Limit Theorems and Applications of Time Series with Varying Coefficients

Limit Theorems and Applications of Time Series with Varying Coefficients

By Xiaochang Wang,

A Thesis Submitted to the School of Graduate Studies in the Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

McMaster University December 13, 2022

McMaster University
Doctor of Philosophy (2022)
Hamilton, Ontario (Department of Mathematics and Statistics)

TITLE: Limit Theorems and Applications of Time Series with Varying Coefficients
AUTHOR: Xiaochang Wang M.Sc., (University of Manchester)
SUPERVISORS: Dr. Shui Feng and Dr. Yiping Guo
NUMBER OF PAGES: ix, 115

Abstract

In this thesis, we study the asymptotic behaviour and applications of a class of time series with varying coefficients. More specifically, we establish the large deviation principles and the moderate deviation principles for the first-order autoregressive models. The ADF and KPSS, two different methods, are then used to test the stationarity and conintegration of the annual average temperature series for the hemispheric, continental, and individual cities. We confirm that these series adhere to varying coefficient model. Finally, we examine some regions' extreme rainfall series and show that they should follow the ARCH model.

Keywords: Central Limit Theorem, Large Deviation Principle, ADF, KPSS, Varying Coefficient Autoregressive Model, Cointegration

Acknowledgements

This thesis took a long time to finish. It could not have happened without the help of Shui Feng and Yiping Guo. I would like to express my gratitude to them in particular for agreeing to supervise me while I wrote the papers for this thesis, as well as for their prompt and thoughtful responses to all of my questions. Thank you very much, Tom Hurd and Anas Abdallah, for being on my committee and taking the time to meet and discuss. Thank you to all of my family members and relatives, especially my wife Yuting Tong, who have always been supportive.

Contents

List of Figures

List of Tables

Declaration of Authorship

No portion of the work referred to in the dissertation has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Chapter 1

Introduction

Time series analysis is a statistical method for processing dynamic data based on stochastic process theory and mathematical statistics. It has become a standard statistical method in many industries and is widely used in many fields of social and natural sciences such as economics, finance, meteorology, and astronomy. One of the main topics of our discussion as a time series model is its stationarity. When performing regression analysis on non-stationary time series, we will reach some incorrect conclusions, which is known as the "Spurious Regression" problem. This means that the two series are linked but not causally. Currently, the most common method for testing time series stationarity is to estimate the time series coefficients.

Because of its simple structure, an autoregressive time series, particularly of order one which is denoted by AR(1), has received a lot of attention recently. In the case of an AR(1) process with a fixed coefficient, the model is as follows:

$$Y_t = \phi Y_{t-1} + \varepsilon_t, t = 1, 2, ..., T$$

where $\phi \in \mathbb{R}$ is unknown. $(Y_t)_{0\leq t\leq T}$ is observed and $(\varepsilon_t)_{0\leq t\leq T}$ is a sequence of centred, independent and identically distributed random variables valued in $\mathbb{R}$. The value of ϕ determines the stationarity of this process:

1. If $|\phi| < 1$, the process $\{Y_t\}$ is a stationary process;

2. If $|\phi| \geq 1$, the process $\{Y_t\}$ is a non-stationary process.

More specifically, if $\phi = 1$, the process is called the Unit Root process.

There is one estimator of ϕ which is widely used in times series analysis. It is called the least squares estimator:

$$\hat{\phi} = \frac{\sum_{t=1}^{T} Y_t Y_{t-1}}{\sum_{t=1}^{T} Y_{t-1}^2}$$

Over the past three decades, there have been many studies on the asymptotic behaviour of $\hat{\phi}$ when the sample size becomes large. If the series is a stationary process, the asymptotic distribution of $\hat{\phi}$ should follow the central limit theorem. But if the series is not stationary, the asymptotic distribution of $\hat{\phi}$ will be different. Dickey and Fuller[9] have been at the forefront of this research. In 1979, they firstly provided the asymptotic distribution result for least squares estimator $\hat{\phi}$ when the value of ϕ is equal to one. Because the parameters have different asymptotic distributions in different cases, we can use the hypothesis testing method to determine whether the time series parameters meet the above conditions using the p-value to confirm whether the series is stationary or not. This method is also known as the Dickey-Fuller method. Phillips[20] applied this method to the case in 1986, allowing for weakly dependent and heterogeneously distributed innovations. Concurrently, many researchers have provided the large deviation theory of the least squares estimator for this model, which goes a step further in helping us understand the estimator's asymptotic behaviour when it is close to the actual value. Bercu[1, 3] established the large deviation principle in the AR(1) with stationary and non-stationary cases first. Miao and Shen[29] established the moderate deviation principle for the least squares estimator in the stationary AR(1) process in 2009. Large and moderate deviation results provide refined information on the central limit theorem and the law of large numbers.

Recent studies have considered models where the regression coefficients depend on the sample size. In other words, one would update the regression coefficients as more observations become available. The first known model, the varying coefficient model (VCM) was introduced by Hastie and Tibshirani[13] to allow the regression coefficients to vary systematically and smoothly in more than one dimension. VCMs are useful tools in applied work in economics as they can be used to model parameter heterogeneity in a general way. The varying coefficient model allows for a significant impact on the dynamic structure of the data and some flexibility in the volatility features, which are not available in a fixed coefficient model. For example, Hong and Lee[14] used a varying coefficient model to forecast nonlinearity in the conditional mean of exchange rate changes, allowing the autoregressive coefficients to vary with investment positions. Li and Chen[28] also built a varying coefficient functional autoregressive model for US Treasury bonds, which characterises non-constant dependence between functional predictors and functional responses using a time-varying operator. Because of the unique structure of the varying coefficient model, the time series' stationarity may change as the coefficient changes. Park[19] discovered that some econometric models with time series have weak unit roots, which are roots that are close to unity. Consider the following AR(1) process:

$$Y_t(n) = \phi_n Y_{t-1}(n) + \varepsilon_t, \quad t = 1, 2, ..., n \tag{1.1}$$

where $\phi_n = 1 - \frac{\gamma}{a_n}$, γ is a positive constant and $\{a_n\}$ is a sequence of constants which increases to infinity. When the number of observations is small, the coefficient in this model is strictly less than one, and the series is stationary. However, the nature of this series will change as the number of data points increases. If we simply judge whether the time series is stationary based on the estimator's asymptotic distribution, the result may be incorrect.

In recent years, there has been a lot of activity in the research on the relationships between temperatures/extreme rainfalls and global warming. One obvious question is whether the varying coefficient models can be used in this context. As we all know, the greenhouse effect can cause significant changes in temperature and rainfall. Stern and Kaufmann[23] demonstrated that the unit root should be present in global and hemispheric annual temperature time series from 1850 to 2000. However, if we only test these time series with data from 1850 to 1900, we will find that they do not contain unit root. As a result, we believe that the coefficient of these time series will change over time and that it should be formulated through a varying coefficient model. Therefore, understanding the asymptotic behaviour of the estimator will significantly aid our future tests of such models. Chan and Wei[4] provided the asymptotic distribution of the least squares estimator of the varying coefficient model (1.1) with the form of ϕ_T as $\phi_T = 1 - \frac{1}{a_T}$.

The contributions of this thesis are two folds. The first main result established in this thesis is the large and moderate deviation results for least squares estimators with the coefficient ϕ_T in the form of $1 - \frac{1}{f(T)}$. For the large deviation principle of the estimator, it can be shown that the result is the same as the critical case. In contrast, for the moderate deviation principle, we need additional conditions. The rate function is also different from the fixed-parameter model. We gain a more refined understanding of the asymptotic behaviour of the model's estimators as a result of these findings. The application of varying coefficient models in data analysis for temperature and extreme rainfalls is the second main result. It turns out that the annual average temperature series of different regions follow the varying coefficient autoregressive model. Since the extreme rainfalls data is localized, the random noise is dominant and the VCM does not seem to be appropriate. The situation is expected to change if the data is not localized.

In Chapter 2, we introduce some definitions and developments in probability theory, such as the law of large numbers, the central limit theorem, and the large deviation principle. Both theories were developed first for independently and identically distributed variable sequences. However, when certain conditions are met, variables that are not independent or differently distributed still satisfy the central limit theorem and the large deviation principle.

In Chapter 3, we define time series and introduce stationarity. Then, based on the distinction between stationary and non-stationary time series models, we present some special time series models. Following that, we introduce the ordinary least squares estimation in the AR(1) model and discuss some asymptotic results. We also present two different methods for testing the stationarity of time series: Augmented Dickey–Fuller(ADF) Test and Kwiatkowski–Phillips–Schmidt–Shin(KPSS) Test. The theory of cointegration is then introduced, which allows us to perform a combinatorial analysis of nonstationary time series. Finally, we present the heteroscedastic time series model, which enables the residual sequence of time series to satisfy the conditional variance structure.

The first main result will be presented in Chapter 4. We start with a definition of the VCM. Then, in the varying coefficient AR(1) model, we present the asymptotic theory of the OLS estimator, including the central limit theorem and the large deviation principle. We first prove the existence of upper and lower bounds for the large deviation in the large deviation principle section. The limit of these two bounds is used to demonstrate that they converge to the same value. Second, we derive the OLS estimator's moderate deviation principle.

We will present the second main result in Chapter 5. We will examine annual average temperature and extreme rainfall data from a variety of locations. It will be shown that the annual average temperature of most countries and cities follows the varying coefficient autoregressive model using various statistical methods. However, the test results for the extreme rainfall data do not match what we expected. We found that the series' characteristics do not match to the varying coefficient after testing them. We are unsure of its development trend due to a lack of data. Furthermore, we can only find data for individual stations and not average data for a country or region. As a result, it is impossible to confirm the actual rainfall sequence model. However, we examined the residual series of the extreme rainfall and discovered that for some locations, the residuals exhibit heteroscedasticity.

Chapter 2

Asymptotic Theory in Probability Theory

Asymptotic theory is an integral part of probability theory. Many results become the foundations for statistical inference. In this chapter, we present the necessary definitions and discuss closely related results on the law of large numbers, the central limit theorem, the large deviation principle, and the moderate deviation principle.

2.1 Law of Large Numbers

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\mathbb{R}, \mathcal{B})$ be the one dimensional Euclidean measurable space with the Borel σ-field $\mathcal{B}$.

Theorem 2.1.1 *(One Dimension Weak Law of Large Numbers) Let $X_1, X_2, ...$ be a sequence of i.i.d. random variables on probability space $(\Omega, \mathcal{F}, P)$ with finite expectation μ and covariance $\sigma^2 < \infty$, then*

$$\bar{X}_n := \frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{P} \mu$$

where $\bar{X}_n$ is called the sample mean and $\xrightarrow{P}$ denotes convergence in probability.

According to the law, the average of random variables should be close to the expected value as n approaches infinity.

2.2 Central Limit Theorem

Theorem 2.2.1 *(Central Limit Theorem) Let $X_1, X_2, ..., X_n$ be a sequence of independent and identically distributed, real-valued random variables with mean μ and covariance $\sigma^2 < \infty$ on probability space $(\Omega, \mathcal{F}, P)$, then*

$$Z_n := \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{D} N(0, 1)$$

where $\xrightarrow{D}$ denotes convergence in distribution.

The central limit theorem (CLT) establishes that, in many situations, when independent random variables are summed up, their properly normalized sum tends toward a normal distribution even if the original variables themselves are not normally distributed. In some cases, the random variables may not be identically distributed. As a result, we have a more general theorem known as the Lindeberg-Feller Central Limit Theorem.

Theorem 2.2.2 *Suppose that $X_1, X_2, ...$ are independent random variables on probability space $(\Omega, \mathcal{F}, P)$ such that $E[X_i] = \mu_i$ and $Var[X_i] = \sigma_i^2 < \infty$. Also let $s_n^2 := \sum_{i=1}^n \sigma_i^2$. If this sequence of independent random variable X_i satisfies the Lindeberg's condition:*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} E[(X_i - \mu_i) 1_{|X_i - \mu_i| > \varepsilon s_n}] = 0$$

for all $\varepsilon > 0$*, then*

$$Z_n := \frac{\sum_{i=1}^{n}(X_i - \mu_i)}{s_n} \xrightarrow{D} N(0,1)$$

Furthermore, if the sequence of random variables X_i are not independent, we have a more general theorem which is called the Martingale Difference Central Limit Theorem.

Definition 2.2.1 *Let* $\{X_n\}$ *and* $\{Y_n\}$ *be sequences of random variables on probability space* $(\Omega, \mathcal{F}, P)$*. We said that* $\{X_n\}$ *are martingale differences with respect to* $\{Y_n\}$ *if*

$$X_n = f(y_n, y_{n-1}, ...)$$

for some function f *and*

$$E[X_{n+1}|Y_n, Y_{n-1}, ...] = 0$$

Theorem 2.2.3 *Let* $\{X_n\}$ *be martingale differences with respect to* $\{Y_n\}$*. Suppose that* $\{X_n\}$ *obeys the conditional Lindeberge condition, namely, for any* $\varepsilon > 0$

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} E[X_i^2 1_{|X_i| > \varepsilon s_n} | Y_{i-1}, Y_{i-2}, ...] = 0$$

and

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{i=1}^{n} E[X_i^2 | Y_{i-1}, Y_{i-2}, ...] = 1$$

then

$$\frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow{D} N(0,1)$$

2.3 Large Deviation Principle

Let $\{P_\varepsilon\}$ be a family of probability measures on the measurable space $(\mathbb{R}, \mathcal{B})$. Large deviation principle characterizes the limiting behavior of the family of probability measures $\{P_\varepsilon\}$ in terms of a rate function as $\varepsilon \to 0$. The behavior is shown by asymptotic upper and lower exponential bounds on the values that P_ε assigns to measurable subsets of $\mathcal{X}$. To provide the complete large deviation principle, we firstly introduce some basic definitions.

Definition 2.3.1 *A rate function I is a lower semi-continues mapping*

$$I : \mathcal{X} \to [0, \infty].$$

A good rate function is a rate function for which all the level sets $\Psi_I(\alpha) \triangleq \{x : I(x) \leq \alpha\}$ are compact subsets of $\mathbb{R}$. The effective domain of I, denoted $\mathcal{D}_I$, is the set of points in $\mathcal{X}$ of finite rate, namely,

$$\mathcal{D}_I \triangleq \{x : I(x) < \infty\}$$

Note that if $\mathcal{X}$ is a metric space, the lower semicontinuity property may be checked on sequences, i.e., I is lower semicontinuous if and only if $\liminf_{x_n \to x} I(x_n) \geq I(x)$ for all $x \in \mathcal{X}$. A consequence of a rate function being good is that its infimum is achieved over closed sets.

Now, we can provide the formal definition of the large deviation principle (LDP) and the moderate deviation principle (MDP).

Definition 2.3.2 *Let $\{P_\varepsilon\}$ be a family of probability measures on the measurable space $(\mathcal{X}, \mathcal{B}_\mathcal{X})$ which satisfies the large deviation principle with a rate function I if for all $\Gamma \in \mathcal{B}_\mathcal{X}$, $\bar{\Gamma}$, the closure of Γ and Γ°, the interior of Γ satisfies*

$$-\inf_{x\in\Gamma^\circ} I(x) \leq \liminf_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\Gamma) \leq \limsup_{\varepsilon\to 0} \varepsilon \log \mu_\varepsilon(\Gamma) \leq -\inf_{x\in\bar{\Gamma}} I(x)$$

The right and left hand sides of above equation are referred to as the upper and lower bounds, respectively. Since $\mu_\varepsilon(\mathcal{X}) = 1$ for all ε, it is necessary that $\inf_{x\in\mathcal{X}} I(x) = 0$ for the upper bound to hold. When I is a good rate function, it means that there exists at least one point x which can satisfies the equation $I(x) = 0$. In many cases, a countable family of measures P_n is considered. Then the LDP corresponds to the following statement

Definition 2.3.3 *A sequence of random variables (X_n) with the probability space $(\Omega, \mathcal{F}, P)$ satisfies an Large Deviation Principle with speed a_n and rate function I if $a_n \to \infty$ and for each $A \in \mathcal{B}$*

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{a_n} \log P(X_n \in A) &\geq -\inf_{x\in A^\circ} I(x) \\ \limsup_{n\to\infty} \frac{1}{a_n} \log P(X_n \in A) &\leq -\inf_{x\in\bar{A}} I(x) \end{aligned}$$

where A° and $\bar{A}$ denote the interior and the closure of A, respectively.

