecasting is getting more and more crucial to guiding people's activities and even to government, like setting disaster prevention budget such as snow-removal cost.

People are seeking methods to modelling the daily temperature. Temperatures in different cities probably need different fitting models. Dornier and Querel [15] proposed a mean-reverting Ornstein-Uhlenbeck stochastic process to model the daily temperature. Some extensions of this model type were studied later. Alaton *et al.* [2] studied the Ornstein-Uhlenbeck model and observed that the quadratic variation $\sigma^2(t)$ is nearly constant over each month in the data set. They chose a piecewise constant function to represent the monthly variation in volatility. However, a statistical test for the normality of the residuals was not provided. Brody *et al.* [13] suggested a fractional Brownian motion replacement, and Benth *et al.* [6] suggested a *Lévy* process replacement. They also suggested to use an autoregressive conditional heteroscedastic (ARCH) dynamics with seasonal and cycle components to describe the residuals. Also, Campbell *et al.* [14] studied the non-structural model to estimate temperature of seven cities in America. They emphasized the capacity of long-horizon forecasting of the model.

In this thesis we try this mean-reverting stochastic process model described in Benth *et al.* [7] to model the daily temperature case and compare the result with that of our periodic ARMA model. Suppose the mean monthly temperature $T^m(t)$ satisfies a deterministic function of time t,

$$T^m(t) = A + Bt + Ccos(\omega t + \phi). \tag{5.4.1}$$

As temperature $T(t)$ varies along its mean value, it is modelled by a stochastic process solution of the following SDE

$$dT(t) = dT^m(t) - [a(T(t) - T^m(t))]\, dt + \sigma(t)dW(t).$$

The term $dT^m(t)$ guarantees that the process really reverts to the mean $T^m(t)$ (Alaton *et al.* [2]). The explicit solution is given as

$$T(t) = T^m(t) + [T(0) - T^m(0)]\, e^{-at} + \int_0^t \sigma(s)e^{-a(t-s)}dW(s).$$

Discretizing the equation, we obtain

$$\begin{aligned}
& T(t+1) - T(t) \\
= \ & T^m(t+1) + [T(0) - T^m(0)]\, e^{-a(t+1)} + \int_0^{t+1} \sigma(s)e^{-a(t+1-s)}dW(s) - \{T^m(t) \\
& + [T(0) - T^m(0)]\, e^{-at} + \int_0^t \sigma(s)e^{-a(t-s)}dW(s)\Big\} \\
= \ & [T^m(t+1) - T^m(t)] - (1 - e^{-a})e^{-at}\, [T(0) - T^m(0)] - (1 - e^{-a}) \int_0^t \sigma(s)e^{-a(t-s)}dW(s) \\
& + e^{-a} \int_t^{t+1} \sigma(s)e^{-a(t-s)}dW(s) \\
= \ & [T^m(t+1) - T^m(t)] - (1 - e^{-a})e^{-at}\, [T(t) - T^m(t)] + e^{-a} \int_t^{t+1} \sigma(s)e^{-a(t-s)}dW(s).
\end{aligned}$$

Approximating the integral part, we have

$$e^{-a} \int_t^{t+1} \sigma(s)e^{-a(t-s)}dW(s) \approx e^{-a}\sigma(t)[W(t+1) - W(t)].$$

Let $\tilde{T}(t) := T(t) - T^m(t)$, then we consider the following model

$$\tilde{T}(t+1) = \phi\tilde{T}(t) + \tilde{\sigma}(t)Z(t), \tag{5.4.2}$$

where $\phi := e^{-a}$, $\tilde{\sigma}(t) := e^{-a}\sigma(t)$ and $Z(t) \sim N(0, 1)$. One can use ARMA(1,0) model to estimate the coefficient ϕ first. $\tilde{\sigma}(t)$ can be estimated from the squared residuals.

Daily temperature can be seen as a good example of random periodic process in real life. If we eliminate the first order trend term $A + Bt$ from the original data, consider the average temperature $T^m(t)$ and $\sigma(t)$ satisfying periodic functions of t, then the solution $T(t)$ is a random periodic solution of SDE. Later we will see the periodicity of $\sigma(t)$ by autocorrelation function of squared residuals of model (5.4.2).

The data set we used in this example is the daily maximum temperature of central England obtained from Met Office. The range of the data process is 140 years from Jan. 1878 to Dec. 2017 with length 51100 and period 365. We eliminate every 29th Feb from the sample in leap years. Part of the data is plotted below. Daily CET values are expressed in tenths of a degree.

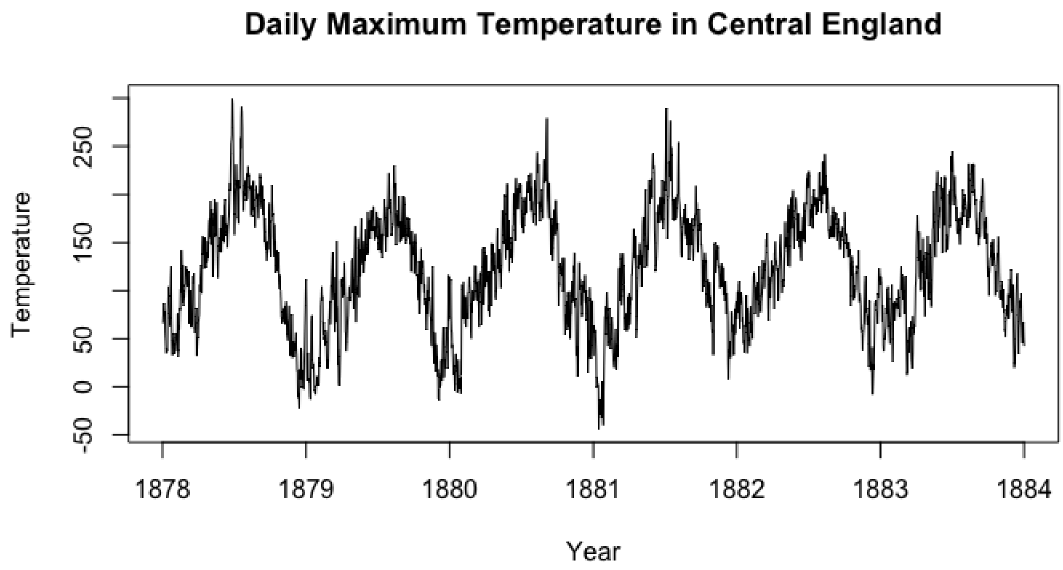


Figure 5.1: Central England Temperature.

We use (5.4.1) to estimate the mean of daily maximum temperature. The regression result we obtained and the plot are displayed below,

$$T^m(t) = 124.3 + 2.853e^{-4}t - 0.72233\cos(\frac{2\pi}{365}t + 24.8). \tag{5.4.3}$$

The coefficient of first order term seems to play a crucial role in the process. In average the temperature of central England will rise around 0.1 °C every ten years. Such trend increase may due to the Greenhouse effect, or development and air pollution which increase the urban temperature in general.

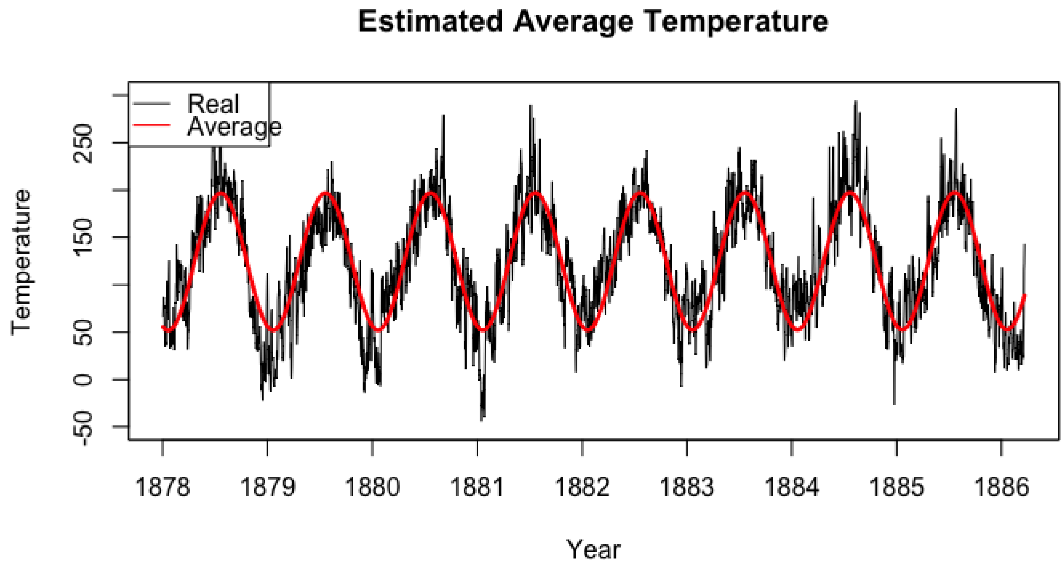


Figure 5.2: Estimated mean values of daily maximum temperature.

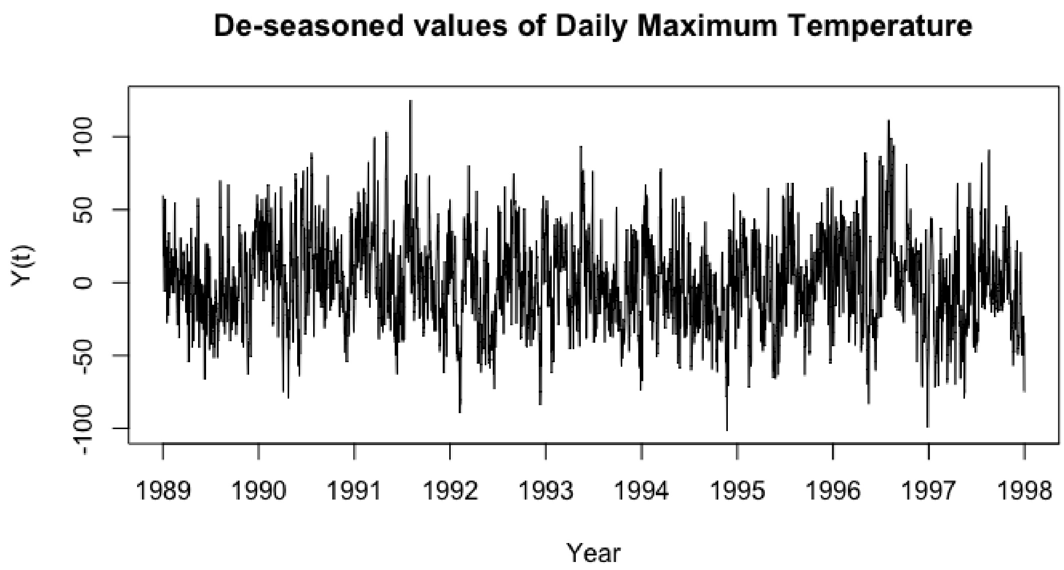


Figure 5.3: De-seasoned values of the daily maximum temperature.

Figure 5.2 displays the estimated average maximum daily temperature by (5.4.3) in red curve. It approximately describes the evolution of daily temperature around an average.

Figure 5.3 shows the de-seasoned values of the daily maximum temperature. There is no obvious trend or non-stationary pattern shown in the figure. Augmented Dickey-Fuller test rejects the non-stationary hypothesis with $p - value = 0.01$. Since we will use the sample from 40151 to 43800 to estimate the coefficients of ARMA(p, q) model for random periodic processes later, here we only use this slot of time to estimate AR(1) model in order to keep consistent with ARMA model for random periodic processes with respect to each quantities. The result is shown below. This gives us the estimated value of ϕ in (5.4.2) is 0.7768.

```
Call:
arima(x = data_tr[c(40151:43800)], order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.7768     1.4581
s.e.  0.0104     1.4609

sigma^2 estimated as 388.9:  log likelihood = -16062.64,  aic = 32131.29
```

Figure 5.4: AR(1) model of de-seasoned temperature.

Figure 5.5 and 5.6 shows the autocorrelation functions for residuals and squared residuals of AR(1) model. The first several lags of the autocorrelation functions for residuals is significant beyond the confidence bounds of zero, which shows that a higher-order autoregression model may be taken consideration. The autocorrelation functions for squared residuals displays a slight periodic pattern, which reveals a time dependency in the variance of the residuals.

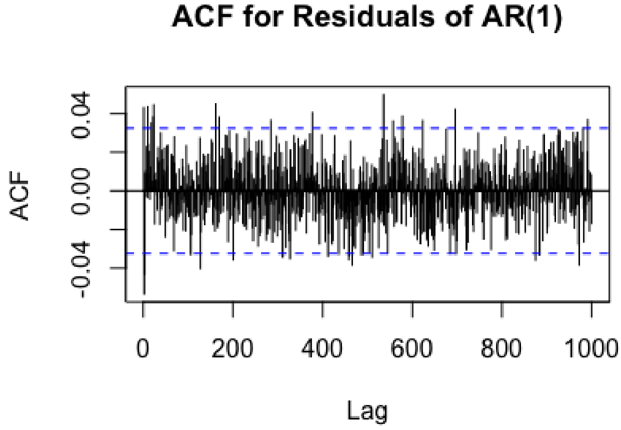


Figure 5.5: Autocorrelation function for residuals.

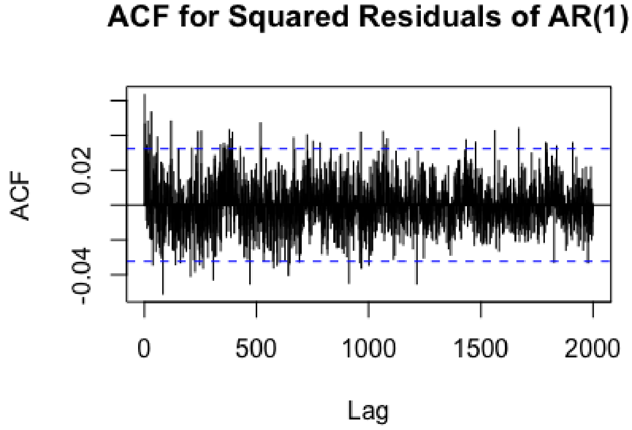


Figure 5.6: Autocorrelation function for squared residuals.

We use the following method described in the paper of Benth *et al.* [7] to estimate $\tilde{\sigma}(t)$. First we calculate the empirical values of variance by averaging the squared residuals in each day. This gives us 365 values. Then we use a Fourier series of lag 4, i.e. (5.4.4), to fit the empirical values. The results is shown in Table 5.1 and Figure 5.7. We observe that the estimated variance by Fourier series shows slight oscillation, which matches the appearance of the autocorrelation function of squared residuals. One may consider the variance to be constant within allowed range of error for simplicity.

$$\begin{aligned}\tilde{\sigma}^2(t) \approx\ & C_1\sin(\frac{2\pi}{365}t) + D_1\cos(\frac{2\pi}{365}t) + C_2\sin(\frac{4\pi}{365}t) + D_2\cos(\frac{4\pi}{365}t) + C_3\sin(\frac{6\pi}{365}t) \\ & + D_3\cos(\frac{6\pi}{365}t) + C_4\sin(\frac{8\pi}{365}t) + D_4\cos(\frac{8\pi}{365}t) + C. \qquad (5.4.4)\end{aligned}$$

Table 5.1: The coefficients of Fourier series.

C	C_1	C_2	C_3	C_4	D_1	D_2	D_3	D_4
388.90	42.10	-57.86	-19.57	49.67	-16.84	41.98	11.59	10.59

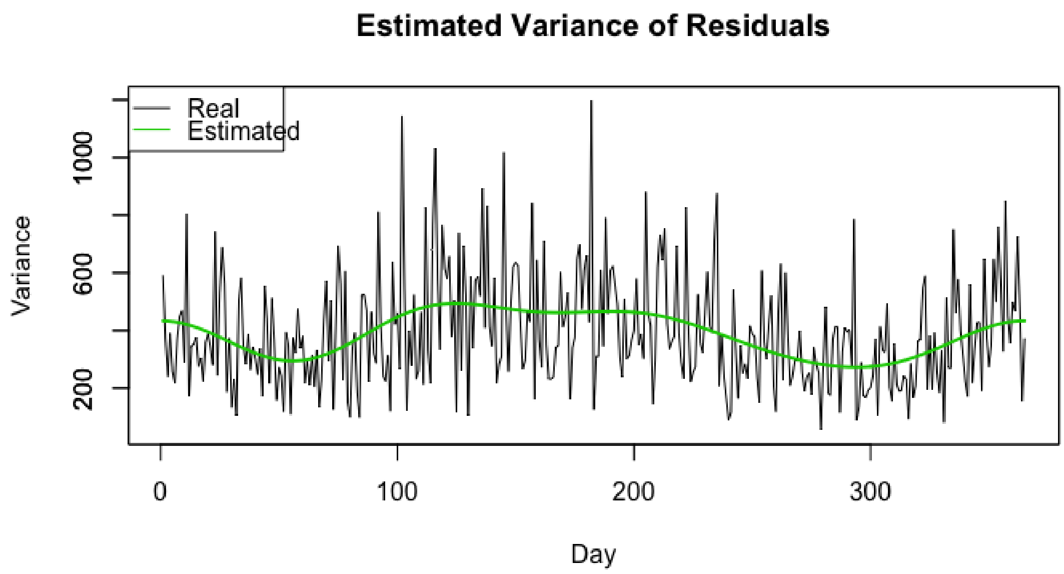


Figure 5.7: Estimated variance of residuals.

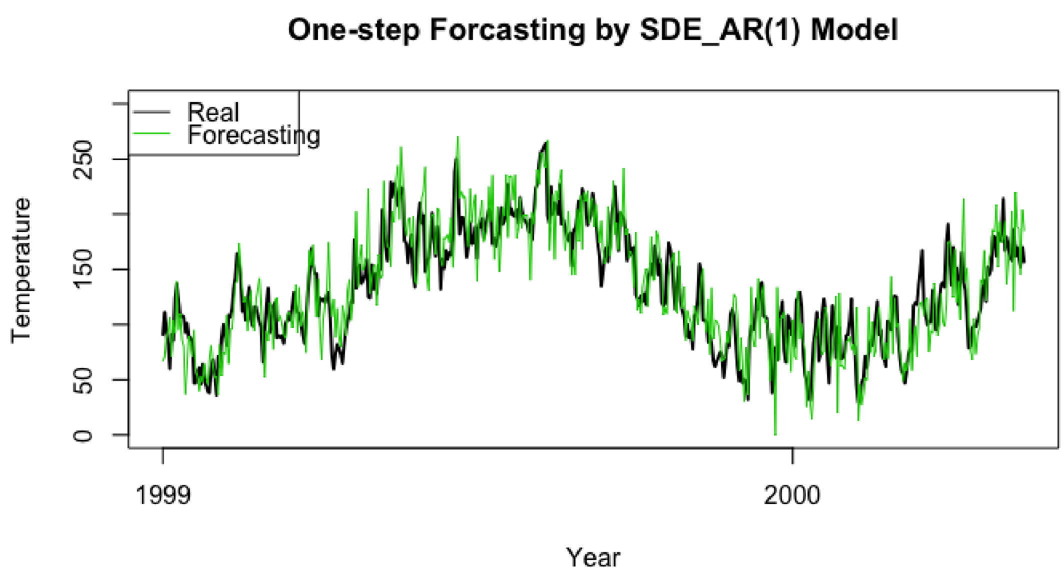


Figure 5.8: One-step forecasting result of SDE-AR(1) model.

We then make forecasting by SDE-AR(1) model (shown in green curve) in Figure 5.8 and compare with the observed data plotted in black curve from time 43801 to 44300. We use the following formula to calculate the relative error between the real data vector $\boldsymbol{T}(t)$ and the forecasting vector $\hat{\boldsymbol{T}}(t)$:

$$err := \frac{< |\boldsymbol{T}(t) - \hat{\boldsymbol{T}}(t)| >}{< |\hat{\boldsymbol{T}}(t)| >}. \tag{5.4.5}$$

Then the relative error for this model is 3.5994%.

Campbell *et al.* [14] studied the autoregression model with residuals described by Fourier series. In some other cases in the early research people gave several ARMA-based models with residuals described by more complicated models, such as ARCH series or Lévy process. But further research showed that these complicated residuals model will add complexity for further price modelling for weather derivatives based on such model. In this thesis we only compare the ARMA model for random periodic processes with the classical ARMA model. For specific applications, one can also consider adding more complicated term in the ARMA model for random periodic processes if necessary.

Now we start to estimate the ARMA model for random periodic processes for this sample. In order to use random periodic process to describe the data, we first eliminate the non-periodic trend component by using the least square estimation (1.1.1).

```
> lm1

Call:
lm(formula = data_daily ~ t)

Coefficients:
(Intercept)            t
  1.241e+02    2.915e-04
```

Figure 5.9: LSE of first order.

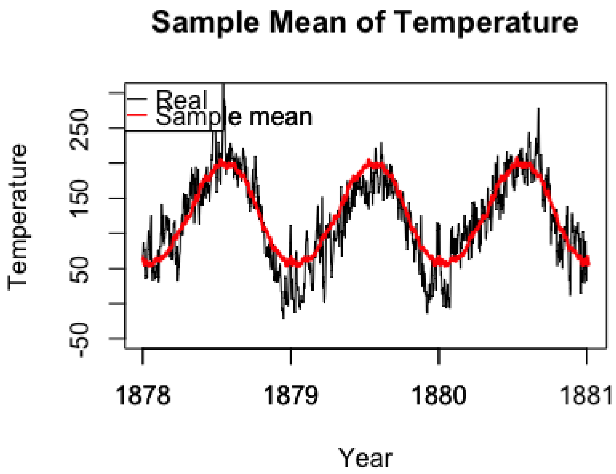
```
> lm2

Call:
lm(formula = data_daily ~ t + I(t^2))

Coefficients:
(Intercept)            t       I(t^2)
  1.254e+02    1.450e-04    2.866e-09
```

Figure 5.10: LSE of second order.

We ignore the second-order coefficient, but the first-order is crucial to the trend of process, which is consistent with the result of the estimated mean values of temperature. After eliminating the trend component, we calculate the sample mean and sample auto-covariance at different time points in one period to estimate the value of w. One sample of the results are shown below.