Let (a_n) be a sequence of increasing positive numbers satisfying $1 = o(a_n^2)$ and $a_n^2 = o(n)$,

$$a_n \to \infty, \ \frac{a_n}{\sqrt{n}} \to 0$$

Definition 2.3.4 *(Moderate Deviation Principle). We say that a sequence of random variables $(X_n)_{n\geq 1}$ on the probability space $(\Omega, \mathcal{F}, P)$ satisfies an MDP with speed a_n^2 such that the inequalities in definition 2.3.3 hold, and the rate function $I : \mathcal{X} \to [0, \infty)$ if the sequence $(\frac{\sqrt{n}M_n}{a_n})_n$ satisfies an LDP with speed a_n^2 and rate function I.*

2.3.1 Cramér's Theorem

In the Large Deviations Principle, the most classical result is Cramér's Theorem. It was established in Cramér[7], which concerns the large deviation principal associated with the empirical mean of independent and identically distributed random variables valued in a finite $d-$dimensional space. Specifically for dimension one, considering the empirical mean

$$\hat{S}_n = \frac{1}{n}\sum_{j=1}^{n} X_j$$

where $X_1, ..., X_n, ...$ are independent and identically distributed one dimensional random variables with X_1 distributed according to the probability law P. The logarithmic moment generating function associated with the law P is defined as

$$\Lambda(\lambda) = \log E[e^{\lambda X_1}].$$

We also call $\Lambda(\cdot)$ as the cumulant generating function. It has the following properties:

Proposition 2.3.1 *Suppose that λ is the cumulant generating function of a random variable X, then $\Lambda(\lambda)$ has the following properties:*

1. $\Lambda(0) = 0$;

2. $\Lambda(\lambda) > -\infty$ *for all* λ;

3. Λ *is convex;*

4. Λ *is a continuous function.*

To get the rate function, we consider the Fenchel-Legendre transformation of the cumulant generating function $\Lambda(\lambda)$ below:

Definition 2.3.5 *The Fenchel-Legendre transform of $\Lambda(\lambda)$ is*

$$\Lambda^*(x) := \sup_{\lambda \in \mathbb{R}} \{\lambda x - \Lambda(\lambda)\}$$

where $D_{\Lambda^*} := \{x \in \mathbb{R} : \Lambda^*(x) < \infty\}$ is the domain of Λ^*.

Proposition 2.3.2 *Suppose that Λ^* is the Fenchel-Legendre transform of the cumulant generating function of a random variable X, then $\Lambda^*(x)$ has the following properties:*

1. $\Lambda^*(x) \geq 0$ *for all* x

2. Λ^* *is convex*

3. Λ^* *is lower semicontinuous.*

We now present the first fundamental large deviation result.

Theorem 2.3.1 *(Cramér′s Theorem) For every closed subset $F \subset \mathbb{R}$,*

$$\limsup_{n\to\infty} \frac{1}{n} \log P(\hat{S}_n \in F) \leq - \inf_{x\in F} \Lambda^*(x)$$

and for every open subset $G \subset \mathbb{R}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log P(\hat{S}_n \in G) \geq - \inf_{x\in G} \Lambda^*(x)$$

2.3.2 Gärtner Ellis Theorem

Cramér's Theorem is limited to the i.i.d cases. However, the large deviation theorem can be extended to the non-i.i.d. case. Consider a sequence of random variables $X_n \in \mathbb{R}$, where X_n possesses the law P_n and logarithmic moment generating function

$$\Lambda_n(\lambda) = \log E[e^{\lambda X_n}]$$

The existence of a limit of properly scaled logarithmic moment generating functions indicates that P_n may satisfy the LDP. Specifically, the following assumption is imposed throughout this section.

Assumption 2.3.1 *For each $\lambda \in \mathbb{R}$, the logarithmic moment generating function, defined as the limit*

$$\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n} \Lambda_n(n\lambda)$$

exists as an extended real number. Further, the origin belongs to the interior of $D_\Lambda = \{\lambda \in \mathbb{R} : \Lambda(\lambda) < \infty\}$.

In particular, if P_n is the law governing the empirical mean $\hat{S}_n$ of i.i.d. random variables $X_i \in \mathbb{R}$, then for every $n \in \mathbb{N}$

$$\frac{1}{n}\Lambda_n(n\lambda) = \Lambda(\lambda) = \log E[e^{\lambda X_1}]$$

and assumption 2.3.1 holds whenever $0 \in D^\circ_\Lambda$ which is the interior of D_Λ.

Let $\Lambda^*(\cdot)$ be the Fenchel-Legendre transform of $\Lambda(\cdot)$ with $D_{\Lambda^*} = \{x \in \mathbb{R} : \Lambda^*(x) < \infty\}$.

Definition 2.3.6 $y \in \mathbb{R}$ *is an exposed point of* Λ^* *if for some* $\lambda \in \mathbb{R}$ *and all* $x \neq y$*,*

$$\lambda y - \Lambda^*(y) > \lambda x - \Lambda^*(x)$$

λ *in above equation is called an exposing hyperplane.*

Definition 2.3.7 *A convex function* $\Lambda : \mathbb{R} \to (-\infty, \infty]$ *is essentially smooth if*

1. D°_Λ *is non-empty;*
2. $\Lambda(\cdot)$ *is differentiable throughout* D°_Λ*;*
3. $\Lambda(\cdot)$ *is steep, namely,* $\lim_{n\to\infty} |\nabla\Lambda(\lambda_n)| = \infty$ *whenever* $\{\lambda_n\}$ *is a sequence in* D°_Λ *converging to a boundary point of* D°_Λ.

Theorem 2.3.2 *(Gärtner-Ellis)[8] Let Assumption 2.3.1 hold. Then the followings hold.*

1. *For any closed set F*

$$\limsup_{n\to\infty} \frac{1}{n} \log P_n(F) \leq -\inf_{x\in F} \Lambda^*(x)$$

2. *For any open set G*

$$\liminf_{n\to\infty} \frac{1}{n} \log P_n(G) \geq -\inf_{x\in G\bigcap \mathcal{F}} \Lambda^*(x)$$

   where $\mathcal{F}$ is the set of exposed points of Λ^* whose exposing hyperplane belongs to D_Λ°

3. *If Λ is an essentially smooth, lower semicontinuous function, then the LDP holds with the good rate function $\Lambda^*(\cdot)$.*

Chapter 3

Topics on Time Series

A time series is a collection of random variables that are indexed in time. It can range between $-\infty$ and $+\infty$. Time series can be found in many fields, including climatology, economics, and finance. This chapter will define time series and discuss some special time series models.

3.1 Definition of Time Series

Definition 3.1.1 *A time series model for the observed data* $\{x_t\}$ *is a specification of the joint distributions of a sequence of random variables* $\{X_t\}$ *of which* $\{x_t\}$ *is postulated to be a realization.*

Some statistical characteristics must be provided for a time series in order to describe its properties. There are three main quantity features:

Definition 3.1.2 *(Mean function) For a stochastic process* $\{X_t : t = 0, \pm1, \pm2, ...\}$*, the mean function is defined by*

$$\mu_t = E[X_t], t = 0, \pm1, \pm2, ...$$

Definition 3.1.3 *(Autocovariance function) The autocovariance function* $\gamma_{t,s}$ *is defined as*

$$\gamma_{t,s} = Cov(X_t, X_s) \text{ for } t, s = 0, \pm 1, \pm 2, ...$$

Definition 3.1.4 *(Autocorrelation function) The autocorrelation function(ACF)* $\rho_{t,s}$ *is defined as*

$$\rho_{t,s} = Corr(X_t, X_s) \text{ for } t, s = 0, \pm 1, \pm 2, ...$$

where

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

In order to conduct statistical inferences based on observed records, it is necessary to make some assumptions on the underlying time series. The most critical assumption is that the series should be stationary time series. Intuitively, "Stationary" means that the statistical characteristics of a series, mean and autocorrelation do not change over time.

Definition 3.1.5 *A time series* $\{X_t\}$ *is said to be weakly (or second-order) stationary if*

1. *The mean function is constant over time:* $\mu_t = E[X_t] = c$ *where* c *is any constant;*

2. *The autocorrelation function* $\rho_{t,t-k} = \rho_{0,k}$ *for all time* t *and lag* k.

For a weakly stationary time series, the variance of the series can be represented as

$$\gamma_{t,t} = \gamma_0$$

and its autocorrelation function can be represented as

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

3.2 Long Memory Time Series

We can divide time series into long memory and short memory series based on the structure of the corresponding ACF.

Definition 3.2.1 *Suppose $\{X_t\}$ is a stationary time series with autocorelation function ρ_τ. If ρ_τ satisfies*

$$\lim_{n\to\infty} \sum_{\tau=-n}^{n} |\rho_\tau| < \infty$$

then $\{X_t\}$ is called a short memory time series.

Therefore, it is easy to show that the ACF is an even function and for every $k \in \mathbb{N}$ we have $|\rho_k| \leq 1$. The ACF decreases rapidly as k increases for a short memory time series, but slowly as k increases for a long memory time series.

3.2.1 R/S Test

We can use the R/S test to determine whether a time series is a long memory time series. Suppose $\{X_t\}$ is a time series. The sample mean of n observations is defined as

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The R/S test statistic is defined as

$$Q_n = \frac{R(n)}{S(n)}$$

where

$$R(n) = \max_{1\leq k\leq n} \sum_{j=1}^{k}(X_j - \bar{X}_n) - \min_{1\leq k\leq n} \sum_{j=1}^{k}(X_j - \bar{X}_n)$$

and

$$S(n) = [\frac{1}{n}\sum_{j=1}^{n}(X_j - \bar{X}_n)^2]^{\frac{1}{2}}.$$

When $n \to \infty$, it can be shown that

$$H = \lim_{n\to\infty} \frac{\log Q_n}{\log n}$$

If $H \leq 0.5$, it means that the time series $\{X_t\}$ is a short memory time series, but if $H > 0.5$, the series will be a long memory time series.

3.3 Model for Stationary Time Series

3.3.1 White Noise Process

The first example of stationary time series is the white noise process.

Definition 3.3.1 *A stochastic process $\{X_t\}$, with mean zero is said to be a white noise process, if all variables in the process are assumed to be uncorrelated with mean 0 and variance σ^2. If all variables in the process are assumed to be independent and identically* *distributed with mean 0 and variance σ^2, then $\{X_t\}$ is called independent white noise process. Furthermore, if all variables in the process are independent and identically distributed as normal with mean zero and variance σ^2, then $\{X_t\}$ is called Gaussian white noise process.*

3.3.2 Autoregressive Model

Autoregressive model is a special time series model which is widely used in economics and weather. It captures the linear relationship between the series at a certain time t and the series values at the previous p time.

Definition 3.3.2 *A stochastic process, $\{X_t\}$, with zero mean is said to be an autoregressive process of order p, AR(p), if it can be represented as*

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a white noise process. And its AR characteristic polynomial is defined as

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_p x^p$$

Furthermore, when $\phi(x) = 0$, it is called an AR characteristic equation.

If the value $p = 1$, the model will be called the first-order autoregressive process:

Definition 3.3.3 *A stochastic process,* $\{X_t\}$*, with zero mean is said to be an autoregressive process of order* 1*, AR(1), if it can be represented as*

$$X_t = \phi X_{t-1} + \varepsilon_t$$

where $\{\varepsilon_t\}$ *is a white noise process. And its AR characteristic polynomial is represented as*

$$\phi(x) = 1 - \phi x$$

Theorem 3.3.1 *For an AR process, it is weakly stationary if and only if the roots of the AR characteristic equation exceed 1 in absolute value.*

As a result of Theorem 3.3.1, we can obtain the following result.

Theorem 3.3.2 *For an AR(1) process, it is an stationary process if and only if* $|\phi| < 1$*.*

3.3.3 Moving Average Model

The moving average model is a linear model of the white noise process that is defined as follows:

Definition 3.3.4 *A stochastic process,* $\{X_t\}$*, with zero mean is said to be an moving average process of order* q*, MA(p), if it can be represented as*

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

where $\{\varepsilon_t\}$ *is a white noise process.*

3.3.4 Mixed Autoregressive Moving Average Model

When we assume that the series is partly autoregressive and partly moving average, we obtain a very general time series model known as the mixed autoregressive moving average model.

Definition 3.3.5 *A stochastic process, $\{X_t\}$, with zero mean is said to be an mixed autoregressive moving average model of orders p and q, $ARMA(p,q)$, if it can be represented as*

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

where $\{\varepsilon_t\}$ is a white noise process.

The stationarity of the ARMA process is determined by its AR components, and we have the following theorem.

Theorem 3.3.3 *For an ARMA process, it is weakly stationary if and only if the roots of the AR characteristic equation exceed 1 in absolute value.*

3.4 Model for Non-Stationary Time Series

3.4.1 Wiener Process

Definition 3.4.1 *A stochastic process $\{W(t)\}$ is called a wiener process if it satisfies the following four conditions:*

1. $W(0) = 0$

2. *$W(t)$ has independent increments: for every $t > 0$, the future increments $W(t+u) - W(t), u \geq 0$ are independent of the past values $W(s), s \leq t$*

3. *$W(t)$ has Gaussian increments: $W(t+u) - W(t) \sim N(0, \sigma^2 u)$ where $\sigma > 0$ is a constant.*

4. *$W(t)$ has continuous paths: $W(t)$ is continuous in t.*

Based on above definition, we could find that

$$Var(W(t)) = t\sigma^2$$

It is clear that the variance of the process depends on time and will go to infinity with time $t \to \infty$. Therefore, the Wiener process is not stationary. If we let $t = 1, 2, ... \in \mathbb{N}$, then we could find

$$\varepsilon_t := W(t) - W(t-1) \sim N(0, \sigma^2) \tag{3.1}$$

which means that the increments of wiener process $\{\varepsilon_t\}$ is a Gaussian white noise process and it is a stationary process.

Definition 3.4.2 *If $W(t)$ is a wiener process with mean zero and variance t, then*

$$B(t) = W(t) - \frac{t}{T} W(T)$$

is called a Brownian bridge for $t \in [0, T]$.

3.4.2 Integrated Series

Definition 3.4.3 *A series with no deterministic component which has a stationary, ARMA representation after differencing d times, is said to be integrated of order d, denoted* $X_t \sim I(d)$.

For ease of exposition, only the values $d = 0$ and $d = 1$ will be considered in much of the paper. For $d = 0$, $\{X_t\}$ will be a stationary series which is referred to as an $I(0)$ or a "level stationary" process. For $d = 1$, the change of $\{X_t\}$ will be a stationary series. Therefore, the process $\{X_t\}$ is referred to as an $I(1)$ process. A process that requires differencing twice to achieve stationarity is referred to as an $I(2)$ process. There are substantial differences in appearance between $I(0)$ series and $I(1)$ series.

1. If $X_t \sim I(0)$
   (a) Variance of X_t is finite;
   (b) Autocorrelation ρ_k decrease steadily in magnitude for large enough k.
2. If $X_t \sim I(1)$
   (a) Variance of X_t goes to infinity as t goes to infinity;
   (b) Autocorrelation ρ_k goes to 1 for all k as $t \to \infty$.

For a first-order autoregressive model, we already know that the model is stationary only when $|\phi| < 1$, and when $\phi = 1$, the model will be an $I(1)$ process. We also name it as Unit Root process. The first difference of this model will be the white noise process, which is stationary.

3.5 Parameter Estimation

For an AR(1) process, the value of ϕ plays a vital role. Because its value determines the stationarity of the process. To estimate its value, least squares method is a most common way which could provide an unbiased estimator. According to the structure of AR(1) model, we can provide the asymptotic distribution of least squares estimator $\hat{\phi}$ under the stationary and non-stationary case.

3.5.1 Ordinary Least Squares Estimation

For an AR(1) process $\{Y_t\}$ where

$$Y_t = \phi Y_{t-1} + \varepsilon_t, \quad t = 1, 2, ..., T \tag{3.2}$$

and

1. ε_t is a Gaussian white noise process;
2. $Y_0 = y_0$ which is any fixed constant;
3. $y_1, y_2, ..., y_T$ are observed values of $\{Y_t\}_{t=1,2,..,T}$.