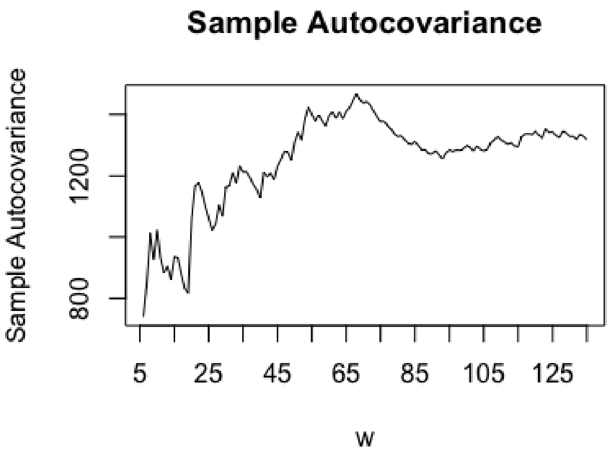


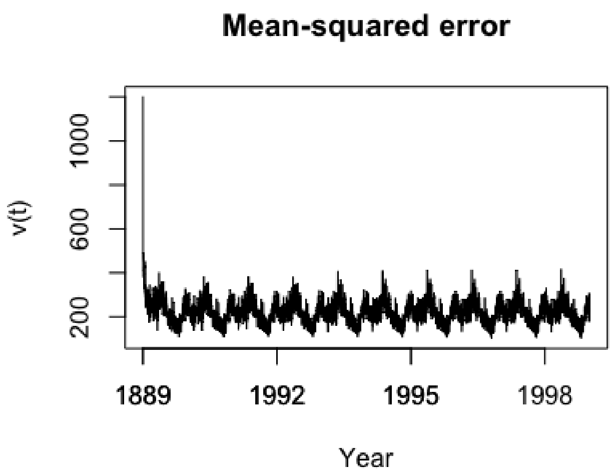
Figure 5.11: Sample mean of temperature. Figure 5.12: Sample autocorrelation w.r.t. w.

Figure 5.12 shows the plot of $\hat{\gamma}(50001, 50001)$ with respect to N. We observe that he sample mean has a cosine pattern but with some small fluctuation and little difference between different cycles. We also observe that after around 100 the autocovariance functions tend to converge. Hence we set $w = 110$ for this case. Since the value of period is far larger than that of w, and the total number of cycles in the sample is less than the period, we observe that the truncated innovation algorithm is hard to achieve. For this special case, we use the original innovation algorithm instead. And in order to avoid the singularity of the sample covariance matrix, we set

$$\hat{\gamma} := \frac{1}{w+t-s} \sum_{i=0}^{w-1} Y(t-iw)Y(s-iw), \quad t \geq s,$$

which makes the large-distant sample autocovariances tend to zero and guarantees the non-singularity of covariance matrix. We then use sample from 40150 to 43800 in the innovation algorithm. The plot of mean-squared error v_t is shown in Figure 5.13.

We observe that $\hat{v}_t$ shows similar pattern with the estimated variance by squared residuals in SDE-AR(1) model. The average of $\hat{v}_t$ described in red curve in Figure 5.14 is smaller than the estimated variance $\tilde{\sigma}^2(t)$ in (5.4.4) in green one, which may be resulted by the periodic coefficients of autoregression part which matches more suitable to the sample than the constant coefficients.



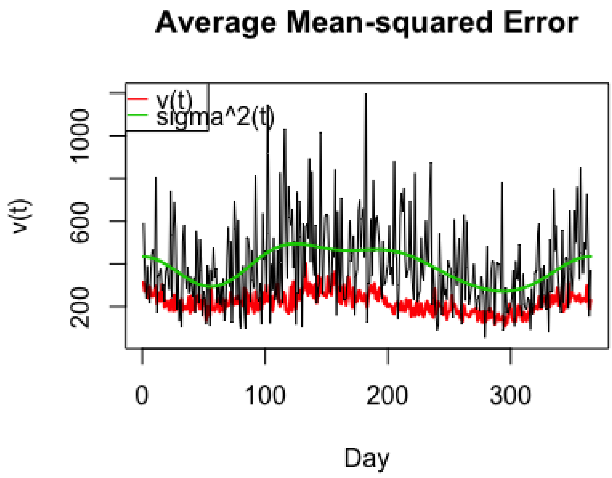
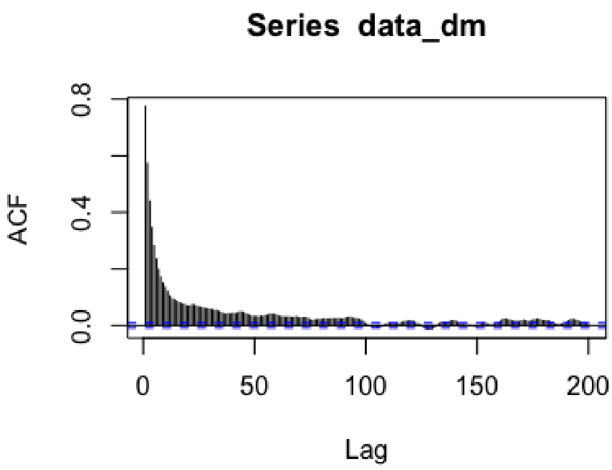


Figure 5.13: Mean-squared error of innovation algorithm.

Figure 5.14: Average mean-squared error.



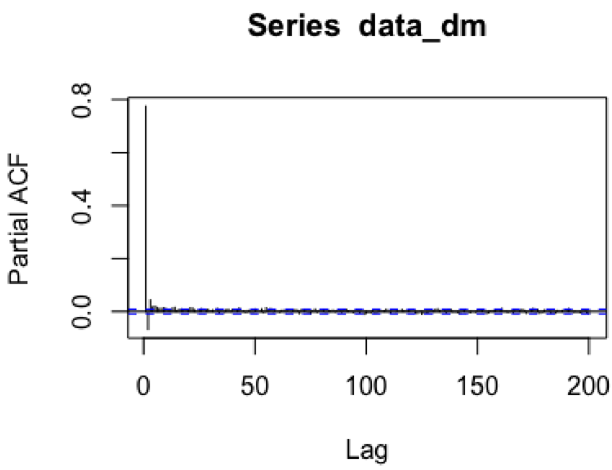


Figure 5.15: Autocorrelation function.

Figure 5.16: Partial autocorrelation function.

Specifically, for an AR(1) process, the sample autocorrelation function should have an exponentially decreasing appearance. However, higher-order autoregressive processes are often a mixture of exponentially decreasing and damped sinusoidal components.

According to the autocorrelation function and the partial autocorrelation function for the de-trended process with mean eliminated in Figure 5.15 and 5.16, we search AR(p) model for random periodic processes to fit the sample. By the performance of the autocorrelation functions for residuals and squared residuals, we choose AR(1) model for random periodic processes for this sample.

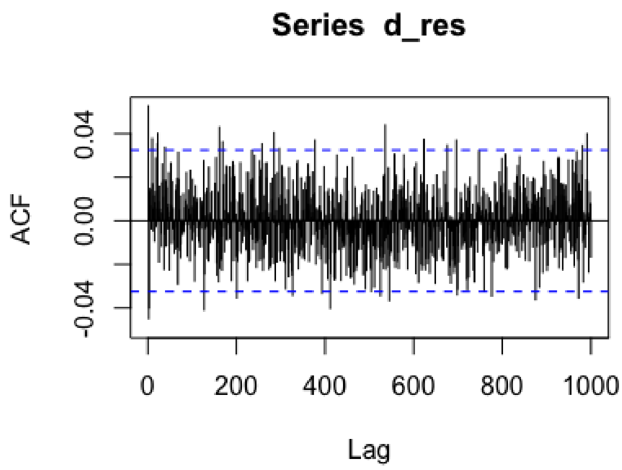


Figure 5.17: Autocorrelation functions for residuals.

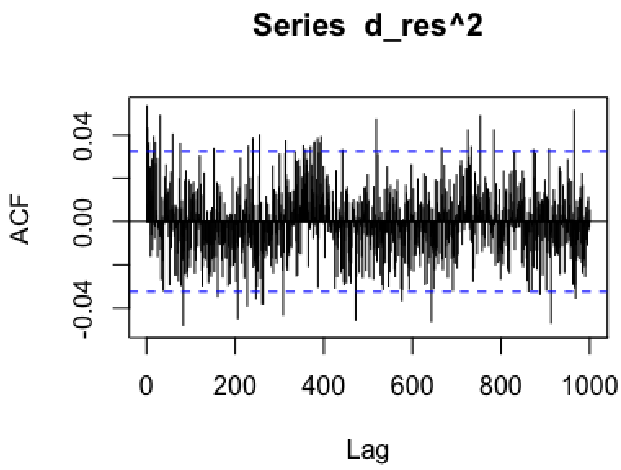


Figure 5.18: Autocorrelation functions for squared residuals.

The slightly periodic pattern of autocorrelation functions for squared residuals reveals the consistent with periodicity of $\hat{v}_t$ and $\hat{\theta}_0(t)$. The autocorrelation function for residuals shows that a higher-order of autoregression may be taken into consideration. But further calculation shows that there is not much improvement of adding more terms in the model. Hence we still choose AR(1) model for random periodic processes for this example:

$$[T(t) - \hat{m}(t)] - \phi_1(t)\,[T(t-1) - \hat{m}(t-1)] = \theta_0(t)Z(t). \tag{5.4.6}$$

The values of $\phi_1(t)$ and $\theta_0(t)$ will not be given in this thesis as the length of the coefficients matrix is too big.

Figure 5.19 shows one sample of the forecasting result of AR(1) model for random periodic processes. The relative error between observations and predictions is 2.7425%, which is a bit smaller than that of SDE-AR(1) model.

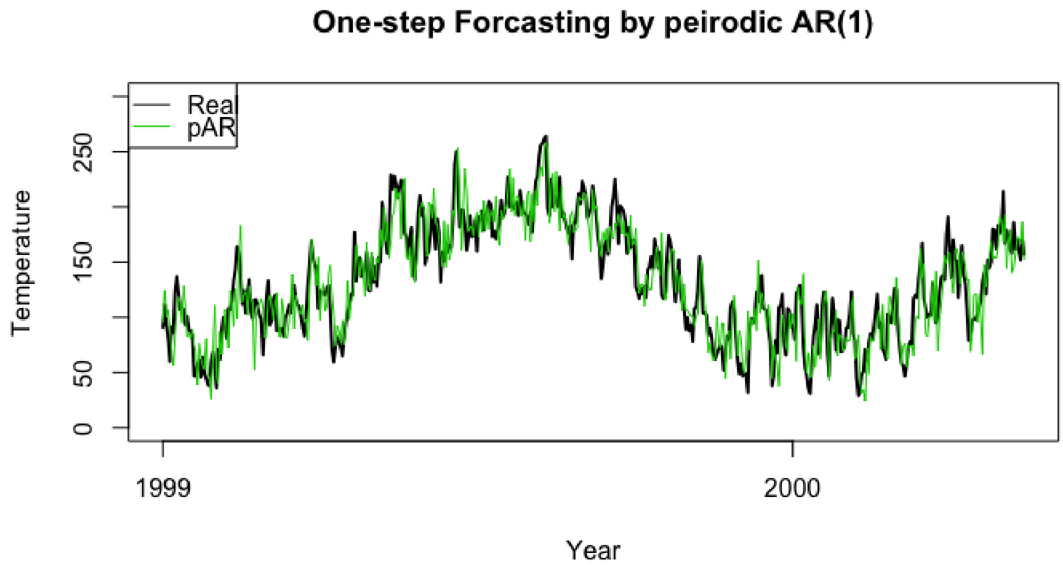


Figure 5.19: One-step forecasting result of AR(1) model for random periodic processes.

For comparisons, we search for suitable ARIMA model for de-trended and de-seasoned data *data_tr* by function *auto.arima* in R, which gives us the result shown in Figure 5.20.

The autocorrelation function of residuals for ARIMA(2,0,2) model in Figure 5.22 is almost within the confidence interval of zero, which is satisfactory. But the autocorrelation function of squared residuals in Figure 5.21 shows small periodic pattern, which implies the variance of noise may depend on time.

Although the autocorrelation function and partial autocorrelation function of residuals give quite satisfactory result, the forecasting given in Figure 5.23 is not more satisfactory, and the relative error is 5.7639%, which is much bigger than that of the AR(1) model for random periodic processes. Shapiro-Wilk normality test rejects the normality hypothesis of residuals with $p - value = 1.791e^{-7}$ in addition to the autocorrelation function of squared residuals. That is to say, the residuals are not " white noise", which implies that there is still some information about temperature in the residuals which ARIMA model fails to figure out.

```
> d_fit1
Series: ts(data_tr[c(40151:43800)])
ARIMA(2,0,2) with zero mean

Coefficients:
         ar1      ar2      ma1      ma2
      1.6047  -0.6214  -0.7876  -0.1111
s.e.  0.0529   0.0448   0.0542   0.0217

sigma^2 estimated as 385.8:  log likelihood=-16045.84
AIC=32101.67   AICc=32101.69   BIC=32132.68
```

Figure 5.20: ARIMA(2,0,2) model.

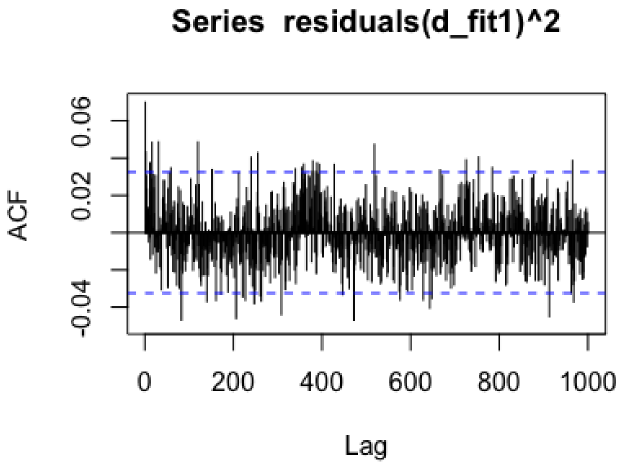


Figure 5.21: Autocorrelation function for squared residuals.

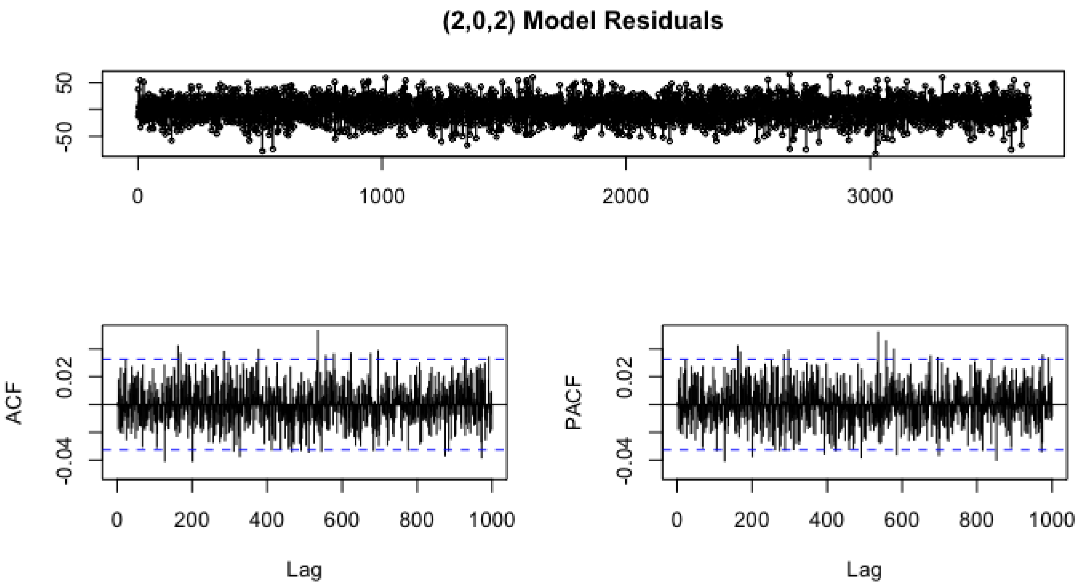


Figure 5.22: Residuals of ARIMA.

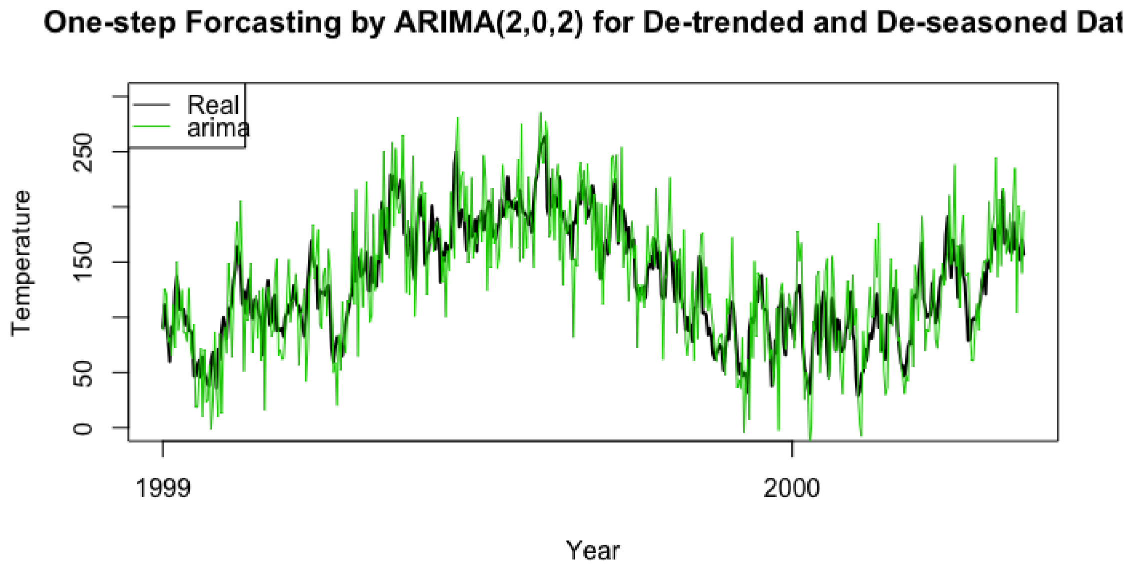


Figure 5.23: Forecasting result of ARIMA(2,0,2) model.

5.5 Example of SDE

We continue with the example of random periodic solution of SDE. From Figure 2.11 we set $w = 650$. Hence the first stage of estimation is from 1 to 13000. The length of the second stage is set to be 4000 after practising several times, i.e. from 13000 to 17000. And we aim to use the last 1000 data to do the prediction.

We observe that there will be extreme points of mean-square error $\hat{v}(t)$ when K is chosen too large, for example in Figure 5.24, we take $w = 5$ and there are several near-sigular points. We find that in application K usually is 1. In Figure 5.25 we plot $\hat{v}(t)$ after applying the truncated innovation algorithm with $K = 1$ and $w = 650$. One can observe that $\hat{v}(t)$ shows periodic pattern clearly. We plot $\hat{v}(t+(i-1)\times 20)$ and $\hat{b}_k(t+(i-1)\times 20)$ for each point t in one periodic, and choose the converged value to determine $v(t)$ and $\psi_k(t)$. One example of time 1 is shown in Figure 5.26 and 5.27.

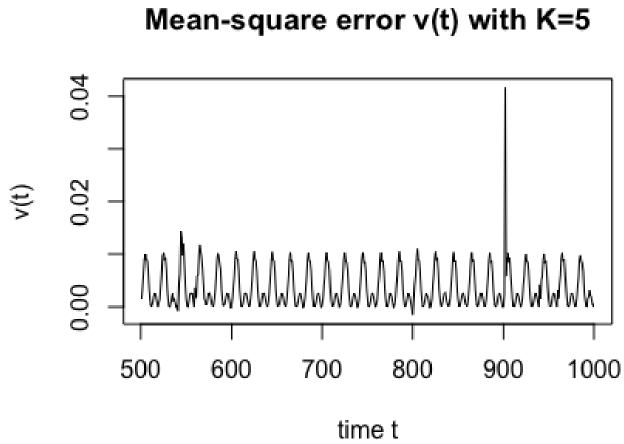


Figure 5.24: $\hat{v}(t)$ with K=5 and w=650.

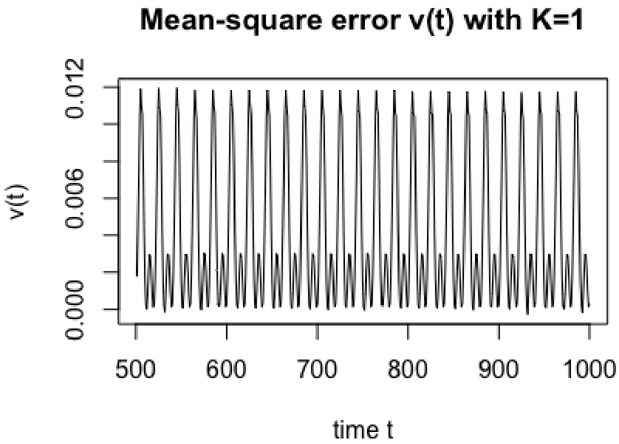


Figure 5.25: $\hat{v}(t)$ with K=1 and w=650.

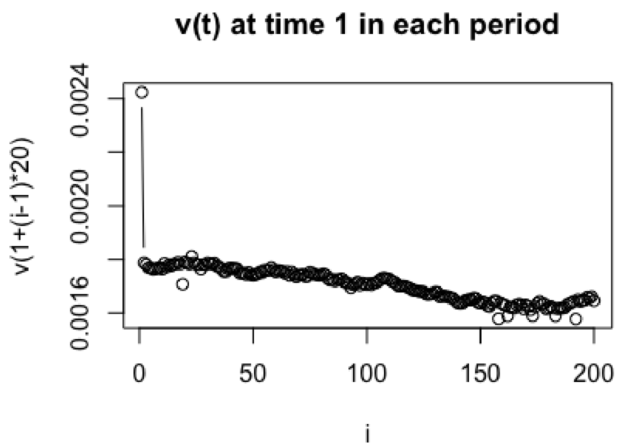


Figure 5.26: $\hat{v}(1 + (i - 1) \times 20)$.