By Ordinary Least Squares (OLS) method, we could estimate the value of ϕ by

$$\hat{\phi}_T = \frac{\sum_{t=1}^{T} Y_{t-1} Y_t}{\sum_{t=1}^{T} Y_{t-1}^2} \tag{3.3}$$

Theorem 3.5.1 *In AR(1) process, the least squares estimator $\hat{\phi}_T$ in equation (3.3) is an unbiased estimator of ϕ with variance $\frac{\sigma^2}{\sum_{t=1}^{T} y_{t-1}^2}$.*

The difference between estimator $\hat{\phi}$ and the true value ϕ is as follows:

$$\hat{\phi}_T - \phi = \frac{\sum_{t=1}^{T} Y_{t-1}\varepsilon_t}{\sum_{t=1}^{T} Y_{t-1}^2}$$

3.5.2 Asymptotic Distribution of Estimator

First of all, based on the Definition 2.2.1, it is easy to show that $\{Y_{t-1}\varepsilon_t\}$ is a martingale difference with respect to $\{\varepsilon_t\}$. Therefore, we can prove that the least squares estimator $\hat{\phi}_T$ satisfies the law of large number and central limit theorem under some specific conditions.

Theorem 3.5.2 *(Weak Law of Large Numbers) For a first order Autoregressive model $\{Y_t\}$, if $|\phi| < 1$, the ordinary least squares estimator $\hat{\phi}_T$ converges in probability to ϕ.*

Theorem 3.5.3 *For an AR(1) process in equation (3.2) with $|\phi| < 1$, $(\hat{\phi}_T - \phi)$ satisfies the Linderberg's condition so that*

$$\tau_T := \frac{(\sum_{t=1}^{T} Y_{t-1}^2)^{\frac{1}{2}}}{\sigma}(\hat{\phi}_T - \phi) \xrightarrow{D} N(0,1)$$

When $\phi = 1$, AR(1) process is an integrated $I(1)$ series. The first difference of this model will be the white noise process $\{\varepsilon_t\}$. We can get the following asymptotic distribution from [21]:

1. $T^{-\frac{1}{2}} \sum_{t=1}^{T} \varepsilon_t \xrightarrow{D} \sigma W(1)$

2. $T^{-\frac{3}{2}} \sum_{t=1}^{T} Y_{t-1} \xrightarrow{D} \sigma \int_0^1 W(r)dr$

3. $T^{-2} \sum_{t=1}^{T} Y_{t-1}^2 \xrightarrow{D} \sigma^2 \int_0^1 W(r)^2 dr$

4. $T^{-1} \sum_{t=1}^{T} Y_{t-1}\varepsilon_t \xrightarrow{D} \frac{1}{2}\sigma^2 (W^2(1) - 1)$

5. $T^{\frac{3}{2}} \sum_{t=1}^{T} t\varepsilon_t \xrightarrow{D} \sigma \int_0^1 r dW(r)$

6. $T^{-\frac{5}{2}} \sum_{t=1}^{T} tY_{t-1} \xrightarrow{D} \sigma \int_0^1 rW(r)dr$

where σ^2 is the variance of ε_t. Based on above result, we could get the following result:

Theorem 3.5.4 *For an AR(1) process in equation (3.2) with* $\phi = 1$

$$\tau_T \xrightarrow{D} \frac{\frac{1}{2}(W^2(1) - 1)}{(\int_0^1 W^2(r)dr)^{1/2}}$$

Therefore, based on the above two theorems, the statistic τ_T will converge to a standard normal distribution if the process is stationary. But when process is non-stationary, the variable τ_T will converge to a lévy process which is a stochastic process with independent, stationary increments. Based on the asymptotic distribution of τ_T, we can determine whether the value of ϕ is less than or equal to one.

3.5.3 Large Deviation Principle

In previous sections, it has been shown that the least squares estimator $\hat{\phi}$ is an unbiased estimator. This estimator $\hat{\phi}$ should converge to its true value ϕ. But there still should be a probability that the estimator $\hat{\phi}$ is far from its true value. We can use the large deviation principle to find the probability of the event that the estimator is far from its true value.

Theorem 3.5.5 *[3] For an AR(1) process in equation (3.2) with $|\phi| < 1$, the law of the Yule-Walker estimator $\tilde{\phi}_T = \frac{\sum_{t=1}^{T} Y_{t-1}Y_t}{\sum_{t=0}^{T} Y_t^2}$ satisfies a large deviation principle with speed T and good rate function*

$$I(x) = \begin{cases} \frac{1}{2}\log\left(\frac{1+\phi^2-2\phi x}{1-x^2}\right) & \text{if } x \in [-1,1] \\ \infty & \text{otherwise} \end{cases}$$

Theorem 3.5.6 *[29] Suppose that the moderate deviation scale (b_T) is a sequence of positive numbers satisfying*

$$b_T \to \infty, \frac{b_T}{\sqrt{T}} \to 0 \text{ as } T \to \infty$$

Then $\frac{\sqrt{T}}{b_T}(\hat{\phi}_T - \phi)$ as well as $\frac{\sqrt{T}}{b_T}(\tilde{\phi}_T - \phi)$ satisfie the large deviation principle with speed b_T^2 and rate function

$$I(x) = \frac{x^2}{2(1-\phi^2)}$$

For the unit root process, even the process is non-stationary, the Yule-Walker estimator $\hat{\phi}_T$ still satisfies the large deviation principle:

Theorem 3.5.7 *[1] For an AR(1) process in equation (3.2), if $\phi = 1$, the law of the Yule-Walker estimator $\tilde{\phi}_T$ satisfies a large deviation principle with speed T and good rate* *function*

$$I(x) = \begin{cases} \frac{1}{2}\log(\frac{2}{1+x}) & \textit{if } x \in (-1,1] \\ \\ \infty & \textit{otherwise} \end{cases}$$

and if $\phi = -1$*, the law of the Yule-Walker estimator* $\tilde{\phi}_T$ *satisfies a large deviation principle with speed* T *and good rate function*

$$I(x) = \begin{cases} \frac{1}{2}\log(\frac{2}{1-x}) & \textit{if } x \in [-1,1) \\ \\ \infty & \textit{otherwise} \end{cases}$$

Based on above theorems, we can find that the rate function could help us determine how fast the estimator converges.

3.6 Deterministic Trend

A trending mean is a common stationarity violation. For non-stationary series with a trending mean, there are two popular models. The first is known as trend stationary. That is, the mean trend is deterministic. The residual series is a stationary stochastic process after the trend is estimated and removed from the data. The second is known as a stochastic trend. This indicates that the mean trend is stochastic. A stationary stochastic process is obtained by differencing the series d times. The distinction between deterministic and stochastic trends has significant implications for a process's long-term behaviour. In the long run, a time series with a deterministic trend will always revert to the trend. However, if the series has a stochastic trend, it will never recover from system shocks. The Mann-Kendall test is the most commonly used test in time series to determine deterministic trend.

3.6.1 Mann-Kendall Test

The Mann-Kendall test is a test which is commonly employed to detect monotonic trends in series of environmental data, climate data or hydrological data. The null hypothesis, H_0, is that the data come from a population with independent realizations and are identically distributed. The alternative hypothesis, H_1, is that the data follow a monotonic trend. The Mann-Kendall test statistic is calculated according to:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^{n} sgn(Y_j - Y_k)$$

with

$$sgn(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \\ -1 & \text{if } x < 0 \end{cases}$$

where n is the length of the sample, X_i and X_j are from $k = 1, 2, ..., n-1$ and $j = k+1, k+2, ..., n$. If n is bigger than 9, statistic S approximates to normal distribution. The mean of S is equal to zero and the variance of S can be acquired as follows:

$$Var(S) = \frac{n(n-1)(2n+5)}{18}$$

Then the test statistic Z is denoted by

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var(S)}} & \text{if } S > 0 \\ 0 & \text{if } S = 0 \\ \frac{S+1}{\sqrt{Var(S)}} & \text{if } S < 0 \end{cases}$$

If $Z > 0$, it indicates an increasing trend, and vice versa. Given a confidence level α, the sequential data would be supposed to experience statistically significant trend if

$$|Z| > Z_{1-\frac{\alpha}{2}}$$

where $Z_{1-\frac{\alpha}{2}}$ is the corresponding value of $P-$value that is equal to $\frac{\alpha}{2}$ following the standard normal distribution.

3.7 Test for Stationarity

There are two most common methods to test the stochastic stationarity of the time series: Dickey-Fuller Test and KPSS Test.

3.7.1 Dickey-Fuller Test

The Dickey-Fuller test in statistics tests the null hypothesis that a unit root exists in an autoregressive time series model. The alternative hypothesis varies depending on the version of the test used, but it is typically stationarity or trend-stationarity. The test was created in 1979 by statisticians David Dickey and Wayne Fuller[9] and is named after them.

Consider an AR(1) process

$$X_t = \phi X_{t-1} + \varepsilon_t, X_0 = 0, t = 1, 2, ..., T$$

where ε_t is a Gaussian white noise process. The regression model can be written as

$$\Delta X_t = (\phi - 1)X_{t-1} + \varepsilon_t = \delta X_{t-1} + \varepsilon_t$$

where Δ is the first difference operator such that

$$\Delta X_t := X_t - X_{t-1}$$

and $\delta = \phi - 1$. This model can be estimated and testing for a unit root is equivalent to testing $\delta = 0$. When $\delta = 0$, it is equivalent to that $\phi = 1$. It means that the series contain a unit root. As we show in the previous section, for an AR(1) process, the statistic τ_T has different asymptotic distribution under stationary and non-stationary conditions. In this test, the null hypothesis of DF test is $H_0 : \phi = 1$ which means the series $\{X_t\}$ is a unit root process. Therefore, we could construct the following two test statistic:

1. $T(\hat{\phi} - \phi)$
2. $t_T = \frac{\hat{\phi}-\phi}{S(\hat{\phi})}$

where $S(\hat{\phi})$ represents the standard deviation of the estimator $\hat{\phi}$. With large sample size, if the process is a stationary process, then the distribution of t_T should follow a standard normal distribution. Otherwise, if $|\phi| = 1$, the test statistic will follow a stochastic process which is shown in the Theorem 3.5.4.

There are three main versions of the test:

1. Tests for a unit root:

$$\Delta X_t = \delta X_{t-1} + \varepsilon_t$$

2. Test for a unit root with constant:

$$\Delta X_t = a_0 + \delta X_{t-1} + \varepsilon_t$$

3. Test for a unit root with constant and deterministic time trend:

$$\Delta X_t = a_0 + a_1 t + \delta X_{t-1} + \varepsilon_t$$

Each version of the test has different asymptotic distribution under the null hypothesis:

For situation 1:

$$T(\hat{\phi} - 1) = \frac{T^{-1} \sum_{t=1}^{T} Y_t Y_{t-1}}{T^{-2} \sum_{t=1}^{T} Y_{t-1}^2}$$

and

$$t_T = \frac{\hat{\phi} - 1}{S(\hat{\phi})} = \frac{\hat{\phi} - 1}{S_\varepsilon (\sum_{t=1}^{T} Y_{t-1}^2)^{-\frac{1}{2}}}$$

where

$$S_\varepsilon^2 = \frac{1}{T-1}\sum_{t=1}^{T}(Y_t - \hat{\phi}Y_{t-1})^2.$$

When $T \to \infty$, it can be shown that

$$t_T \xrightarrow{D} \frac{\frac{1}{2}(W^2(1)-1)}{(\int_0^1 W^2(r)dr)^{\frac{1}{2}}}$$

$$T(\hat{\phi}-1) \xrightarrow{D} \frac{\frac{1}{2}(W^2(1)-1)}{\int_0^1 W^2(r)dr}$$

For situation 2, the null hypothesis is that $a_0 = 0$ and $\phi = 1$ and the least squares estimator of a_0 and ϕ can be represented by

$$\begin{bmatrix} \hat{a}_0 \\ \hat{\phi} \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^T Y_{t-1} \\ \sum_{t=1}^T Y_{t-1} & \sum_{t=1}^T Y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T Y_{t-1} \\ \sum_{t=1}^T Y_{t-1}Y_t \end{bmatrix}$$

when null hypothesis is true, it can be shown that

$$\begin{bmatrix} \hat{a}_0 \\ \hat{\phi}-1 \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^T Y_{t-1} \\ \sum_{t=1}^T Y_{t-1} & \sum_{t=1}^T Y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T \varepsilon_t \\ \sum_{t=1}^T Y_{t-1}\varepsilon_t \end{bmatrix}$$

and

$$\begin{bmatrix} T^{\frac{1}{2}}\hat{a}_0 \\ T(\hat{\phi}-1) \end{bmatrix} = \begin{bmatrix} 1 & T^{-\frac{3}{2}}\sum_{t=1}^T Y_{t-1} \\ T^{-\frac{3}{2}}\sum_{t=1}^T Y_{t-1} & T^{-2}\sum_{t=1}^T Y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}}\sum_{t=1}^T \varepsilon_t \\ T^{-1}\sum_{t=1}^T Y_{t-1}\varepsilon_t \end{bmatrix}$$

If $T \to \infty$, it can be shown that

$$\begin{bmatrix} 1 & T^{-\frac{3}{2}}\sum_{t=1}^{T} Y_{t-1} \\ T^{-\frac{3}{2}}\sum_{t=1}^{T} Y_{t-1} & T^{-2}\sum_{t=1}^{T} Y_{t-1}^{2} \end{bmatrix} \xrightarrow{D} \begin{bmatrix} 1 & \sigma\int_0^1 W(r)dr \\ \sigma\int_0^1 W(r)dr & \sigma^2\int_0^1 W^2(r)dr \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix} \times \begin{bmatrix} 1 & \int_0^1 W(r)dr \\ \int_0^1 W(r)dr & \int_0^1 W^2(r)dr \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & \sigma \end{bmatrix}$$

and

$$\begin{bmatrix} T^{-\frac{1}{2}}\sum_{t=1}^{T} \varepsilon_t \\ T^{-1}\sum_{t=1}^{T} Y_{t-1}\varepsilon_t \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \sigma W(1) \\ \frac{1}{2}\sigma^2(W^2(1)-1) \end{bmatrix} = \begin{bmatrix} \sigma & 0 \\ 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} W(1) \\ \frac{1}{2}(W^2(1)-1) \end{bmatrix}$$

Therefore, under null hypothesis

$$T(\hat{\phi}-1) \xrightarrow{D} \frac{\frac{1}{2}(W^2(1)-1) - W(1)\int_0^1 W(r)dr}{\int_0^1 W^2(r)dr - (\int_0^1 W(r)dr)^2}$$

and

$$S^2(\hat{\phi}) = S_\varepsilon^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} T & \sum_{t=1}^{T} Y_{t-1} \\ \sum_{t=1}^{T} Y_{t-1} & \sum_{t=1}^{T} Y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

where S_ε^2 is the least squares estimator of variance of $\{\varepsilon_t\}$ which is represented as

$$S_\varepsilon^2 = \frac{1}{T-2}\sum_{t=1}^{T}(Y_t - \hat{a}_0 - \hat{\phi}Y_{t-1})^2$$

Furthermore,

$$\begin{aligned} T^2S^2(\hat{\phi}) &= S_\varepsilon^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} T & \sum_{t=1}^T Y_{t-1} \\ \sum_{t=1}^T Y_{t-1} & \sum_{t=1}^T Y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-\frac{1}{2}} & 0 \\ 0 & T \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ &= S_\varepsilon^2 \begin{bmatrix} 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & T^{-\frac{3}{2}}\sum_{t=1}^T Y_{t-1} \\ T^{-\frac{3}{2}}\sum_{t=1}^T Y_{t-1} & T^{-2}\sum_{t=1}^T Y_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{aligned}$$

Since S_ε^2 is a consistent estimator of σ^2, it can be shown that

$$T^2S^2(\hat{\phi}) \xrightarrow{D} \frac{1}{\int_0^1 W^2(r)dr - (\int_0^1 W(r)dr)^2}$$