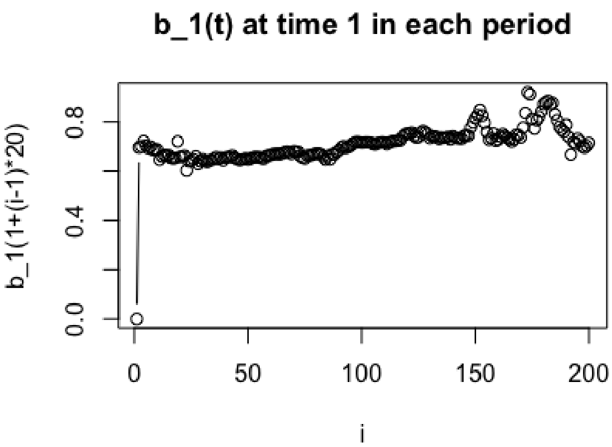


Figure 5.27: $\hat{b}_1(1 + (i - 1) \times 20)$.

For each pair of (p, q) with $p > 0, q > 0, p + q \leq K\tau$, calculate the coefficients of the ARMA(p, q) model for random periodic processes by $\hat{\boldsymbol{\psi}}(t)$ and then calculate the model fit criteria. Also calculate them for the MA(20) model for random periodic processes. We found that the MA(20) model for random periodic processes has the smallest value of model fit criteria. Hence we use it to do model prediction. For $0 < m \leq K\tau$, calculate the coefficients of each corresponding AR(m) model for random periodic processes and use this model to estimate the history noise by history data. To find the proper value for m, we define $\tilde{Y}_m$ as the predicted value of the corresponding AR(m) model for random periodic processes and choose m^* which minimises the following error

$$err(m) := \frac{1}{n}\sum_{t=1}^{n}\Big(Y(t) - \tilde{Y}_m(t)\Big)^2.$$

For this example $m^* = 12$. We found that not the larger m causes the better prediction result. One sample of the forecasting result by the MA(20) model for random periodic processes is given in Figure 5.28. The relative error between the predictions and the observations is 24.18%. The relative error is quite large since the absolute value of sample is too small.

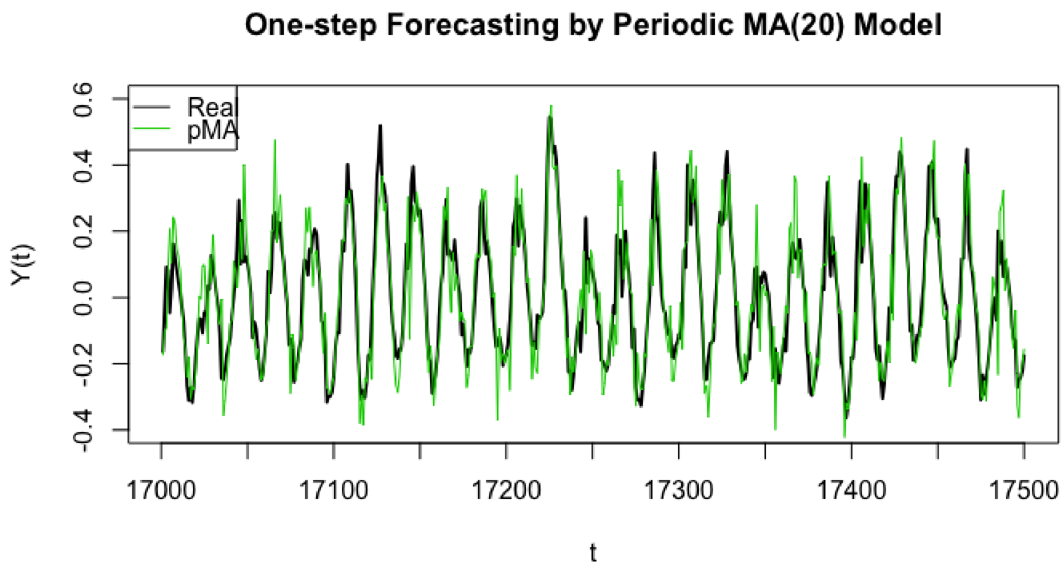


Figure 5.28: Forecasting result by periodic MA(20) model.

In comparison we apply *auto.arima* function in R to simulate the sample by ARIMA model with *season* parameter in the function to be TRUE. The *auto.arima* function searches through combinations of order parameters and picks the set that optimizes model fit criteria AIC, AICc, and BIC. The result is shown in Figure 5.29. In Figure 5.30, the function *tsdisplay* tests the residuals of the given model and displays the ACF and PACF plots of the residuals. If the model order parameters and the structure are correctly specified, there should be no significant autocorrelations of residuals present.

```
> v_fit_no_holdout
Series: ts(data_v1[c(13001:17000)])
ARIMA(2,0,3) with zero mean

Coefficients:
         ar1      ar2      ma1     ma2     ma3
      1.8614  -0.9550  -1.0306  0.0838  0.1881
s.e.  0.0059   0.0059   0.0184  0.0215  0.0193

sigma^2 estimated as 0.005087:  log likelihood=4887.23
AIC=-9762.46   AICc=-9762.44   BIC=-9724.69
```

Figure 5.29: The results of function *auto.arima* for original data.

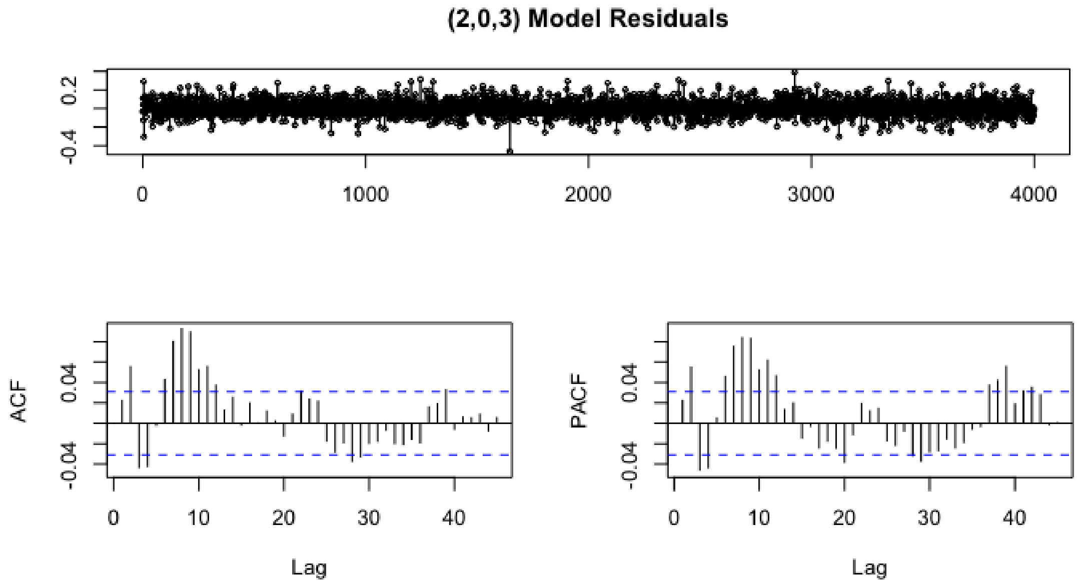


Figure 5.30: Test of residuals by ARIMA(2,0,3).

The function *auto.arima* gives ARIMA(2,0,3) model for this sample. But the autocorrelation function of residuals shows that there are more parameters expected to add in. We forecast next 500 data by ARIMA(2,0,3) model and the result is shown in Figure 5.31. The relative error is 37.40%, which is much bigger than that of the MA(20) model for random periodic processes.

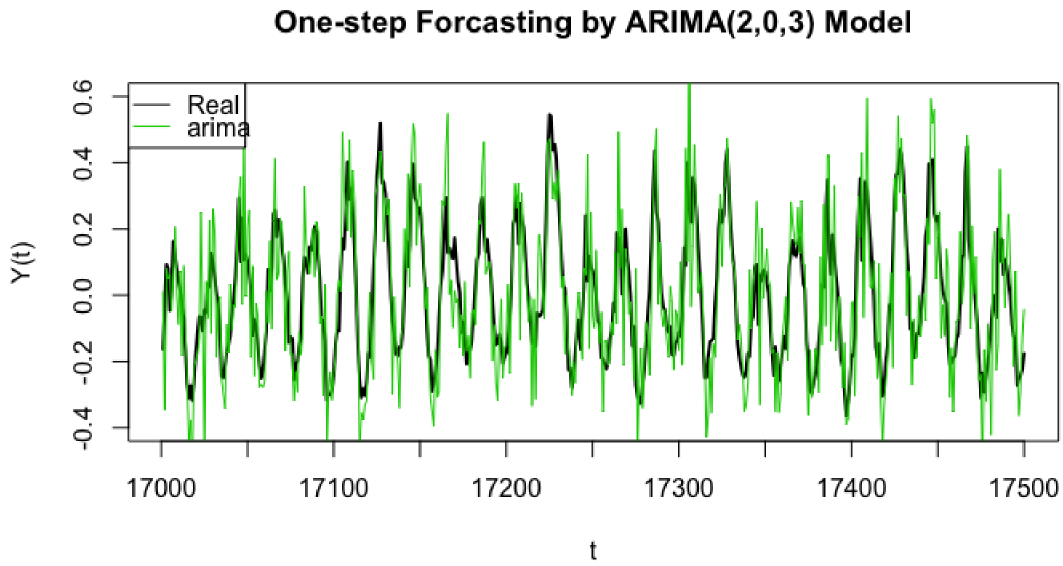


Figure 5.31: Forecasting result by ARIMA(2,0,3) model.

Next we consider SDE-AR model for this example according to the construction of the data set. The solution of this SDE is (integral from zero to t)

$$X(t) = e^{-\pi t} X(0) + \int_0^t e^{-\pi(t-s)} \sin(\pi s) ds + \int_0^t \sigma(s) e^{-\pi(t-s)} dW_t,$$

where $\sigma(t) := 0.1 + 0.3 \sin(\pi t)$. Then

$$X(t+1) = e^{-\pi} X(t) + e^{-\pi} \int_t^{t+1} e^{-\pi(t-s)} \sin(\pi s) ds + e^{-\pi} \int_t^{t+1} \sigma(s) e^{-\pi(t-s)} dW_t.$$

Set $\tilde{X}(t) := X(t) - \left[\frac{1}{2\pi}\left(\sin(\pi t) - \cos(\pi t)\right)\right]$, and by similar approximation of the integral of Brownian motion, we have

$$\tilde{X}(t+1) \approx e^{-\pi} \tilde{X}(t) + e^{-\pi} \sigma(t) \epsilon_t, \tag{5.5.1}$$

where $\epsilon_t \sim N(0,1)$. This approximation of the solution implies us to establish AR(1) model with periodic-variance noise for the sample data as

$$[X(t) - s(t)] - \phi\, [X(t-1) - s(t-1)] = Z(t), \tag{5.5.2}$$

where $Z(t) \sim N(0, \tilde{\sigma}(t))$.