Therefore, we can get

$$t_T \xrightarrow{D} \frac{\frac{1}{2}(W^2(1)-1) - W(1)\int_0^1 W(r)dr}{\{\int_0^1 W^2(r)dr - (\int_0^1 W(r)dr)^2\}^{\frac{1}{2}}}$$

For situation 3, the null hypothesis is that $a_1 = 0$ and $\rho = 1$. The model can be represented as

$$Y_t = \theta^T Z_{t-1} + \varepsilon_t$$

where

$$Z_t = \begin{bmatrix} 1 \\ Y_t - a_0 t \\ t \end{bmatrix}$$

$$\theta = \begin{bmatrix} a_0 + a_1 \\ \phi \\ a_1 + \phi 0 \end{bmatrix}$$

The least squares estimator of vector θ can be represented as

$$\hat{\theta} = (\sum_{t=2}^{T} Z_{t-1} Z_{t-1}^T)^{-1} \sum_{t=2}^{T} Y_t Z_{t-1}$$

We define

$$T_T = \begin{bmatrix} T^{\frac{1}{2}} & 0 & 0 \\ 0 & T & 0 \\ 0 & 0 & T^{\frac{3}{2}} \end{bmatrix}$$

It can be shown that

$$T_T(\hat{\theta} - \theta) = V_T^{-1} \phi_T$$

where

$$V_T = T_T^{-1} \sum_{t=2}^{T} Z_{t-1} Z_{t-1}^T T_T^{-1}$$

$$\phi_T = T_T^{-1} \sum_{t=2}^{T} Z_{t-1} \varepsilon_t$$

Under the null hypothesis, it can be shown that

$$
\begin{aligned}
V_{T,1,1} &= T^{-1}\sum_{t=2}^{T} 1 \rightarrow 1 \\
V_{T,1,2} &= T^{-\frac{3}{2}}\sum_{t=2}^{T} S_{t-1} \xrightarrow{D} \sigma\int_0^1 W(r)dr \\
V_{T,1,3} &= T^{-2}\sum_{t=2}^{T}(t-1) \rightarrow \frac{1}{2} \\
V_{T,2,2} &= T^{-2}\sum_{t=2}^{T} S_{t-1}^2 \xrightarrow{D} \sigma^2\int_0^1 (W(r))^2 dr \\
V_{T,2,3} &= T^{-\frac{5}{2}}\sum_{t=2}^{T}(t-1)S_{t-1} \xrightarrow{D} \sigma\int_0^1 rW(r)dr \\
V_{T,3,3} &= T^{-3}\sum_{t=2}^{T}(t-1)^2 \rightarrow \frac{1}{3}
\end{aligned}
$$

and

$$
\begin{aligned}
\phi_{T,1} &= T^{-\frac{1}{2}}\sum_{t=2}^{T} \varepsilon_t \xrightarrow{D} \sigma W(1) \\
\phi_{T,2} &= T^{-1}\sum_{t=2}^{T} S_{t-1}u_t \xrightarrow{D} \frac{\sigma^2}{2}[(W(1))^2 - 1] \\
\phi_{T,3} &= T^{-\frac{3}{2}}\sum_{t=2}^{T}(t-1)\varepsilon_t \xrightarrow{D} \sigma\int_0^1 r dW(r)
\end{aligned}
$$

where

$$S_t = \sum_{t=1}^{t} \varepsilon_t$$

Therefore, it is easy to find that

$$\begin{aligned} t_T &= (S_\varepsilon^2 V_{T,2,2})^{-\frac{1}{2}} T(\hat{\phi}-1) \\ T(\hat{\phi}-1) &= (V_T^{-1}\phi_T)_2 \end{aligned}$$

when $T \to \infty$, we can get

$$\begin{aligned} t_T \quad \overset{D}{\to} \quad & \sqrt{3}|A|^{-\frac{1}{2}}\{\frac{1}{3}W(1)\int_0^1 W(r)dr - W(1)\int_0^1 rW(r)dr \\ & + \frac{1}{12}(W^2(1)-1) + 2\int_0^1 W(r)dr \int_0^1 rW(r)dr - (\int_0^1 W(r)dr)^2\} \end{aligned}$$

$$\begin{aligned} T(\hat{\phi}-1) \quad \overset{D}{\to} \quad & \frac{1}{|A|}\{\frac{1}{6}W(1)\int_0^1 W(r)dr - \frac{1}{2}W(1)\int_0^1 rW(r)dr \\ & + \frac{1}{24}(W^2(1)-1) + \int_0^1 W(r)dr \int_0^1 W(r)dr \int_0^1 rW(r)dr \\ & - \frac{1}{2}(\int_0^1 W(r)dr)^2\} \end{aligned}$$

where

$$|A| = \frac{1}{12}\int_0^1 W^2(r)dr + \int_0^1 W(r)dr \int_0^1 rW(r)dr - (\int_0^1 rW(r)dr)^2 - \frac{1}{3}(\int_0^1 W(r)dr)^2$$

For above three situations, we can find that the asymptotic distribution of t_T and $T(\hat{\phi}-1)$ do not depend on the variance of ε_t. By the Monte Carlo method, we could get the critical value of the above three versions' tests. The augmented Dickey-Fuller test (ADF test) is an extension of the Dickey-Fuller test that can be used to test the stationarity of the higher-order autoregressive model. This method can also be applied to the ARMA model and is known as the Said-Dickey Test. For hypothesis testing, we can use the same table of critical values as the Dickey-Fuller test statistic for augmented Dickey-Fuller and Said-Dickey tests.

3.7.2 KPSS Test

KPSS test was firstly provided in 1992 by Kwiatkowski, Phillips, Schmidt and Shin[16]. Contrary to Dickey-Fuller test, the presense of a unit root is not the null hypothesis but the alternative. For a time series $\{X_t\}$, it can be represented as follows:

$$X_t = \mu + \alpha t + U_t$$

where $\{U_t\}$ is the difference of $\{X_t\}$ and its deterministic trend $\{\mu + \alpha t\}$. The process $\{U_t\}$ can be represented as:

$$U_t = R_t + \epsilon_t$$

where $\{R_t\}$ is a random walk process which is independent with $\{\epsilon_t\}$ and it can be represented as follows:

$$R_t = R_{t-1} + \xi_t$$

where $\{\xi_t\}$ is a Gaussian white noise which is independent with $\{\epsilon_t\}$.

If the variance of $\{\xi_t\}$ is equal to zero, then there is no random walk in the series $\{X_t\}$ and the series is a trend stationary process. To construct the statistic, we firstly define the partial sum process S_t as follows:

$$S_t = \sum_{i=1}^{t} e_t, t = 1, 2, ..., T$$

where $e_t = X_t - \bar{X}$.

Secondly, we define the "long-run variance" as

$$\sigma^2 = \lim_{T\to\infty} \frac{E[S_T^2]}{T}$$

which will enter into the asymptotic distribution of the test statistics. We can find a consistent estimator of σ^2, say $s^2(l)$ which is constructed from the residuals e_t:

$$s^2(l) = \frac{1}{T}\sum_{t=1}^{T} e_t^2 + 2\frac{1}{T}\sum_{s=1}^{l} w(s,l) \sum_{t=s+1}^{T} e_t e_{t-s}$$

where

$$w(s,l) = 1 - \frac{s}{l+1}$$

So we can build a test statistic as

$$\eta = \frac{\sum_{t=1}^{T} S_t^2}{Ts^2(l)}$$

If the process is a stationary time series with constant mean, the asymptotic distribution of η will be as

$$\eta \xrightarrow{D} \int_0^1 V(r)^2 dr$$

where $V(r)$ is a standard Brownian bridge: $V(r) = W(r) - rW(1)$. If the process is a trend stationary series, then

$$\eta \xrightarrow{D} \int_0^1 V_2(r)^2 dr$$

where $V_2(r)$ is a second level Brownian bridge which is given by

$$V_2(r) = W(r) + (2r - 3r^2)W(1) + (-6r + 6r^2)\int_0^1 W(s)ds$$

If the process is a non-stationary series, the test statistic η will diverge.

3.8 Cointegration

Many traditional asymptotic theories based on the least squares method are based on series stationarity, whereas many actual time series are not. When we perform regression analysis on non-stationary time series, we will arrive at some incorrect conclusions, which we refer to as the "Spurious Regression" problem.

3.8.1 Spurious Regression Problem

Consider the following two time series:

$$\begin{aligned} X_t &= X_{t-1} + u_t, t = 1, 2, ..., T \\ Y_t &= Y_{t-1} + v_t, t = 1, 2, ..., T \end{aligned}$$

where $\{u_t\}$ and $\{v_t\}$ are Gaussian white noise processes. The processes $\{X_t\}$ and $\{Y_t\}$ are uncorrelated. Since there is no influence from $\{X_t\}$ on $\{Y_t\}$, if we do the regression from $\{Y_t\}$ on $\{X_t\}$, the relation can be represented as follows

$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t, t = 1, 2, ..., T$$

where β_1 should be zero. Since $\{u_t\}$ and $\{v_t\}$ are Gaussian white noises, it can be shown that when $T \to \infty$,

1. $\frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^{T} Y_t \xrightarrow{D} \sigma_u \int_0^1 W_u(r)dr$
2. $\frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^{T} X_t \xrightarrow{D} \sigma_u \int_0^1 W_v(r)dr$
3. $\frac{1}{T^2} \sum_{t=1}^{T} Y_t^2 \xrightarrow{D} \sigma_u^2 \int_0^1 (W_u(r))^2 dr$
4. $\frac{1}{T^2} \sum_{t=1}^{T} X_t^2 \xrightarrow{D} \sigma_v^2 \int_0^1 (W_u(r))^2 dr$

where σ_u and σ_v are the standard deviations of white noise process $\{u_t\}$ and $\{v_t\}$. Based on above results, it is known[16] that

1. $\frac{1}{T^2} \sum_{t=1}^{T} (Y_t - \bar{Y})^2 \xrightarrow{D} \sigma_u^2 [\int_0^1 (W_u(r))^2 dr - (\int_0^1 (W_u(r)) dr)^2]$
2. $\frac{1}{T^2} \sum_{t=1}^{T} (X_t - \bar{X})^2 \xrightarrow{D} \sigma_u^2 [\int_0^1 (W_v(r))^2 dr - (\int_0^1 (W_v(r)) dr)^2]$
3. $\frac{1}{T^2} \sum_{t=1}^{T} Y_t X_t \xrightarrow{D} \sigma_u \sigma_v \int_0^1 W_u(r) W_v(r) dr$

Therefore, we can get the asymptotic distribution of $\hat{\beta}_1$, the OLS estimator of β_1, as follows:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{t=1}^{T}(Y_t - \bar{Y})(X_t - \bar{X})}{\sum_{t=1}^{T}(X_t - \bar{X})^2} \\ &\xrightarrow{D} \frac{\sigma_u}{\sigma_v}\xi\end{aligned}$$

where

$$\xi = \frac{\int_0^1 W_u(i)W_v(i)di - \int_0^1 W_u(i)di \cdot \int_0^1 W_v(i)di}{\int_0^1 (W_v(i))^2 di - (\int_0^1 W_v(i)di)^2}$$

We can also find that the value of $\hat{\beta}_1$ does not converge to the constant zero when $T \to \infty$. As a result, the asymptotic theory based on least squares estimation is predicated on the assumption that the time series of the regression variable is stationary, and regression that does not satisfy stationarity may result in spurious regression.

3.8.2 Test for Cointegration

Since most time series are nonstationary, we usually transform them into stationary time series. The most common method is to differentiate the time series in order to make it stationary. However, this method discards some important long-term information. Engle and Granger[11] introduced the cointegration theory in 1987, which can reflect the equilibrium relationship of two or more nonstationary time series.

Definition 3.8.1 *The components of the vector time series $\{X_t\}$ are said to be cointegrated of order d, b, denoted $X_t \sim CI(d, b)$, if*

1. *all components of X_t are $I(d)$*

2. there exists a non zero vector α so that

$$Z_t = \alpha^T X_t \sim I(d-b), b > 0$$

where the vector α is called the cointegrating vector.

The Engle-Granger Two-Step method begins by generating residuals based on the static regression, which are then tested for the presence of unit roots. To test for stationarity in time series, it employs the Augmented Dickey-Fuller Test (ADF) or other tests. The Engle-Granger method will demonstrate residual stationarity if the time series are cointegrated.

3.9 Time Series Models of Heteroscedasticity

Time series with the following characteristics are common in some macroeconomic and financial fields: After removing the influence of deterministic non-stationary factors, the fluctuation of the residual series is stationary in most periods, but the fluctuation will remain significant at times and small at others, indicating a volatility cluster effect. This means that the variance of white noise is not constant in this model. Engle[10] first proposed the autoregressive conditional heteroscedasticity (ARCH) model for modeling the changing variance of a time series.

3.9.1 Definition of ARCH model

Definition 3.9.1 *A p^{th} order autoregressive conditional heteroskedasticity(ARCH) process is a model for the residual process where $\{r_t\}$ satisfying:*

1. $r_t = \sigma_t \varepsilon_t$

2. $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2$

where ω is any constant and $\{\varepsilon_t\}$ is a Gaussian white noise process with zero mean and variance of one.

It is not uncommon that p needs to be very big in order to capture all the serial correlation in r_t^2. The generalized ARCH or GARCH mdoel is a parsimonious alternative to an ARCH(p) model.

Definition 3.9.2 *A GARCH(p,q) is a model for the residual process $\{r_t\}$ satisfying:*

1. $r_t = \sigma_t \varepsilon_t$

2. $\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$

where ω is any constant and $\{\varepsilon_t\}$ is a Gaussian white noise process with zero mean and variance of one.

3.9.2 Test for ARCH model

The Ljung-Box test is commonly used to determine whether a residual series is an ARCH process. The Ljung-Box test's null hypothesis is that the squared residual process is white noise. The squared residual process is autocorrelated, according to the alternative hypothesis. The test statistics is

$$Q := n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

where n is the sample size, $\hat{\rho}_k^2$ is the sample autocorrelation at lag k, and h is the number of lags being tested. Under the null hypothesis, the statistic Q asymptotically follows a χ_h^2. For a significance level α, the critical region for rejection of the hypothesis of randomness is:

$$Q > \chi_{1-\alpha,h}^2$$

Chapter 4

Varying Coefficient Autoregreessive Model

In contrast to the fixed coefficient autoregressive model, the varying coefficient autoregressive model allows the coefficients in the model to vary over time. Consider the following Gaussian AR(1) process:

$$X_t(T) = \phi_T X_{t-1}(T) + \varepsilon_t, \quad t = 1, 2, ..., T$$

where $\phi_T = 1 - \frac{\gamma}{a_T}$, γ is a positive constant and $\{a_T\}$ is a sequence of constants which increases to infinity.

When we only have a few observations and the value of n is small in this model, the value of ϕ_T is strictly less than one. As a result, this model represents a weakly stationary process. However, as the number of observations increases, the value of ϕ_T approaches one, transforming this weakly stationary AR(1) process into a unit root process. As a result, in this chapter, we will present the asymptotic theory of the OLS estimator of the varying coefficient model with the form $\phi_T = 1 - \frac{1}{T}$.

4.1 Asymptotic Distribution of OLS estimator

When T goes to infinity, the value of ϕ_T will go to one, and the process will become a unit root process. The variance of $\hat{\phi}_T$ will go to infinity. As a result, it will fail the central limit theorem because it will not satisfy the Linderberg condition. However, we could find the limit distribution of the estimator $\hat{\phi}_T$ as follows:

Theorem 4.1.1 *[22] Let $\phi_T = 1 - \frac{1}{T}$. For $t = 1, 2, ...T$, suppose $\{X_t\}$ satisfies the AR(1) model*

$$X_t(T) = \phi_T X_{t-1}(T) + \varepsilon_t, X_0(T) = 0 \text{ for all } T \tag{4.1}$$

and $\{\varepsilon_t\}$ is a i.i.d. white noise process where the variance is equal to $\sigma^2 < \infty$. Then, as $T \to \infty$, the least squares estimator $\hat{\phi}_T$ will converge to 1 in probability.

Theorem 4.1.2 *[4] Let $\phi_T = 1 - \frac{\gamma}{T}$. For $t = 1, 2, ...T$, suppose $\{X_t\}$ satisfies the AR(1) model*

$$X_t(T) = \phi_T X_{t-1}(T) + \varepsilon_t, X_0(T) = 0 \text{ for all } T$$

and $\{\varepsilon_t\}$ is a i.i.d. white noise process where the variance is equal to $\sigma^2 < \infty$. Then, as $T \to \infty$

$$\tau_T = (\sum_{t=1}^{T} X_{t-1}^2)^{1/2}(\hat{\phi}_T - \phi) \xrightarrow{D} \zeta(\gamma)$$

where

$$\begin{aligned} \zeta(\gamma) &= \frac{\int_0^1 \frac{W(r)}{1+br} dW(r)}{(\int_0^1 (\frac{W(r)}{(1+br)})^2 dr)^{1/2}} \\ b &= e^{2\gamma} - 1 \end{aligned}$$

and $\{W(r) : 0 \leq r \leq 1\}$ *is a standard Brownian motion.*

As a special case, $\phi_T = 1 - \frac{1}{T}$ is the case where $\gamma = 1$.

4.2 Large Deviation Principle

Before we present the large deviation principle of the varying coefficient AR(1) process, one proposition must be introduced.

Proposition 4.2.1 *[27] Let* $a, n, \tau \in R$, *and* $M_n(\tau, a, b)$ *denotes the tridiagonal matrix*

$$M_n(\tau, a, b) := \begin{bmatrix} a & b & 0 & \cdots \\ b & a & b & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & b & a & b \\ \cdots & 0 & b & a-\tau \end{bmatrix} \tag{4.2}$$

If $a^2 > 4b^2$, *then*

$$\det M_n(\tau, a, b) = \frac{(\lambda_2 - \tau)\lambda_2^n - (\lambda_1 - \tau)\lambda_1^n}{\lambda_2 - \lambda_1} \tag{4.3}$$

where

1. $2\lambda_1 = a - \sqrt{a^2 - 4b^2}$
2. $2\lambda_2 = a + \sqrt{a^2 - 4b^2}$

We are now ready to state our large deviation result.

Theorem 4.2.1 *For an AR(1) process in equation (4.1) with $\phi_T = 1 - \frac{1}{T}$, the law of the Yule-Walker estimator $\tilde{\phi}_T - \phi_T$ satisfies the large deviation principle with speed T and good rate function*

$$I(r) = \begin{cases} \frac{1}{2}\log(\frac{2}{2+r}) & r \in (-2, 0] \\ +\infty & \textit{otherwise} \end{cases}$$

The law of the Least Squares estimator $\hat{\phi}_T - \phi_T$ satisfies the upper and lower estimate with

$$\limsup_{T\to\infty} \frac{1}{T} \log P(\hat{\phi}_T - \phi_T \le r) = \liminf_{T\to\infty} \frac{1}{T} \log P(\hat{\phi}_T - \phi_T < r) = -I(r)$$

where

$$I(r) = \frac{1}{2}\log(\frac{2}{2+r}), r \in (-1.5, 0]$$

When $r \in (-1.5, 0]^c$, the law of the least squares estimator $\hat{\phi}_T - \phi_T$ satisfies the upper estimate with

$$\limsup_{T\to\infty} \frac{1}{T} \log P(\hat{\phi}_T - \phi_T \le r) = -I(r)$$

where

$$I(r) = \frac{1}{2}\log|-2r-1|$$

Remark: To establish a full large deviation principle for $\hat{\phi}_T - \phi_T$, one needs to have the lower estimate for $r \in (-1.5, 0]^c$.

Proof: First of all, define the vector $\mathbf{X}_T = [X_1(T), ..., X_T(T)]^{\mathrm{T}}$ and the joint p.d.f of vector $\mathbf{X}_T$ is as follows

$$f(\mathbf{x}_T^{\mathrm{T}}) = \frac{1}{(2\pi\sigma^2)^{\frac{T}{2}}} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{T}(x_i - \phi_T x_{i-1})^2\}$$

For the convenience of calculation, we assume that the variance of $\{\varepsilon_t\}$ is equal to one and the covariance matrix of random vector $\mathbf{X}_T$ is

$$\Sigma_T = \begin{bmatrix} 1 & \phi_T & \cdots & \phi_T^T \\ \phi_T & \phi_T^2+1 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \phi_T^T & \cdots & \cdots & \phi_T^{2T} + \phi_T^{2(T-1)} + \cdots + 1 \end{bmatrix}$$

For all $r \in \mathbb{R}$, we have

$$\begin{aligned} P(\tilde{\phi}_T - \phi_T \leq r) &= P(\sum_{t=1}^{T} X_t(T)X_{t-1}(T) - r\sum_{t=0}^{T} X_t^2(T) \leq 0) \\ &= P(\mathbf{X}_T^{\mathrm{T}}(A_T - rB)\mathbf{X}_T \leq 0) \end{aligned}$$

where

1. $\mathbf{X}_T^{\mathrm{T}} = (X_1(T), X_2(T), ..., X_T(T))$

2.

$$A_T := -\frac{1}{2}\begin{bmatrix} 2\phi_T & -1 & 0 & \cdots & \cdots & \cdots & \cdots \\ -1 & 2\phi_T & -1 & \cdots & \cdots & \cdots & \cdots \\ 0 & -1 & 2\phi_T & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & -1 & 2\phi_T & -1 \\ \cdots & \cdots & \cdots & \cdots & 0 & -1 & 2\phi_T \end{bmatrix}$$

3.

$$B := \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & 0 & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 1 & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & 0 & 1 & 0 \\ \cdots & \cdots & \cdots & \cdots & 0 & 0 & 1 \end{bmatrix}$$

By Chernoff inequality, we get

$$\begin{aligned} & P(\mathbf{X}_T^{\mathrm{T}}(A_T - rB)\mathbf{X}_T \leq 0) \\ \leq \quad & \inf_{t<0} E[\exp\{t(\mathbf{X}_T^{\mathrm{T}}(A_T - rB)\mathbf{X}_T)\}] \end{aligned}$$

Now we set

$$Z_T(r) = \mathbf{X}_T^{\mathrm{T}}(A_T - rB)\mathbf{X}_T$$

According to [24] equation 2.8, based on the joint distribution of random vector $\mathbf{X}_T$, we could find the m.g.f. of $Z_T(r)$ which is

$$\begin{aligned} M_{Z_T(r)}(t) =& E[\exp\{tZ_T(r)\}] \\ =& D(T)^{-\frac{1}{2}} \end{aligned}$$

where $D(T)$ is the determinant of matrix D_T

$$D_T := \begin{bmatrix} p(T) & q(T) & 0 & \cdots & \cdots & \cdots & \cdots \\ q(T) & p(T) & q(T) & \cdots & \cdots & \cdots & \cdots \\ 0 & q(T) & p(T) & q(T) & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & q(T) & p(T) & q(T) \\ \cdots & \cdots & \cdots & \cdots & 0 & q(T) & \tau(T) \end{bmatrix}$$

and

$$\begin{aligned} p(T) =& 1 + \phi_T^2 + 2rt + 2\phi_T t \\ q(T) =& -\phi_T - t \\ \tau(T) =& p(T) - \phi_T^2 \end{aligned}$$

Based on above results, we have

1. $\lambda_2(T) \geq 1$

2. $p(T)^2 > 4q(T)^2$

where

1. $\lambda_1(T) = \frac{p(T)-\sqrt{p(T)^2-4q(T)^2}}{2}$

2. $\lambda_2(T) = \frac{p(T)+\sqrt{p(T)^2-4q(T)^2}}{2}$

Then the c.g.f of $Z_T(r)$ is

$$\log E[e^{tZ_T(r)}] = -\frac{1}{2}\log D(T)$$

Therefore, we could get

$$\begin{aligned}\frac{1}{T}\log P(\tilde{\phi}_T - \phi_T \leq r) \leq & \frac{1}{T}\log E[e^{tZ_T(r)}] \\ = & -\frac{1}{2T}\log D(T)\end{aligned}$$

Based on the Proposition 4.2.1 equation 4.2 and 4.3, we could get

$$D(T) = \frac{(\lambda_2(T) - \phi_T^2)\lambda_2(T)^T - (\lambda_1(T) - \phi_T^2)\lambda_1(T)^T}{\lambda_2(T) - \lambda_1(T)}$$

Then we can get

$$\log D(T) = T\log\lambda_2(T) + \log(\frac{\lambda_2(T) - \phi_T^2}{\lambda_2(T) - \lambda_1(T)}) + \log(1 - \eta(T)\rho(T)^T)$$

where

1. $\eta(T) = \frac{\lambda_1(T)-\phi_T^2}{\lambda_2(T)-\phi_T^2}$

2. $\rho(T) = \frac{\lambda_1(T)}{\lambda_2(T)}$

Since

$$\rho(T) < 1$$

Therefore, it can be shown that

$$\lim_{T\to\infty} 1 - \eta(T)\rho(T)^T = 1$$

Furthermore,

$$\frac{D(T-1)}{D(T)} = \frac{1}{\lambda_2(T)} - \left(\frac{\lambda_2(T) - \lambda_1(T)}{\lambda_2(T)\lambda_1(T)}\right)\frac{\eta(T)\rho(T)^T}{1 - \eta(T)\rho(T)^T}$$

When $T \to \infty$,

$$\begin{aligned} p(T) &\to p = 2 + 2rt + 2t \\ q(T) &\to q = -1 - t \end{aligned}$$

and $\lambda_1(T)$ and $\lambda_2(T)$ will converge to the constant if we fix the value of r and t. Therefore, we have

$$\lim_{T\to\infty} \frac{1}{T}\frac{D(T-1)}{D(T)} = 0$$

Therefore,

$$\begin{aligned} \lim_{T\to\infty} \frac{1}{T}\log D(T) &= \lim_{T\to\infty} \frac{1}{T}(T\log\lambda_2(T) + \log(\frac{\lambda_2(T) - \phi_T^2}{\lambda_2(T) - \lambda_1(T)}) + \log(1 - \eta(T)\rho(T)^T)) \\ &= \lim_{T\to\infty} \log\lambda_2(T) \end{aligned}$$

Putting all these together, we obtain

$$\lim_{T\to\infty} \frac{1}{T} \log P(\tilde{\phi}_T - \phi_T \le r) \le \lim_{T\to\infty} -\frac{1}{2T} \log D(T) = L(t)$$

where

$$L(t) = -\frac{1}{2} \log \lambda_2$$

and

$$\lambda_2 = \frac{1}{2}(2 + 2rt + 2t + \sqrt{(2 + 2rt + 2t)^2 - 4(-1 - t)^2})$$

Now we use Girsanov theorem to define a new probability measure Q_T. Let us set

$$L_T(t) = \frac{1}{T} \log E[e^{tZ_T(r)}]$$

Then we could define the new probability measure Q_T such that

$$\frac{dQ_T}{dP} = \exp\{tZ_T(r) - TL_T(t)\}$$

Under the new probability measure Q_T, for a fixed point $y \in (-2, r)$,

$$\begin{aligned}
P(\tilde{\phi}_T - \phi_T < r) =& P(Z_T(r) < 0) \\
=& P(\frac{Z_T(r)}{T} < 0) \\
\ge& P(\frac{Z_T(r)}{T} \in (y - \varepsilon, y + \varepsilon)) \\
=& E_{Q_T}[\exp\{-tZ_T(r) + TL_T(t)\} 1_{\frac{Z_T(r)}{T} \in (y-\varepsilon, y+\varepsilon)}]
\end{aligned}$$

Therefore, we have

$$\frac{1}{T}\log P(\tilde{\phi}_T - \phi_T < r) \geq L_T(t) - ty - |t|\varepsilon + \frac{1}{T}\log Q_T(\frac{Z_T(r)}{T} \in (y-\varepsilon, y+\varepsilon))$$

where

$$y = L'(t)$$

Since

$$L'(t) = -\frac{1}{2}\frac{1}{\lambda_2}(\frac{1}{2}((2r+2) + \frac{1}{\sqrt{p^2-4q^2}}(2p(2r+2)+8q)))$$

It can be shown that $L'(t) \in \mathbb{R}$. Therefore,

$$\liminf_{T\to\infty} \tfrac{1}{T}\log P(\tilde{\phi}_T - \phi_T < r)$$
$$\geq L(t) - ty + \lim_{\varepsilon\to 0}\liminf_{T\to\infty} \tfrac{1}{T}\log Q_T(\tfrac{Z_T(r)}{T} \in (y-\varepsilon, y+\varepsilon))$$
$$\geq -L^*(y) + \lim_{\varepsilon\to 0}\liminf_{T\to\infty} \tfrac{1}{T}\log Q_T(\tfrac{Z_T(r)}{T} \in (y-\varepsilon, y+\varepsilon))$$

where

$$L^*(y) = ty - L(t)$$

If we define $\tilde{L}(\cdot) = L(\cdot + t) - L(t)$, under the new probability measure Q_T, for every $\lambda \in \mathbb{R}$, we have

$$\tilde{L_T}(t) = L_T(\lambda + t) - L_T(t) \to \tilde{L}(t)$$

Define

$$\tilde{L}^*(x) = \sup_{\lambda\in\mathbb{R}}(\lambda x - \tilde{L}(\lambda)) = L^*(x) - tx + L(t)$$

In particular,

$$\limsup_{T\to\infty} \frac{1}{T} \log Q_T(\frac{Z_T(r)}{T} \in (y-\varepsilon, y+\varepsilon)^c) \leq - \inf_{x\in(y-\varepsilon,y+\varepsilon)^c} \tilde{L}^*(x) = -\tilde{L}^*(x_0)$$

for some $x_0 \neq y$. Since y is an exposed point of L^*, and $L^*(y) \geq ty - L(t)$, so we have

$$\tilde{L}^*(x_0) \geq [L^*(x_0) - tx_0] - [L^*(y) - ty] > 0$$

Thus, for every $\varepsilon > 0$,

$$\limsup_{T\to\infty} \frac{1}{T} \log Q_T(\frac{Z_T(r)}{T} \in (y-\varepsilon, y+\varepsilon)^c) < 0$$

and it implies that $Q_T(\frac{Z_T(r)}{T} \in (y-\varepsilon, y+\varepsilon)^c) \to 0$ and hence $Q_T(\frac{Z_T(r)}{T} \in (y-\varepsilon, y+\varepsilon)) \to 1$. According to [2] corollary 5.10, we take the first derivative of $L(t)$ with respect to t.

$$t_r = \frac{r}{r+2}, \text{ for } r \in (-2, 0)$$

such that

$$L'(t_r) = 0$$

And we can find that

$$L''(t_r) > 0$$

Let $t \to t_r$

$$\liminf_{T\to\infty} \frac{1}{T} \log P(\tilde{\phi}_T - \phi_T < r) \geq \lim_{T\to\infty} L_T(t_r) = L(t_r)$$

We also know that

$$\begin{aligned}
& P(\tilde{\phi}_T - \phi_T < r + \varepsilon) \\
= \; & P((\tilde{\phi}_T - \phi_T \le r - \varepsilon) \bigcup (\tilde{\phi}_T - \phi_T \in (r - \varepsilon, r + \varepsilon))) \\
\le \; & 2 \max\{P(\tilde{\phi}_T - \phi_T \le r - \varepsilon), P(\tilde{\phi}_T - \phi_T \in (r - \varepsilon, r + \varepsilon)\}
\end{aligned}$$

Since $P(\tilde{\phi}_T - \phi_T < r + \varepsilon) > P(\tilde{\phi}_T - \phi_T \le r - \varepsilon)$, we can find

$$\begin{aligned}
& \lim_{T \to \infty} \frac{1}{T} \log P(\tilde{\phi}_T - \phi_T < r)) \\
= \; & \lim_{\varepsilon \to 0} \lim_{T \to \infty} \frac{1}{T} \log P(\tilde{\phi}_T - \phi_T < r + \varepsilon)) \\
= \; & \lim_{\varepsilon \to 0} \lim_{T \to \infty} \frac{1}{T} \log P(\tilde{\phi}_T - \phi_T \in (r - \varepsilon, r + \varepsilon))
\end{aligned}$$

We take t_r into λ_2,

$$I(r) = \frac{1}{2} \log(\frac{2}{2 + r})$$

Since $|\tilde{\phi}_T| \le 1$, we have

$$I(r) = +\infty, r \in (-2, 0]^c$$

If we replace the estimator by least squares estimator, we can get the similar result when $r \in (-1.5, 0]$. When $r \in (-1.5, 0]^c$, by the sharp study of the domain, $t = 2r$. Therefore, the law of the least squares estimator $\hat{\phi}_T - \phi_T$ satisfies the upper estimate with

$$I(r) = \log |-2r - 1|$$

For the lower estimate, other arguments are required. The proof is complete.