To estimate the mean value of $X(t)$, we do regression to the sample set to fit the following function:

$$\hat{s}(t) = A \sin(\frac{2\pi}{20} t) + B \cos(\frac{2\pi}{20} t) + C.$$

The regression result is shown in Table 5.2. The plots of estimated mean values and de-seasoned values of $\tilde{X}(t) = [X(t) - s(t)]$ are in Figure 5.32 and 5.33.

Table 5.2: The coefficients of $\hat{s}(t)$.

A	B	C
0.158867	- 0.161542	-0.002256

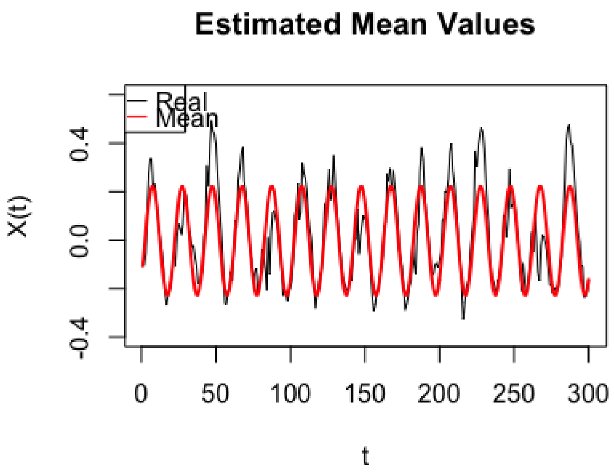


Figure 5.32: Estimated mean values.

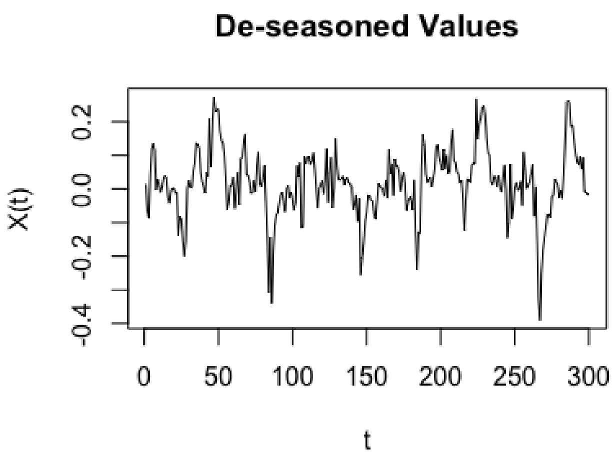


Figure 5.33: De-seasoned values.

```
Call:
arima(x = data_vr[c(13001:17000)], order = c(1, 0, 0))

Coefficients:
         ar1  intercept
      0.7450     0.0058
s.e.  0.0105     0.0041

sigma^2 estimated as 0.004411:  log likelihood = 5170.93,  aic = -10335.86
```

Figure 5.34: The results for SDE-AR(1) model.

Then we use AR(1) model to fit the $\tilde{X}(t)$ ranging from 13001 to 17000. Figure 5.34 shows that the estimated value of ϕ is 0.745, which is close to the real value $e^{-\pi/10}$. The period of the original process is two, but for the sample the period is 20, hence the frequency is $\pi/10$, which is consistent with model (5.5.1) as well.

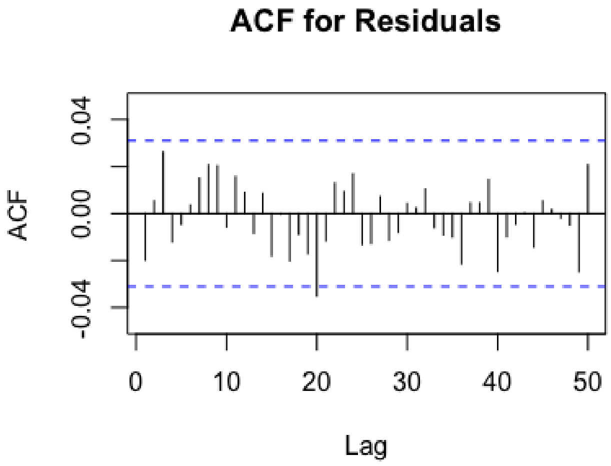


Figure 5.35: ACF for residuals.

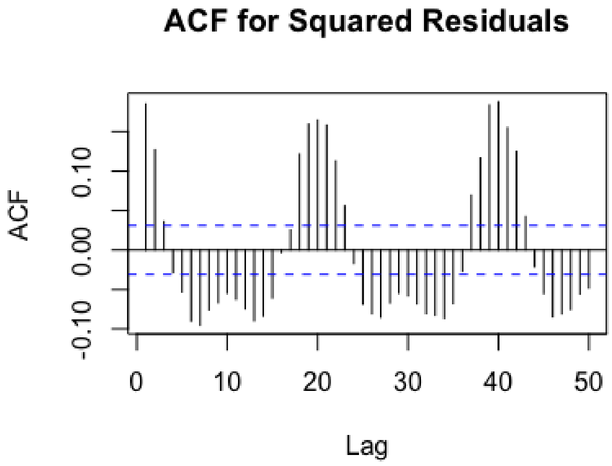


Figure 5.36: ACF for squared residuals.

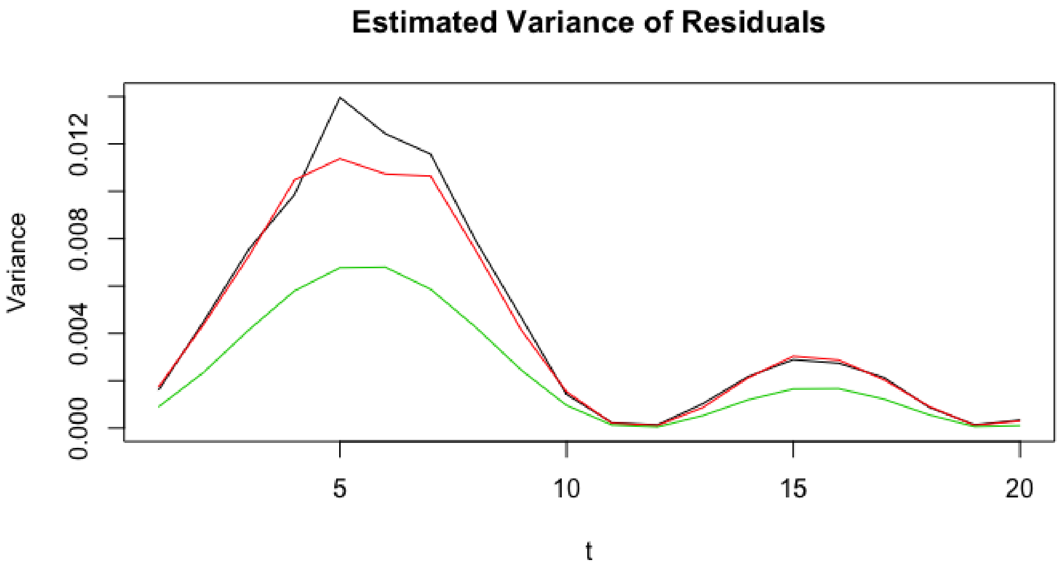


Figure 5.37: The estimated variance of residuals.

The autocorrelation function for residuals in Figure 5.35 shows that the residuals may be regarded as independent with each other, while the autocorrelation function for squared residuals in Figure 5.36 shows that there is significant time dependence in the variance of the residuals.

We use the same method stated in example of temperature and plot the estimated variance in Figure 5.37 with comparison of the mean-squared error $\hat{v}_t$ (red curve) calculated by truncated innovation algorithm and $e^{-\pi}\left(0.1 + 0.3\sin(-\frac{2\pi}{20}t + \frac{\pi}{0.95})\right)$ (green curve) in original solution of this SDE. A phase angle $\frac{\pi}{0.95}$ is introduced here since the sample is taken discretely and there may be some offset with respect to the original process. We observe that the patterns are similar. Since we approximated the integral part of the solution, the green curve may not be considered as the exact standard of the variance, but a good contrast.

Then we use SDE-AR(1) model to forecast next 500 data (represented by green curve) and compare with the observed values (represented by black curve) from 17001 to 17500 in Figure 5.38. The relative error between the forecasting and the observed values is 22.07%, which is smaller than the previous two models.

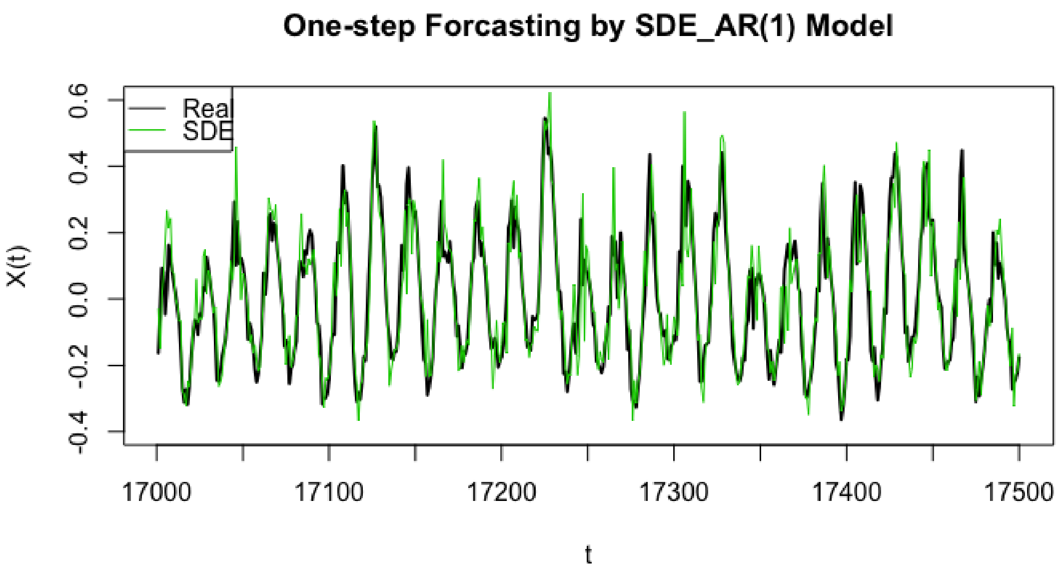


Figure 5.38: Forecasting result by SDR-AR(1) model.

Inspired by the form of the solution, we consider the AR(1) model for random periodic processes as follows,

$$[X(t) - \hat{m}(t)] - \phi_1(t)\,[X(t-1) - \hat{m}(t-1)] = \theta_0(t)Z(t), \tag{5.5.3}$$

where $\hat{m}(t)$ is the sample mean by (2.3.1) and $Z(t) \sim N(0,1)$. The values of coefficients $\hat{\phi}_1(t)$ is given in Table 5.3. We observed that these values are around the real value $e^{-\pi/10}$, and the volatility of these values is quite large in Figure 5.39. One sample of the forecasting result of the AR(1) model for random periodic processes is shown in Figure 5.40. The relative error is 25.36%, which is a little bigger than that of SDE-AR(1) model. It may be better to set $\hat{\phi}_1(t)$ to be constant. This example inspires us to study the determination criterion for periodicity in real world cases.

Table 5.3: The coefficients of periodic AR(1) model.

$\phi_1(1)$	$\phi_1(2)$	$\phi_1(3)$	$\phi_1(4)$	$\phi_1(5)$	$\phi_1(6)$	$\phi_1(7)$	$\phi_1(8)$	$\phi_1(9)$	$\phi_1(10)$
0.6815	0.6951	0.6731	0.7322	0.7395	0.7435	0.7040	0.6952	0.7160	0.7320
$\phi_1(11)$	$\phi_1(12)$	$\phi_1(13)$	$\phi_1(14)$	$\phi_1(15)$	$\phi_1(16)$	$\phi_1(17)$	$\phi_1(18)$	$\phi_1(19)$	$\phi_1(20)$
0.7166	0.7255	0.7359	0.7577	0.7040	0.7327	0.7237	0.7388	0.7193	0.7059

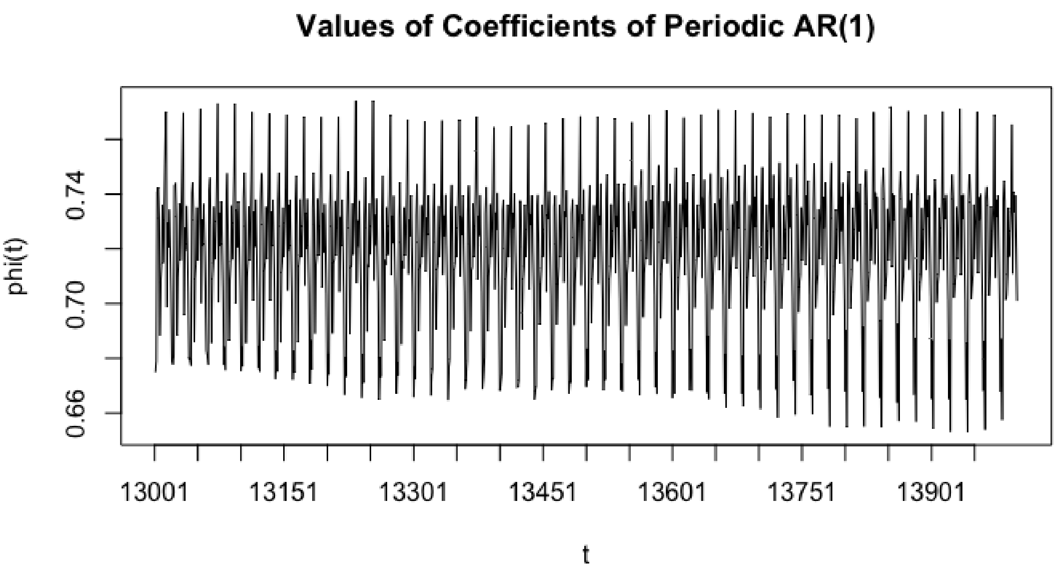


Figure 5.39: Estimated values of $\hat{\phi}_1(t)$.

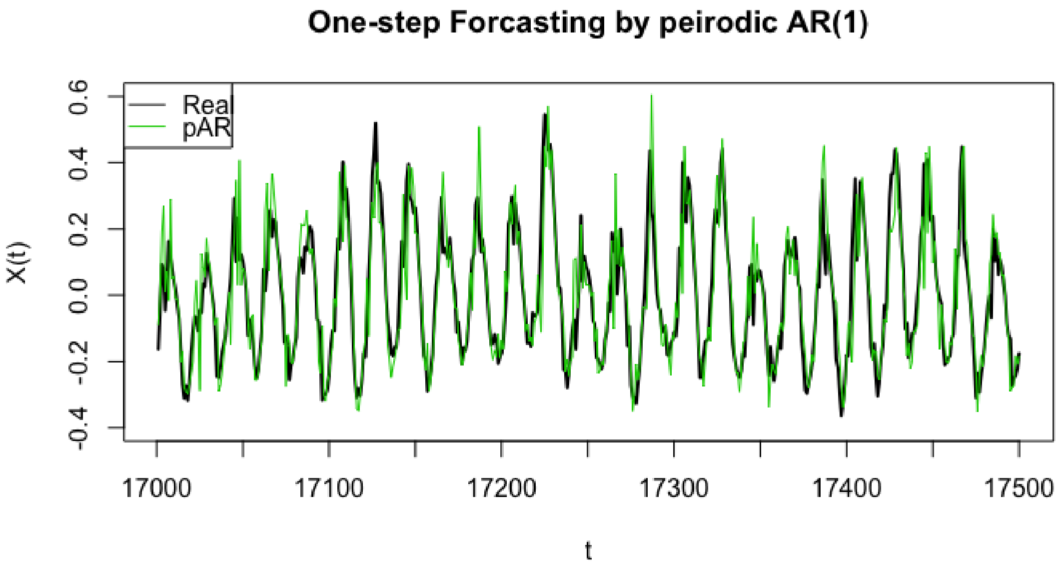


Figure 5.40: Forecasting result by periodic AR(1) model.