4.3 Moderate Deviation Principle of Varying Coefficient AR(1) model

Before we can establish the moderate deviation principle of varying coefficient models, we must first define and prove some theorems.

Definition 4.3.1 *Let m be a given positive integer, a sequence $(Z_n)_{n\geq 1}$ of strictly stationary random variables is called $m-$dependent if for every $k \geq 1$ the two collections $\{Z_1, ..., Z_k\}$ and $\{Z_{k+m}, Z_{k+m+1}, ...\}$ are independent.*

Theorem 4.3.1 *[6] Let $(Z_n)_{n\geq 1}$ be a stationary sequence of $m-$dependent random variables taking values in $\mathbb{R}^d$, such that*

$$E[e^{\alpha|Z_1|}] < \infty \text{ for some } \alpha > 0$$

then for all $\lambda \in \mathbb{R}^d$

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{b_n^2} \log E[e^{b_n^2 <\lambda, \frac{1}{\sqrt{n}b_n} \sum_{k=1}^{n}(Z_k - E[Z_k])>}] &= \frac{1}{2} \lim_{n\to\infty} \frac{1}{n} E < \lambda, \sum_{k=1}^{n}(Z_k - E[Z_k]) >^2 \\ &= \frac{1}{2}(E < \lambda, Z_1 >^2 \\ &\quad +2 \sum_{k=2}^{m+1} E < \lambda, Z_1 >< \lambda, Z_k >)\end{aligned}$$

Theorem 4.3.2 *[18] For each $k = 1, 2, ...,$ let $n = n(k)$ and $m = m(k)$ be specified and let $\{X_1^k, ..., X_n^k\}$ be a sequence of strict stationary $m-$dependent random variables with* *zero means such that*

$$\sup_k E[\exp\{\alpha|X_1^k|\}] < \infty \textit{ for some } \alpha > 0$$

In addition, we assume that

1. *there exists a constant* $0 < \sigma^2 < \infty$*, such that*

$$\lim_{k\to\infty} m^{-1}Var[X_1^k + \cdots + X_n^k] = \sigma^2$$

   and

$$\lim_{k\to\infty} m^{-1}\sum_{i=1}^{m} iE[X_1^k X_{i+1}^k] = 0$$

2. *as* $k \to \infty$*(i.e.,* $n \to \infty$*), the moderate deviation scale* (b_n) *is a sequence of positive numbers satisfying,*

$$b_n \to \infty, \frac{b_n m^2}{\sqrt{n}} \to 0$$

Then for any $r > 0$*, we have*

$$\lim_{k\to\infty} \frac{1}{b_n^2} \log P(\frac{1}{b_n\sqrt{n}}|\sum_{i=1}^{n} X_i^k| \geq r) = -\frac{r^2}{2\sigma^2}$$

We must also add some constraints to this model in order to follow the moderate deviation principle of the varying coefficient AR(1) model. Consider the following model:

$$X_t(T) = \phi_T X_{t-1}(T) + \varepsilon_t \tag{4.4}$$

where

1. $t = 1, 2, ..., T$
2. $\phi_T = 1 - \frac{1}{f(T)}$, $f(T) > 0$ and $f(T) \to \infty$ as $T \to \infty$
3. The sequence of $\{\varepsilon_t\}$ is i.i.d. with mean zero and finite variance σ_ε^2

There are two assumptions

Assumption 4.3.1 *1. $\{\varepsilon_t\}$ is a sequence of centered i.i.d. random variables with mean zero and finite variance σ_ε^2 and there exists $\alpha > 0$ such that*

$$E(e^{\alpha|\varepsilon_0|^2}) < \infty$$

2. the moderate deviation scale (b_T) satisfies

$$\begin{aligned} b_T &\to \infty \\ \frac{\sqrt{T}}{b_T} &\to \infty \\ \frac{\sqrt{T}}{b_T f(T)^{\frac{3}{2}}} &\to \infty \end{aligned}$$

Remark: If $f(T)$ is form with T^α for some $\alpha > 0$, the domain of validity of the speed of the MDP will be $1 < b_T < T^{\frac{1}{2}(1-3\alpha)} < \sqrt{T}$.

The moderate deviation principle obtained in this thesis is as follow.

Theorem 4.3.3 *For an AR(1) process in equation (4.4) with* $\phi_T = 1 - \frac{1}{f(T)}$*, under the assumption 4.3.1, the law of the least squares estimator* $|\hat{\phi}_T - \phi_T|$ *satisfies a moderate deviation principle with speed* b_T^2 *and good rate function*

$$I(r) = \begin{cases} \frac{r^2}{2} & r \geq 0 \\ \\ \infty & otherwise \end{cases}$$

Proof: When time T is fixed, the time series $\{X_t(T)\}$ is stationary. Therefore, the distribution of $X_{t+l}(T)X_t(T)$ is the same with $X_l(T)X_0(T)$. So we set

$$C_l := E(X_{i+l}(T)X_i(T)) = E(X_l(T)X_0(T))$$
$$C_{T,l}^* := \frac{\sum_{i=1}^{T-l} X_{i+l}(T)X_i(T)}{T-l}$$

Then we can see that

$$E(C_{T,l}^*) = C_l$$

Let us set

$$Z_{i,l} = X_{i+l}(T)X_i(T) - C_l$$

$$U_{i,l} = \phi_T X_{i+l-1}(T)\varepsilon_i + \phi_T X_{i-1}(T)\varepsilon_{i+l} + \varepsilon_{i+l}\varepsilon_i - \phi_T^l \sigma_\varepsilon^2$$

We could get

$$
\begin{aligned}
Z_{i,l} &= \phi_T^2 Z_{i-1,l} + U_{i,l} \\
C_{T,l}^* - C_l &= (1-R)^{-1}(\bar{U}_{T,l}) + (1-R)^{-1}R(\frac{Z_{0,l} - Z_{T-l,l}}{T-l}) \qquad (4.5)
\end{aligned}
$$

where

1. R is a linear operator:

$$R : x \to \phi_T^2 x$$

2.

$$\bar{U}_{T,l} = \frac{\sum_{i=1}^{T-l} U_{i,l}}{T-l}$$

For the second part $(1-R)^{-1}R(\frac{Z_{0,l}-Z_{T-l,l}}{T-l})$ in equation (4.5), we have the following result

$$
\begin{aligned}
P(|Z_{0,l} - Z_{T-l,l}| > \frac{rb_T\sqrt{T-l}}{\sqrt{1-\phi_T^2}}) &= P(|X_l(T)X_0(T) - X_T(T)X_{T-l}(T)| > \frac{rb_T\sqrt{T-l}}{\sqrt{1-\phi_T^2}}) \\
&\leq 2P(|X_l(T)X_0(T)| > \frac{rb_T\sqrt{T-l}}{2\sqrt{1-\phi_T^2}}) \\
&\leq 4P(|X_0(T)|^2 > \frac{rb_T\sqrt{T-l}}{2\sqrt{1-\phi_T^2}})
\end{aligned}
$$

Since

$$|X_0(T)|^2 = |\sum_{p=0}^{\infty} \phi_T^p \varepsilon_{-p}|^2 \leq (\sum_{p=0}^{\infty} |\phi_T^p||\varepsilon_{-p}|)^2 = K_{\phi_T}^2 (\sum_{p=0}^{\infty} \frac{|\phi_T^p|}{K_{\phi_T}} |\varepsilon_{-p}|)^2$$

where $K_{\phi_T} = f(T)$ and

$$\begin{aligned} P(|X_0(T)|^2 > \frac{rb_T\sqrt{T-l}}{2\sqrt{1-\phi_T^2}}) &\leq P(K_{\phi_T}^2(\sum_{p=0}^{\infty}\frac{|\phi_T^p|}{K_{\phi_T}}|\varepsilon_{-p}|)^2 > \frac{rb_T\sqrt{T-l}}{2\sqrt{1-\phi_T^2}})) \\ &\leq P(K_{\phi_T}^2\sum_{p=0}^{\infty}\frac{|\phi_T^p|}{K_{\phi_T}}|\varepsilon_{-p}|^2 > \frac{rb_T\sqrt{T-l}}{2\sqrt{1-\phi_T^2}})) \\ &\leq P(\sum_{p=0}^{\infty}\frac{|\phi_T^p|}{K_{\phi_T}}\alpha|\varepsilon_{-p}|^2 > \frac{\alpha}{K_{\phi_T}^2}\frac{rb_T\sqrt{T-l}}{2\sqrt{1-\phi_T^2}}) \\ &\leq \exp\{-\frac{\alpha}{K_{\phi_T}^2}\frac{rb_T\sqrt{T-l}}{2\sqrt{1-\phi_T^2}}\}\prod_{p=0}^{\infty}E((e^{\alpha|\varepsilon_0|^2})^{\frac{|\phi_T|^p}{K_{\phi_T}}}) \\ &\leq \exp\{-\frac{\alpha}{K_{\phi_T}^2}\frac{rb_T\sqrt{T-l}}{2\sqrt{1-\phi_T^2}}\}E(e^{\alpha|\varepsilon_0|^2}) \end{aligned}$$

From above results, we could get

$$P(|Z_{0,l} - Z_{T-l,l}| > \frac{rb_T\sqrt{T-l}}{\sqrt{1-\phi_T^2}}) \leq 4\exp\{-\alpha\frac{rb_T\sqrt{T-l}}{2K_{\phi_T}^2\sqrt{1-\phi_T^2}}\}E(e^{\alpha|\varepsilon_0|^2})$$

Since $(1-R)^{-1}$ is the linear operator, based on the assumption 4.3.1, we could get

$$\lim_{T\to\infty}\sup\frac{1}{b_{T-l}^2}\log P(\frac{\sqrt{T-l}\sqrt{1-\phi_T^2}}{b_{T-l}}\frac{|(1-R)^{-1}R(Z_{0,l}-Z_{T-l,l})|}{T-l} > r) = -\infty$$

which means that $Z_{0,l}$ and $Z_{T-l,l}$ are equal in probability, as $T\to\infty$.

Secondly, for all $T \geq 1$, $0 \leq l \leq M$, $1 \leq i \leq T-l$, $m \geq 2M$ and $m \geq l$, set

$$X_{i-1,m}(T) = \varepsilon_{i-1} + \phi_T\varepsilon_{i-2} + ... + \phi_T^{m-2}\varepsilon_{i-m+1} = \sum_{j=0}^{m-2}\phi_T^j\varepsilon_{i-1-j}$$

and

$$\begin{aligned}U_{i,l,m}(T) &= \phi_T X_{i+l-1,m}(T)\varepsilon_i + \phi_T X_{i-1.m}(T)\varepsilon_{i+l} + \varepsilon_{i+l}\varepsilon_i - \phi_T^l \sigma_\varepsilon^2 \\ &= \sum_{j=1}^{m-1} \phi_T^j \varepsilon_{i+l-j}\varepsilon_i + \sum_{j=1}^{m-1} \phi_T^j \varepsilon_{i+l}\varepsilon_{i-j} + \varepsilon_{i+l}\varepsilon_i - \phi_T^l \sigma_\varepsilon^2\end{aligned}$$

For any $0 \le l \le M$ and $1 \le i \le T - l$, it is easy to show that

$$E[U_{i,l,m}(T)] = 0$$

Since $E[U_{i,0,m}(T)|\mathcal{F}_{i-1}] = 0$ and $U_{j,0,m}(T)$ is measurable with respect to $\mathcal{F}_{i-1}$, when $i \neq j$, we can get

$$E[U_{i,0,m}(T)U_{j,0,m}(T)] = E[U_{i,0,m}(T)E(U_{j,0,m}(T)|\mathcal{F}_{i-1})] = 0$$

When $l \neq 0$ and $i > j$, let

$$\begin{aligned}\delta_{1,i,l} &= \sum_{j=1}^{m-1} \phi_T^j \varepsilon_{i+l-j}\varepsilon_k \\ \delta_{2,i,l} &= \sum_{j=1}^{m-1} \phi_T^j \varepsilon_{i+l}\varepsilon_{i-j}\end{aligned}$$

It can be shown that

$$E[U_{i,l,m}(T)|\mathcal{F}_{i+l-1}] = \delta_{1,i,l} - \phi_T^l E[\varepsilon_0^2]$$

therefore, we have

$$E[U_{i,l,m}(T)U_{j,l,m}(T)] = E[U_{j,l,m}(\delta_{1,i,l} - \phi_T^l E[\varepsilon_0^2])] = E[U_{j,l,m}(T)\delta_{1,i,l}]$$

Let $A_1 = \{j + l > i\}$, $A_2 = \{j + l = i\}$. Therefore, we have

$$\begin{aligned}
E[\delta_{1,i,l}\delta_{1,j,l}] =& (\phi_T^l E[\varepsilon_0^2])^2(1 + 1_{A_1}) \\
E[\delta_{1,i,l}\delta_{2,j,l}] =& (E[\varepsilon_0^2])^2 1_{A_2} \sum_{q=1}^{m-1-2l} \phi_T^{2q+2l} \\
E[\delta_{1,i,l}\varepsilon_{j+l}\varepsilon_j] =& (\phi_T^l E[\varepsilon_0^2])^2 1_{A_2} \\
E[\delta_{1,i,l}]\phi_T^l E[\varepsilon_0^2] =& (\phi_T^l E[\varepsilon_0^2])^2
\end{aligned}$$

Finally,

$$E[U_{i,l,m}(T)U_{j,l,m}(T)] = \theta_T^{2l}(E[\varepsilon_0^2])^2(1_{A_1} + \sum_{q=0}^{m-1-2l} \theta_T^{2q} 1_{A_2})$$

Let $m := m(T)$ denote the subsequence of T such that $m(1 - \theta_T) \to \infty$ as $n \to \infty$. When $l \neq 0$,

$$\lim_{n\to\infty} \frac{1 - \theta_T^2}{m} \sum_{i=1}^{m} iE[U_{1,l,m}(T)U_{i+1,l,m}(T)] = 0$$

and

$$\begin{aligned}
Var[U_{1,l,m}(T)+\cdots+U_{m,l,m}(T)] &= \sum_{i=1}^{m} E[U_{i,l,m}(T)^2] \\
&\quad +2\sum_{i=1}^{m-1}\sum_{q=i+1}^{m} E[U_{i,l,m}(T)U_{q,l,m}(T)] \\
&= \sum_{i=1}^{m} E[U_{i,l,m}(T)^2] \\
&\quad +2\sum_{i=1}^{m-l}\sum_{q=i+1}^{i+l} E[U_{i,l,m}(T)U_{q,l,m}(T)] \\
&= m\theta_T^{2l}E[\varepsilon_0^4]+(m+[2(m-l)l-2m]\theta_T^{2l})(E[\varepsilon_0^2])^2 \\
&\quad +(2m\sum_{j=1}^{m-1}\theta_T^{2j}+2(m-l)\theta_T^{2l}\sum_{j=1}^{m-1-2l}\theta_T^{2j})(E[\varepsilon_0^2])^2
\end{aligned}$$

Therefore, when $T \to \infty$, we have

$$\lim_{T\to\infty}\frac{1-\theta_T^2}{m}Var[U_{1,l,m}(T)+\cdots+U_{m,l,m}(T)] = 4(E[\varepsilon_0^2])^2$$

When $l=0$, it can be shown that

$$\lim_{T\to\infty}\frac{1-\theta_T^2}{m}\sum_{i=1}^{m} iE[U_{1,0,m}(T)U_{i+1,0,m}(T)] = 0$$

and

$$\begin{aligned}
Var[U_{1,0,m}(T)+\cdots+U_{m,0,m}(T)] &= \sum_{i=1}^{m} E[U_{i,0,m}(T)^2] \\
&= mE[\varepsilon_0^4]+(E[\varepsilon_0^2])^2[4\sum_{j=1}^{m-1}\theta_T^{2j}-1]
\end{aligned}$$

Therefore, we have

$$\lim_{T\to\infty}\frac{1-\theta_T^2}{m}Var[U_{1,0,m}(T)+\cdots+U_{m,0,m}(T)]=4(E[\varepsilon_0^2])^2$$

Let

$$\overrightarrow{U}_{i,m}=(U_{i,l,m}(T))_{0\le l\le T}$$

Therefore, we could find that $\{\overrightarrow{U}_{i,m}\}_{i\ge 1}$ is a strictly stationary sequence with $(T+m)$-dependent structure and we could set

$$\bar{U}_{T-l,l,m}=\frac{1}{T-l}\sum_{i=1}^{T-l}U_{i,l,m}(T)\text{ and }Q_{T-l,l,m}=(\bar{U}_{T-l,l,m})_{0\le l\le T}$$

The next step is to show that the process $\frac{\sqrt{T-l}}{b_{T-l}}\bar{U}_{T-l,l,m}$ is b_{T-l}^2-exponentially good approximation of the sequence $\{\frac{\sqrt{T-l}}{b_{T-l}}\bar{U}_{T-l,l}\}$.

For all $p\ge 0$, $i\ge 1$, we set

$$W_{i,p}=\frac{\phi_T^p}{|\phi_T^p|}\varepsilon_{i-p}\varepsilon_i$$

Lemma 4.3.1 *[17] If Assumption 4.3.1 condition 1 holds, then, there exists $\alpha_0>0$ and $\beta_0>0$ such that, for all $p\ge 1$, $T\ge 1$ and $t\ge 0$,*

$$P(\max_{j\le T}|\sum_{k=1}^{j}W_{i,p}|\ge t)\le 36\exp\{-\frac{t^2}{\alpha_0 T+\beta_0 t}\}$$

Since

$$
\begin{aligned}
|\sum_{i=1}^{T-l}(U_{i,l}(T)-U_{i,l,m}(T))| =& |\sum_{i=1}^{T-l}[\phi_T(X_{i+l-1}(T)-X_{i+l-1,m}(T))\varepsilon_i \\
&+\phi_T(X_{i-1}(T)-X_{i-1,m}(T))\varepsilon_{i+l}]| \\
\leq& |\sum_{i=1}^{T-l}[\phi_T(X_{i+l-1}(T)-X_{i+l-1,m}(T))\varepsilon_i]| \\
&+|\sum_{i=1}^{T-l}[\phi_T(X_{i-1}(T)-X_{i-1,m}(T))\varepsilon_{i+l}]|
\end{aligned}
$$

and

$$
X_{i+l-1}(T)-X_{i+l-1,m}(T)=\sum_{p=m-1}^{\infty}\phi_T^p\varepsilon_{i+l-1-p}=\phi_T^{m-1}(\sum_{p=0}^{\infty}\phi_T^p\varepsilon_{i+l-m-p})
$$

therefore

$$
\begin{aligned}
|\sum_{i=1}^{T-l}\phi_T(X_{i+l-1}(T)-X_{i+l-1,m}(T)\varepsilon_i| &= |\sum_{p=0}^{\infty}\sum_{i=1}^{T-l}\phi_T^m(\phi_T^p\varepsilon_{i+l-m-p}\varepsilon_i)| \\
&\leq |\phi_T^m|\sum_{p=0}^{\infty}|\phi_T^p||\sum_{i=1}^{T-l}W_{i,m+p-l}|
\end{aligned}
$$

Now we set

$$
K_1:=\sum_{p=0}^{\infty}(p+1)|\phi_T^p|\leq\infty
$$

and

$$
t_{m,p}(r)=\frac{r(p+1)}{2K_1|\phi_T^m|}
$$

Based on Lemma 4.3.1, we could get

$$
\begin{aligned}
&P(\max_{j\leq T-l}|\sum_{i=1}^{j}\phi_T(X_{i+l-1}(T)-X_{i+l-1,m}(T))\varepsilon_i|>\frac{rb_{T-l}\sqrt{T-l}}{2\sqrt{1-\phi_T^2}})\\
\leq &P(\sum_{p=0}^{\infty}(p+1)\frac{|\phi_T^p|}{p+1}\max_{j\leq T-l}|\sum_{i=1}^{T-l}W_{i,m+p-l}|>\sum_{p=0}^{\infty}(p+1)|\phi_T^p|\frac{rb_{T-l}\sqrt{T-l}}{2K_1|\phi_T^m|\sqrt{1-\phi_T^2}})\\
\leq &\sum_{p=0}^{\infty}P(\max_{j\leq T-l}|\sum_{i=1}^{j}W_{i,m+p-l}|>\frac{(p+1)rb_{T-l}\sqrt{T-l}}{2K_1|\phi_T^m|\sqrt{1-\phi_T^2}})\\
\leq &36\sum_{p=0}^{\infty}\exp\{-\frac{b_{T-l}^2t_{m,p}^2(r)}{\alpha_0+\beta_0t_{m,p}(r)\frac{b_{T-l}}{\sqrt{T-l}}}\}
\end{aligned}
$$

But by the assumption of b_{T-l}, there exists constants $N\in N^+$, $A,B>0$, such that for all $T-l\geq N$, $m\geq 1$ and $l\geq 0$, $\frac{\sqrt{T-l}}{b_{T-l}}\geq 1$, and we could get

$$
\frac{t_{m,p}^2(r)}{\alpha_0+\beta_0t_{m,p}(r)\frac{b_{T-l}}{\sqrt{T-l}}}\geq c(r)\frac{p+1}{|\phi_T^m|\sqrt{1-\phi_T^2}}\text{ and }c(r):=\frac{r^2}{Ar+B}
$$

Hence, we could get

$$
\begin{aligned}
&P(\max_{j\leq T-l}|\sum_{i=1}^{j}\phi_T(X_{i+l-1}(T)-X_{i+l-1,m}(T))\varepsilon_i|>\frac{rb_{T-l}\sqrt{T-l}}{2\sqrt{1-\phi_T^2}})\\
\leq\quad &36\sum_{p=0}^{\infty}\exp\{-b_{T-l}^2\frac{c(r)}{|\phi_T^m|\sqrt{1-\phi_T^2}}(p+1)\}
\end{aligned}
$$

So combine above results, we could get

$$
\limsup_{m\to\infty}\limsup_{T\to\infty}\frac{1}{b_{T-l}^2}\log P(\frac{\sqrt{T-l}\sqrt{1-\phi_T^2}}{b_{T-l}}|\bar{U}_{n,l}-\bar{U}_{n,l.m}|>r)=-\infty
$$

which means that the process $\frac{\sqrt{T-l}\sqrt{1-\phi_T^2}}{b_{T-l}}\bar{U}_{T-l,l,m}$ is b_{T-l}^2-exponentially good approximation of the sequence $\{\frac{\sqrt{T-l}\sqrt{1-\phi_T^2}}{b_{T-l}}\bar{U}_{T-l,l}\}$. Based on theorem 4.3.2, we can show that $P(\frac{\sqrt{T-l}\sqrt{1-\phi_T^2}}{b_{T-l}}Q_{T-l,l,m} \in A)$ satisfies the large deviation principle with speed b_{T-l}^2. Since $Q_{T-l} = \{\bar{U}_{T-l,l}\}$ can be approximated by $Q_{T-l,l,m}$, it implies that $P(\frac{\sqrt{T-l}\sqrt{1-\phi_T^2}}{b_{T-l}}Q_{T-l} \in A)$ satisfies the large deviation principle with speed b_{T-l}^2. Therefore, it means that $P(\frac{\sqrt{T-l}(1-\phi_T^2)^{\frac{3}{2}}}{b_{T-l}}(C_{T,l}^* - C_l) \in A)$ satisfies the large deviation principle.

Secondly, we will show that $\frac{\sqrt{T}}{b_T\sqrt{1-\phi_T^2}}(\hat{\phi}_T - \phi_T)$ satisfies the moderate deviation principle. Let us set

$$\begin{aligned} r_T &= \frac{\sqrt{T}}{b_T\sqrt{1-\phi_T^2}}(\hat{\phi}_T - \phi_T) \\ R_T &= \frac{1}{\sqrt{1-\phi_T^2}\sqrt{T}b_T}\frac{\sum_{i=1}^T(X_i(T)X_{i-1}(T) - \phi_T X_{i-1}(T)^2)}{E(X_0(T)^2)} \end{aligned}$$

The main idea is to show that r_T and R_T are equal in probability. We can show that

$$\begin{aligned} r_T - R_T =& \frac{\sqrt{T}}{b_T\sqrt{1-\phi_T^2}}(\hat{\phi}_T - \phi_T) - \frac{1}{\sqrt{1-\phi_T^2}\sqrt{T}b_T}\frac{\sum_{i=1}^T(X_i(T)X_{i-1}(T) - \phi_T X_{i-1}(T)^2)}{E(X_0(T)^2)} \\ =& \frac{1}{\sqrt{1-\phi_T^2}b_T\sqrt{T}}[\frac{T\sum_{i=1}^T(X_i(T)X_{i-1}(T) - \phi_T X_{i-1}(T)^2)}{\sum_{i=1}^T X_{i-1}(T)^2} \\ &- \frac{\sum_{i=1}^T(X_i(T)X_{i-1}(T) - \phi_T X_{i-1}(T)^2)}{E(X_0(T)^2)}] \\ =& \frac{1}{\sqrt{1-\phi_T^2}b_T\sqrt{T}}[\frac{\sum_{i=1}^T(X_i(T)X_{i-1}(T) - \phi_T X_{i-1}(T)^2)}{E(X_0(T))^2}][E(X_0(T)^2) - \frac{1}{T}\sum_{i=1}^T X_{i-1}(T)^2] \\ &\times \frac{T}{\sum_{i=1}^T X_{i-1}(T)^2} \end{aligned}$$

Therefore

$$\begin{aligned}P(|r_T - R_T| > r) \leq & P(|\frac{1}{\sqrt{1-\phi_T^2} b_T \sqrt{T}} \frac{\sum_{i=1}^T (X_i(T)X_{i-1}(T) - \phi_T X_{i-1}(T)^2)}{E(X_0(T)^2)}| \geq L\sqrt{\delta r}) \\ & + P(|E(X_0(T)^2) - \frac{1}{T}\sum_{i=1}^T X_{i-1}(T)^2| \geq \frac{\sqrt{\delta r}}{L}) \\ & + P(|\frac{T}{\sum_{i=1}^T X_{i-1}(T)^2}| > \frac{1}{\delta})\end{aligned}$$

For δ, r sufficiently small but fixed, the first term at the right hand side above is negligible by the moderate deviation principle of R_T by letting $L \to \infty$. The second and third terms are negligible by the moderate deviation principle of $\frac{(1-\phi_T^2)^{\frac{3}{2}}}{\sqrt{T} b_T} \sum_{i=1}^T (X_i(T)^2 - E[X_0^2])$. Since

$$\frac{\sum_{i=1}^T (X_i(T)X_{i-1}(T) - \phi_T X_{i-1}(T)^2)}{E(X_0(T)^2)}$$

is a martingale with stationary differences and R_n satisfies the MDP with the speed b_T^2 and with the rate function

$$I(r) = \sup_{\lambda} (\lambda r - \frac{1}{2} E(\lambda \frac{X_1(T)X_0(T) - \phi_T X_0(T)^2}{\sqrt{1-\phi_T^2} E(X_0(T)^2)}))^2$$

Since

$$X_1(T)X_0(T) - \phi_T X_0(T)^2 = X_0(T)(X_1(T) - \phi_T X_0(T)) = X_0(T)\varepsilon_1$$

We then can calculate

$$\begin{aligned} E(\lambda\frac{X_1(T)X_0(T)-\phi_T X_0(T)^2}{\sqrt{1-\phi_T^2}E(X_0(T)^2)}))^2 &= E(\lambda\frac{X_0(T)\varepsilon_1}{\sqrt{1-\phi_T^2}E(X_0(T)^2)})^2 \\ &= \lambda^2\frac{1}{1-\phi_T^2}E(\frac{X_0(T)\varepsilon_1}{E(X_0(T)^2)})^2 \\ &= \lambda^2 \end{aligned}$$

Therefore, we could get the rate function as follows

$$I(r)=\frac{r^2}{2}$$

When $r<0$, $|\hat{\phi}_T-\phi_T|>r$ is always true. Therefore,

$$I(r)=+\infty$$

and the proof is complete.

Chapter 5

Applications of Varying Coefficient Models

As we can see in Chapter 3, the ADF and KPSS tests use different statistics to determine a time series' stationarity. Some applications of these two tests will be made in this chapter. After confirming the stochastic trend of each time series, we perform cointegration tests to see if the linear combinations of these series are stationary.

5.1 Temperature Series

Temperatures in many cities have changed significantly over the last 100 years as a result of the greenhouse gas effect. Analysing and forecasting various countries' or cities' average temperature series will assist us in warning about and controlling extreme weather events. Detecting trends in time series and assessing their magnitude and statistical significance is also an important task in geophysical research.

Woodward and Gray[25, 26] investigated whether a stochastic trend exists in global average temperature series. Fatichi et al[12] used the KPSS method to examine temperature time series for trend stationery and difference stationary behaviours. They confirmed an increase in uncertainty when there are pronounced stochastic trends in the data. In 2016, Chang et al.[5] also analyzed the time series of global average temperature anomaly distributions to identify and estimate persistent features in climate change. They used a formal test for functional unit roots in the distributions' time series.

We retrieved average annual temperature data from the U.S. Department of Energy website[15] and Berkelery Earth website ($http://berkeleyearth.org/$). The data can be divided into four categories: global data, hemispheric data, continental data, and data for individual cities. We performed a structural analysis on the aforementioned datasets and discovered that the majority of them should follow a varying coefficient model.

5.1.1 Global Annual Temperature Series

We examined global annual average temperature data from 1900 to 2010. The time series contains 111 observed values, which are shown in Figure 5.1.

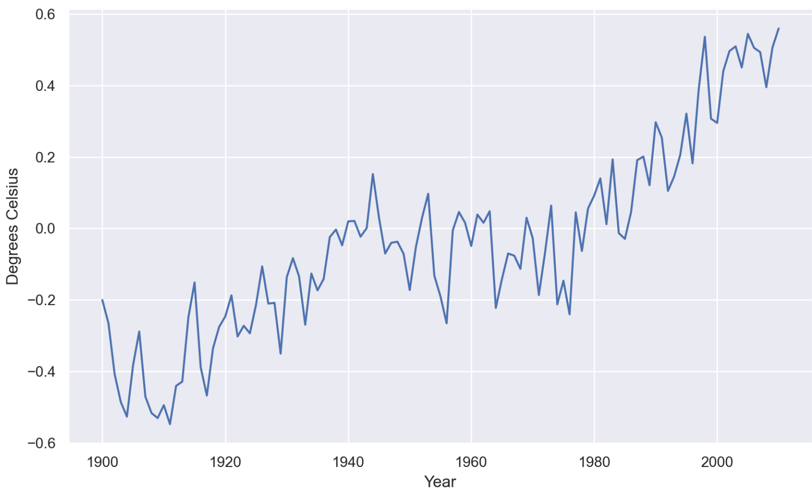


Figure 5.1: Global Annual Average Temperature Time Series

Figure 5.1 clearly shows that the series has an increasing trend. The Mann-Kendall test also confirms this result. It satisfies the third situation of the ADF test, which is that the time series contains constant and time trend terms, according to the unit root test introduced in Chapter 3. Based on this result, we employ two distinct methods to test for the stochastic trend in this series: ADF and KPSS. The outcome is shown in Table 5.1.

Table 5.1: Stochastic Trend Test of Global Temperature Series

No.	Region	Period	ADF result	P-Value (ADF)	KPSS result	P-Value (KPSS)
1	Global	1900-2000	Unit Root	0.3572	Unit Root	0.0293
2	Global	1900-2010	Unit Root	0.8305	Unit Root	0.0148

We used two different time intervals for testing to confirm whether this time series will change with time and whether its stationarity will change as well. Table 5.1 shows that by varying the time interval, the ADF and KPSS tests can provide the same conclusion for the global average temperature series, i.e., it has a unit root.