Appendix A

R Language Code

```
# Calculating the covariance of data between time t and s with s<=t
rCovariance<-function(data,period,t,s,w){
  scov <- 0
  temp <- s-(w-1)*period
  if(temp<1) stop("There isn't enough backward data")
  for(i in 0:(w-1)){
    scov <- scov+data[t-i*period]*data[s-i*period]
    }
  scov<-temp/(w-1)
  return(scov)
  }
# Truncated Innovation algorithm
rInno<-function(data,period,s,t,w,K){
  N <- t-s
  scov <- array(0,c(N,N))
  for(i in 1:N){
    for(j in 1:i){
      scov[i,j] <- rCovariance(data,period,s+i,s+j,w)
      }
  }
  dtheta <- array(0,c(N,N))
  v <- numeric(N)
```

```
v[1] <- scov[1,1]
dtheta[1,1] <- scov[2,1]/v[1]
v[2] <- scov[2,2] - dtheta[1,1]^2*v[1]
for(n in 2:(K*period)){
  dtheta[n+1,n] <- scov[n+1,1]/v[1]
  for(k in 1:(n-1)){
    temp <- 0
    for(j in 0:(k-1)){
      temp <- temp + dtheta[k+1,k-j]*dtheta[n+1,n-j]*v[j+1]
    }
    dtheta[n+1,n-k] <- (scov[n+1,k+1]-temp)/v[k+1]
  }
  temp <- 0
  for(j in 0:(n-1)){
    temp <- temp + dtheta[n+1,n-j]^2*v[j+1]
  }
  v[n+1] <- scov[n+1,n+1] - temp
  if(v[n+1]==0){
    cat("v(",s+n+1,") is zero.")
    break
  }
}
for(n in (K*period+1):(N-1)){
  for(k in (n-K*period):(n-1)){
    temp <- 0
    if(k <= (K*period)){
      for(j in 0:(k-1)){
        temp <- temp + dtheta[k+1,k-j]*dtheta[n+1,n-j]*v[j+1]
      }
      dtheta[n+1,n-k] <- (scov[n+1,k+1]-temp)/v[k+1]
    }
    else{
      for(j in (k-K*period):(k-1)){
```

```
          temp <- temp + dtheta[k+1,k-j]*dtheta[n+1,n-j]*v[j+1]
        }
        dtheta[n+1,n-k] <- (scov[n+1,k+1]-temp)/v[k+1]
      }
    }
    temp <- 0
    for(j in (n-K*period):(n-1)){
      temp <- temp + dtheta[n+1,n-j]^2*v[j+1]
    }
    v[n+1] <- scov[n+1,n+1] - temp
    if(v[n+1]==0){
      cat("v(",s+n+1,") is zero.")
      break
    }
  }
  result<-list(scov=scov,dtheta=dtheta,v=v)
  return(result)
}
# For any (p,q) such that p+q<=K\tau, calculating phi and theta by psi
# p,q are not zero, and p+q<=K*period
coeff_parma <- function(period,p,q,K,psi,v){
  phi <- array(0,c(period,p))
  theta <- array(0,c(period,q))
  if(p>=(q+1)){
    for(t in 1:period){
      t1 <- t+K*period
      A <- array(0,c(p,p))
      b <- numeric(p)
      if((2*p-1)<=(K*period)){
        for(i in 1:p){
          b[i] <- psi[t,p+i-1]
        }
      }
```

```
else{
  for(i in 1:(K*period+1-p)){
    b[i] <- psi[t,p+i-1]
  }
}
s <- (t1-p)%%period
if(s==0) s <- period
A[1,p] <- v[s]
for(i in 2:p){
  A[i,p] <- psi[s,i-1]
}
for(j in 1:(p-1)){
  s <- (t1-j)%%period
  if(s==0) s <- period
  if((K*period+j-p+1)>=p){
    for(i in 1:p){
      A[i,j] <- psi[s,p-1+i-j]
    }
  }
  else{
    for(i in 1:(K*period+j-p+1)){
      A[i,j] <- psi[s,p-1+i-j]
    }
  }
}
if(det(A)==0){
  cat("The determinant of",t," is zero.")
  break
}
phi[t,] <- solve(A)%*%b
t2 <- t-1
if(t2==0) t2 <- period
theta[t,1] <- psi[t,1] - phi[t,1]*v[t2]
```

```
      if(q>1){
        for(j in 2:q){
          temp <- psi[t,j]
          for(i in 1:(j-1)){
            s <- (t1-i)%%period
            if(s==0) s <- period
            temp <- temp - phi[t,i]*psi[s,j-i]
          }
          s <- (t1-j)%%period
          if(s==0) s <- period
          theta[t,j] <- temp - phi[t,j]*v[s]
        }
      }
    }
  }
  else if(p==1){
    for(t in 1:period){
      t1 <- t-1
      if(t1==0) t1 <- period
      phi[t,1] <- psi[t,q+1]/psi[t1,q]
      theta[t,1] <-  psi[t,1] - phi[t,1]*v[t1]
      if(q>1){
        # theta[t,1] <- psi[t,1] - phi[t,1]*v[t1]
        for(j in 2:q){
          theta[t,j] <- psi[t,j] - phi[t,1]*psi[t1,j-1]
        }
      }
    }
  }
  else{
    for(t in 1:period){
      t1 <- t+K*period
      A <- array(0,c(p,p))
```

```
b <- numeric(p)
for(i in 1:p){
  b[i] <- psi[t,q+i]
}
for(j in 1:p){
  s <- (t1-j)%%period
  if(s==0) s<-period
  for(i in 1:p){
    A[i,j] <- psi[s,q+i-j]
  }
}
if(det(A)==0){
  cat("The determinant of",t," is zero.")
  break
}
phi[t,] <- solve(A)%*%b
t2 <- t-1
if(t2==0) t2 <- period
theta[t,1] <-  psi[t,1] - phi[t,1]*v[t2]
for(j in 2:p){
  temp <- psi[t,j]
  for(i in 1:(j-1)){
    s <- (t1-i)%%period
    if(s==0) s <- period
    temp <- temp - phi[t,i]*psi[s,j-i]
  }
  s <- (t1-j)%%period
  if(s==0) s <- period
  theta[t,j] <- temp - phi[t,j]*v[s]
}
if(q>p){
  for(j in (p+1):q){
    temp <- psi[t,j]
```

```
          for(i in 1:p){
            s <- (t1-i)%%period
            if(s==0) s <- period
            temp <- temp - phi[t,i]*psi[s,j-i]
          }
          theta[t,j] <- temp
        }
      }
    }
  }
  result<-list(phi=phi,theta=theta)
  return(result)
}
# Innovation algorithm for Wt with p,q not zero.
rInno_W <- function(data,period,s,t,scov,p,q,phi,theta,v){
  m <- max(p,q)
  N <- t-s
  scov_W <- array(0,c(N,N))
  # Calculate the covariance of Wt
  for(i in 1:m){
    I <- (s+i)%%period
    if(I==0) I <- period
    for(j in 1:i){
      J <- (s+j)%%period
      if(J==0) J <- period
      scov_W[i,j] <- scov[i,j]/(v[I]*v[J])
    }
  }
  for(i in (m+1):N){
    I <- (s+i)%%period
    if(I==0) I <- period
    for(j in 1:m){
      J <- (s+j)%%period
```

```
      if(J==0) J <- period
      temp <- scov[i,j]
      for(k in 1:p){
        temp <- temp - phi[I,k]*scov[i-k,j]
      }
      scov_W[i,j] <- temp/(v[I]*v[J])
    }
    for(j in (m+1):i){
      if(j==i){
        temp <- v[I]^2
        for(k in 1:q){
          temp <- temp + theta[I,k]^2
        }
        scov_W[i,i] <- temp/(v[I]^2)
      }
      else if((i-j)<q){
        J <- (s+j)%%period
        if(J==0) J <- period
        temp <- theta[I,i-j]*v[J]
        for(k in 1:(j-i+q)){
          temp <- temp + theta[I,i-j+k]*theta[J,k]
        }
        scov_W[i,j] <- temp/(v[I]*v[J])
      }
      else if((i-j)==q){
        scov_W[i,j] <- theta[I,q]/v[I]
      }
    }
  }
  for(i in 1:(N-1)){
    for(j in (i+1):N){
      scov_W[i,j] <- scov_W[j,i]
    }
```

```
}
# calculating the coefficients b
dtheta <- array(0,c(N,N))
v <- numeric(N)
v[1] <- scov_W[1,1]
dtheta[2,1] <- scov_W[2,1]/v[1]
v[2] <- scov_W[2,2] - dtheta[2,1]^2*v[1]
if(m>=2){
  for(n in 2:m){
    dtheta[n+1,n] <- scov_W[n+1,1]/v[1]
    for(k in 1:(n-1)){
      temp <- 0
      for(j in 0:(k-1)){
        temp <- temp + dtheta[k+1,k-j]*dtheta[n+1,n-j]*v[j+1]
      }
      dtheta[n+1,n-k] <- (scov_W[n+1,k+1]-temp)/v[k+1]
    }
    temp <- 0
    for(k in 0:(n-1)){
      temp <- temp + dtheta[n+1,n-k]^2*v[k+1]
    }
    v[n+1] <- scov_W[n+1,n+1] - temp
    if(v[n+1]==0){
      cat("v(",s+n+1,") is zero.")
      break
    }
  }
  for(n in (m+1):(N-1)){
    for(k in (n-q):(n-1)){
      temp <- 0
      if(k<=q){
        for(j in 0:(k-1)){
          temp <- temp + dtheta[k+1,k-j]*dtheta[n+1,n-j]*v[j+1]
```

```
          }
        }
        else{
          for(j in (k-q):(k-1)){
            temp <- temp + dtheta[k+1,k-j]*dtheta[n+1,n-j]*v[j+1]
          }
        }
        dtheta[n+1,n-k] <- (scov_W[n+1,k+1]-temp)/v[k+1]
      }
      temp <- 0
      for(k in (n-q):(n-1)){
        temp <- temp + dtheta[n+1,n-k]^2*v[k+1]
      }
      v[n+1] <- scov_W[n+1,n+1] - temp
      if(v[n+1]==0){
        cat("v(",s+n+1,") is zero.")
        break
      }
    }
  }
  else{
    for(n in 2:(N-1)){
      for(k in (n-q):(n-1)){
        temp <- 0
        if(k<=q){
          for(j in 0:(k-1)){
            temp <- temp + dtheta[k+1,k-j]*dtheta[n+1,n-j]*v[j+1]
          }
        }
        else{
          for(j in (k-q):(k-1)){
            temp <- temp + dtheta[k+1,k-j]*dtheta[n+1,n-j]*v[j+1]
          }
```

```
        }
        dtheta[n+1,n-k] <- (scov_W[n+1,k+1]-temp)/v[k+1]
      }
      temp <- 0
      for(k in (n-q):(n-1)){
        temp <- temp + dtheta[n+1,n-k]^2*v[k+1]
      }
      v[n+1] <- scov_W[n+1,n+1] - temp
      if(v[n]==0){
        cat("v(",s+n+1,") is zero.")
        break
      }
    }
  }
  result<-list(scov_W=scov_W, dtheta=dtheta,v=v)
  return(result)
}
# Function of calculating \hat{Y} from s+1 to t.
Y_predictor <- function(data, period, s, t, p, q, phi, theta, v, b){
  m <- max(p,q)
  N <- t-s
  hat_Y <- numeric(N)
  if(m>=2){
    for(n in 1:(m-1)){
      S <- (s+n+1)%%12
      if(S==0) S <- 12
      for(j in 1:n){
        hat_Y[n+1] <- hat_Y[n+1] + b[S,j]*(data[n+1-j] - hat_Y[n+1-j])
      }
    }
    for(n in m:(N-1)){
      S <- (s+n+1)%%12
      if(S==0) S <- 12
```

```
      for(i in 1:p){
        hat_Y[n+1] <- hat_Y[n+1] + phi[S,i]*data[n+1-i]
      }
      for(j in 1:q){
        hat_Y[n+1] <- hat_Y[n+1] + b[S,j]*(data[n+1-j] - hat_Y[n+1-j])
      }
    }
  }
  else{
    for(n in 1:(N-1)){
      S <- (s+n+1)%%12
      if(S==0) S <- 12
      for(i in 1:p){
        hat_Y[n+1] <- hat_Y[n+1] + phi[S,i]*data[n+1-i]
      }
      for(j in 1:q){
        hat_Y[n+1] <- hat_Y[n+1] + b[S,j]*(data[n+1-j] - hat_Y[n+1-j])
      }
    }
  }
  return(hat_Y)
}
# Likelihood function
log_L <- function(data,hat_Y,v,r,s,t){
  n <- t-s
  L <- -n*log(2*pi)/2 - sum(log(v)) - 0.5*sum(r)
  for(i in 1:n){
    L <- L-0.5*(data[s+i]-hat_Y[i])^2/(v[i]^2*r[i])
  }
  return(L)
}
```

References

[1] H. Akaike, A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, vol. 19, issue: 6, pp. 716-723, Dec 1974.

[2] P. Alaton, B. Djehiche, and D. Stillberger, On modelling and pricing weather derivatives, *Applied Mathematical Finance*, vol. 9, no. 1, pp. 1-20, 2002.

[3] P. L. Anderson, M. M. Meerschaert, A. V. Vecchia, Innovations algorithm for periodically stationary time series, *Stochastic Processes and their Applications* 83, pp. 149-169, 1999.

[4] P. L. Anderson, L. Kavalieris, M. M. Meerschaert, Innovations algorithm asymptotics for periodically stationary time series with heavy tails, *Multivariate Analysis* 99, pp. 94-116, 2008.

[5] C. F. Ansley, An Algorithm for the Exact Likelihood of a Mixed Autoregressive Moving Average Process, *Biometrika*, Vol. 66, No. 1, pp. 59-65, 1979.

[6] F. E. Benth and J. Šaltytė-Benth, Stochastic modelling of temperature variations with a view towards weather derivatives, *Applied Mathematical Finance*, vol. 12, no. 1, pp. 53-85, 2005.

[7] F. E. Benth and J. Šaltytė-Benth, The volatility of temperature and pricing of weather derivatives, *Quantitative Finance*, vol. 7, no. 5, pp. 553-561, 2007.

[8] T. Bollerslev, Generalized autoregressive conditional heteroskedasticity, *Econometrics* 31, pp. 307-327, 1986.

[9] G. E. P. Box and G. M. Jenkins, Time Series Analysis: Forecasting and Control. *Journal of the Operational Research Society*, 1971.

[10] F. J. Breidt, R.A. Davis, K.S. Lii, and M. Rosenblatt. Maximum likelihood estimation for non-causal autoregressive processes. *Journal of Multivariate Analysis* 36, 175-198, 1991.

[11] Peter H. Brockwell, Richard A. Davis, Simple consistent estimation of the coefficients of a linear filter, *Stochastic Processes and their Applications*, pp. 47-59, 1988

[12] Peter H. Brockwell, Richard A. Davis, Time Series: Theory and Methods (second edition), *Springer-Verlag*, 1991.

[13] D. C. Brody, J. Syroka, and M. Zervos, Dynamical pricing of weather derivatives, *Quantitative Finance*, 3, pp. 189-198, 2002.

[14] S. D. Campbell and F. X. Diebold, Weather forecasting for weather derivatives, *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 6-16, 2005.

[15] F. Dornier, and M. Querel, Caution to the wind, *Energy Power Risk Management*, Weather risk special report, August, pp. 30-32, 2000.

[16] R. Durrett, *Probability: Theory and Examples - 3rd edition*, Duxbury Press, 2004.

[17] R. F. Engle, Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, Vol. 50, No. 4, 1982

[18] FENG, C., LIU, Y. and ZHAO, H. Numerical approximation of random periodic solutions of stochastic differential equations. *Zeitschrift für Angewandte Mathematik und Physik*, 68 (5), 119, 2017.

[19] C. Feng, H. Zhao, B. Zhou, Pathwise random periodic solutions of stochastic differential equations, *Journal of Differential Equations*, 251, pp.119-149,2011.

[20] C. Feng, H. Zhao, Random Periodic Processes, Periodic Measures and Ergodicity, *arXiv:1408.1897*, 2015.

[21] P. H. Franses, Periodicity and Stochastic Trends in Economic Time Series, *Oxford University Press*, 1999.

[22] E. G. Gladyshev, Periodically and almost-periodically correlated random processes with continuous time parameter, *Theory of Probability and Its Applications*, 8, pp. 173-177, 1963.

[23] M. Lanne, P. Saikkonen *Journal of Time Series Econometrics*, Vol. 3, Issue 3, 1941-1928, 2011.

[24] G. Schwarz, Estimating the Dimension of a Model, *The Annals of Statistics*, Vol. 6, No. 2, 461-464, 1978.

[25] S. J. Taylor, Modeling financial times series, *The Economic Journal* 97, 1987.

[26] A. V. Vecchia, Maximum likelihood estimation for periodic autoregressive moving average models, *Technometrics*, Vol. 27, No. 4, 1985.

[27] P. Whittle, P, Hypothesis testing in time series analysis. Thesis, Uppsala University, Almqvist and Wiksell, Uppsala, 1951.

[28] H. Zhao, Z. Zheng, Random periodic solutions of random dynamical systems, *Journal of Differential Equations*, Vol. 246, pp.2020-2038, 2009.