5.1.2 Hemispheric Annual Temperature Series

We also examined annual average temperature data from the Northern and Southern Hemispheres from 1900 to 2010. Each time series contains 111 observed values, which are shown in Figure 5.2.

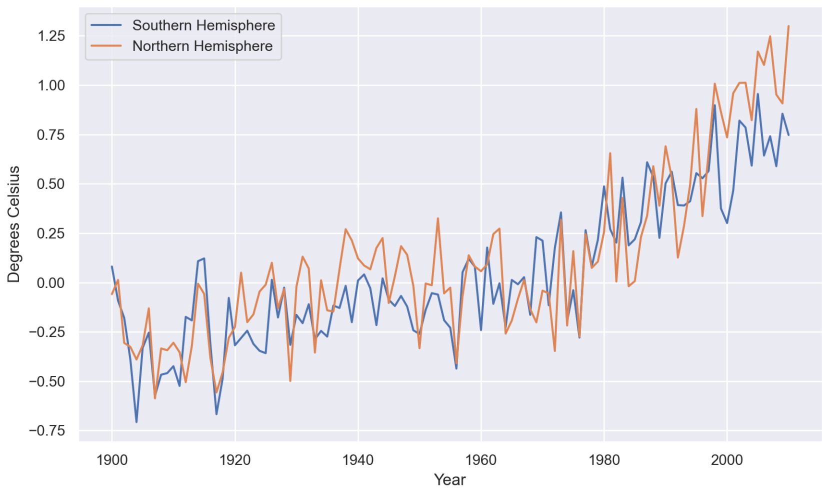


FIGURE 5.2: Hemispheric Annual Average Temperature

Figure 5.2 clearly shows that they both have an increasing trend. The Mann-Kendall test also confirms this result. As a result, the constant and time trend terms should be included in the ARMA models for these two time series. Based on this result, we used two different methods, ADF and KPSS, to test for the stochastic trend in these two series. The outcome is shown in Table 5.2.

Table 5.2: Stochastic Trend Test of Hemispheric Temperature Series

No.	Region	ADF result	P-Value (ADF)	KPSS result	P-Value (KPSS)
1	Southern Hemisphere	Trend Stationary	1.39×10^{-9}	Unit Root	0.01
2	Northern Hemisphere	Unit Root	0.3269	Unit Root	0.01831

Table 5.2 shows that ADF and KPSS tests could give the same conclusion for the Northern Hemisphere's average temperature series, which has a unit root. However, the test result for the Southern Hemisphere series is different. The ADF test method demonstrates that the Southern Hemisphere's annual average temperature series does not contain a unit root. But the KPSS method reveals that the series has a unit root. We ran the tests again for the series with data from various time intervals. The ADF and KPSS tests first assisted us in identifying the stochastic trend in these series. The ARMA models for these series was then chosen using the AIC criterion. These series can be represented by ARMA models under trend stationary conditions using the AIC criterion. The ordinary least squares method was used to estimate the AR component coefficients. Table 5.3 displays the regression result.

Table 5.3: Southern Hemisphere Temperature Series Estimation Results

No.	Region	Time Period	Model	Estimated Value of AR coefficient
1	Southern Hemisphere	1900-2000	ARMA(1,2)	0.9883
2	Southern Hemisphere	1900-2001	ARMA(1,2)	0.9887
3	Southern Hemisphere	1900-2002	ARMA(1,2)	0.9910
4	Southern Hemisphere	1900-2003	ARMA(1,2)	0.9921
5	Southern Hemisphere	1900-2004	ARMA(1,2)	0.9916

The results in Table 5.3 show that increasing the number of observations from 101 to 105 brings the series' AR component coefficient closer to one. When the amount of observed data is small, however, the coefficient of the AR component is strictly less than one; thus, ADF and KPSS tests will show that the series is a trend stationary series with no unit root. When we increase the number of observations, the estimated coefficient value of the AR component approaches one but remains less than one. As a result, based on the ADF method, the test still suggests that the time series does not contain a unit root, which explains why the ADF test's p-value is less than 0.05. The KPSS method demonstrates that the series contains a stochastic trend. Based on the above analysis, we have enough evidence to believe that the series can be well represented by a varying coefficient model. It will eventually become a unit root process as time passes.

We used linear regression to obtain a long-term equilibrium relationship between the two hemispheric temperature series after confirming that both series contain unit roots. The relationship is based on data from 1900 to 2000, with each series containing 101 observations. To test the residual series to see if it is stationary or containing any trend, we used the ADF test method. We discovered that the long-term equilibrium residual between the two hemispheres is a stationary series. Thus, according to the Engle-Granger cointegration testing method, the two hemispheric temperature series are cointegrated. Equation 5.1 depicts their long-term equilibrium relationship.

$$\text{Southern Hemisphere}_t = -0.0472 + 0.7791 \times \text{Northern Hemisphere}_t \tag{5.1}$$

The coefficient in front of the Northern Hemisphere term is a positive number, indicating that the two temperature series have a positive relationship. It is simple to see that when the temperature of one hemisphere rises, the temperature of the other hemisphere rises as well. Between the two time series, the linear correlation coefficient is 0.8390. We also used this long-term equilibrium relationship to predict Southern Hemisphere's temperatures from 2001 to 2010 based on Northern Hemisphere's temperatures observed in the corresponding years, and then compared the predicted values to the actual observed values. Figure 5.3 depicts the final results.

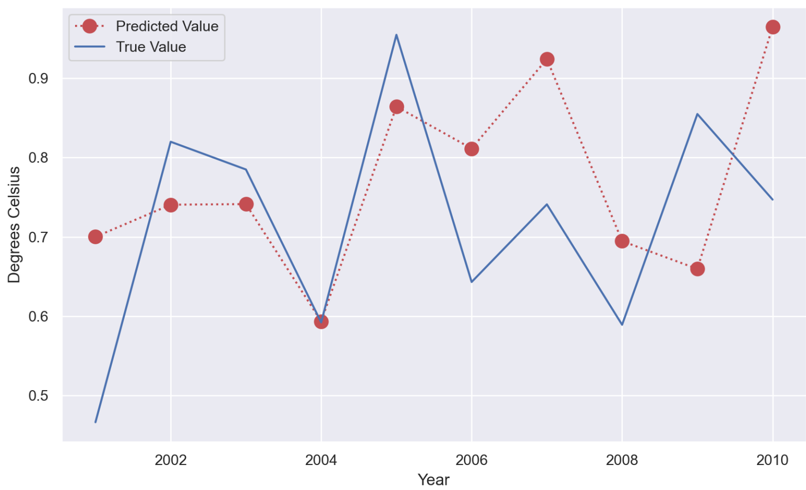
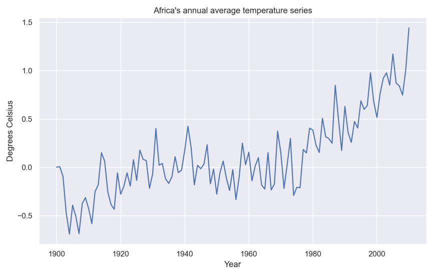


FIGURE 5.3: Predicted Values versus Actual Observed Values

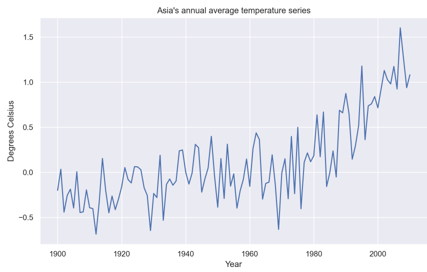
Figure 5.3 shows that the observed temperature fluctuations can be reasonably predicted from year to year. This prediction has a mean squared error of only 0.023.

5.1.3 Continental Annual Average Temperature Series

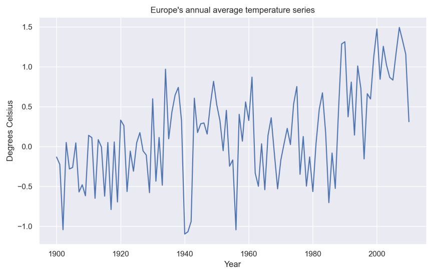
The annual average temperature of each continent is the third object to be studied. We analyzed the annual average temperature of six continents from 1900 to 2010, with 111 observations for each time series. Figure 5.4 depicts the temperature series for each continent.



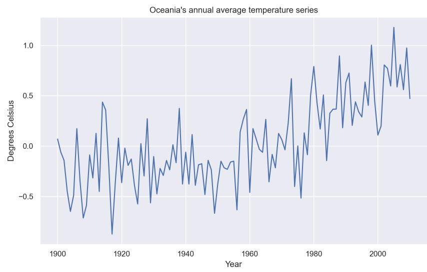
(a) Africa



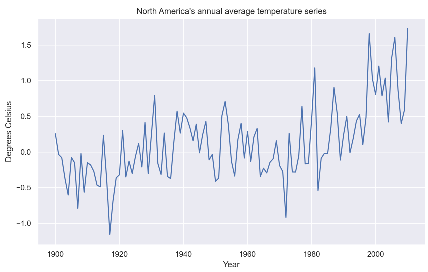
(b) Asia



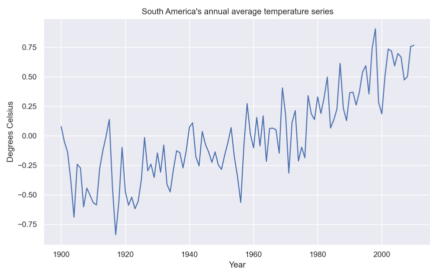
(c) Europe



(d) Oceania



(e) North America



(f) South America

FIGURE 5.4: Continental Annual Average Temperature Series

It is also clear that they all have an increasing trend. Mann-Kendall test confirmed that all of the continental average temperature series contain a deterministic increasing trend. We tested the existence of stochastic trend in the above time series using ADF and KPSS tests. The results are summarised in Table 5.4.

TABLE 5.4: Stochastic Trend Test of Continental Temperature Series

No.	Region	P-Value (ADF)	ADF result	P-Value (KPSS)	KPSS result
1	North America	7.29×10^{-10}	Trend Stationary	0.0451	Unit Root
2	Asia	3.52×10^{-5}	Trend Stationary	0.0124	Unit Root
3	Europe	7.30×10^{-6}	Trend Stationary	0.1	Trend Stationary
4	South America	1.63×10^{-9}	Trend Stationary	0.0163	Unit Root
5	Oceania	0.5771	Unit Root	0.01	Unit Root
6	Africa	0.0223	Trend Stationary	0.0164	Unit Root

As can be seen from Table 5.4, North America, Asia, South America and Africa have different test results when the two different tests were used. So we used the AIC criterion to select the ARMA models for these time series with observations from two different time intervals. North America, Asia and Africa have ARMA models where the AR component is with first order, and estimation results are presented in Table 5.5.

TABLE 5.5: Continental Temperature Series Estimation Result

No.	Region	Time Period	Model	Estimated Value of AR component
1	North America	1900-2000	ARMA(1,2)	0.9573
2	North America	1900-2010	ARMA(1,1)	0.9789
3	Asia	1900-2000	ARMA(1,1)	0.9866
4	Asia	1900-2010	ARMA(1,1)	0.9919
5	Africa	1900-2000	ARMA(1,2)	0.9827
6	Africa	1900-2010	ARMA(1,3)	0.99

As the length of the observation time increases, so does the number of observed data, and the estimated AR coefficient approaches to one. This confirms that the time series in these three regions should follow varying coefficient time series models. And we should be able to confirm that the annual average temperature series in North America, Asia, Oceania, and Africa are nonstationary time series with unit roots. We will not do much analysis for the annual average temperature series in South America because the order of the AR component exceeds 1 as determined by the AIC criterion.

After confirming that the continental series, with the exception of Europe, contain unit roots, we used linear regression to obtain the long-term equilibrium relationships between the continents excluding Europe based on data from 1900 to 2000, with 101 observations for each series. The ADF method is used to test the residual series obtained from the regression analysis. Through testing, we discovered that the long-term equilibrium residuals between eight pairs of continents are stationary series.

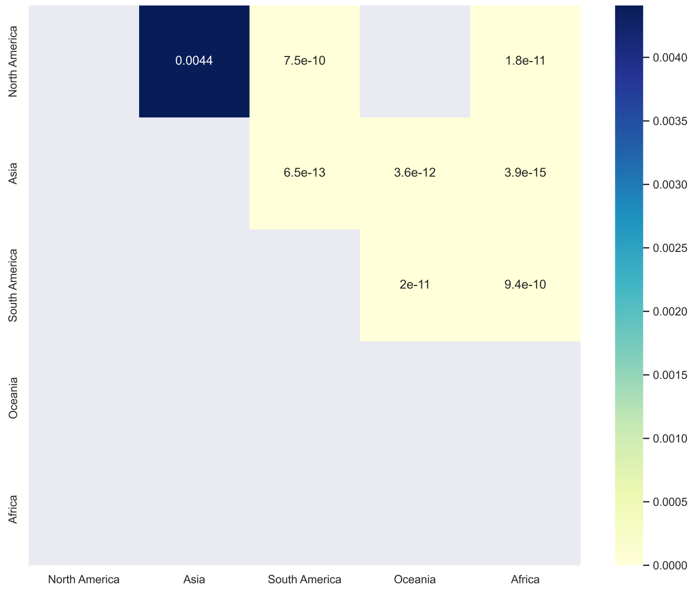


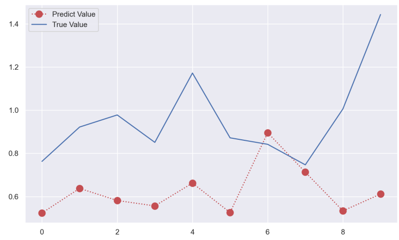
FIGURE 5.5: Continental Temperature Series Cointegration Test Results

We can confirm that there are cointegration relationships between those eight pairs of continents based on the results shown in Figure 5.5. Figure 5.5 depicts the cointegration test statistics as a heatmap, while Table 5.6 lists in detail the long-term equilibrium relationships.

Table 5.6: Long-Term Equilibrium Relationships

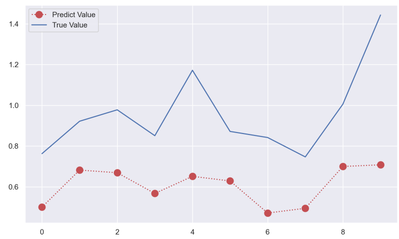
No.	Continents	Relationship
1	Africa and Asia	$\text{Africa}_t = 0.0219 + 0.5451 \times \text{Asia}_t$
2	South America and Asia	$\text{South America}_t = -0.0817 + 0.5693 \times \text{Asia}_t$
3	Oceania and Asia	$\text{Oceania}_t = -0.0438 + 0.6184 \times \text{Asia}_t$
4	Africa and North America	$\text{Africa}_t = 0.0150 + 0.4819 \times \text{North America}_t$
5	Oceania and South America	$\text{Oceania}_t = 0.0287 + 0.8506 \times \text{South America}_t$
6	South America and North America	$\text{South America}_t = -0.0861 + 0.4335 \times \text{North America}_t$
7	Africa and South America	$\text{Africa}_t = 0.0897 + 0.8074 \times \text{South America}_t$
8	Asia and North America	$\text{Asia}_t = 0.0057 + 0.4210 \times \text{North America}_t$

According to Table 5.6, all coefficients in front of the predictor continent are positive. It means that as the temperature of one continent rises, so will the temperature of the other. This demonstrates that the Earth's temperature rise is a global phenomenon, not just a regional phenomenon. The heatmap in Figure 5.5 also shows that there is no cointegration relationship between North America and Oceania, and no cointegration relationship between Oceania and Africa. The residual series of their linear combination is not stationary, as demonstrated by the EG two-step method and confirmed by the ADF test. Afterwards, we used the long-term equilibrium relationships to forecast the temperatures of each continent for the next ten years, from 2001 to 2010, and the results are shown in Figure 5.6.

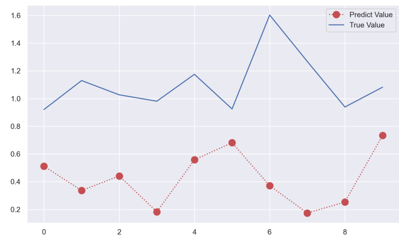


(a) African Prediction based on Asian observation

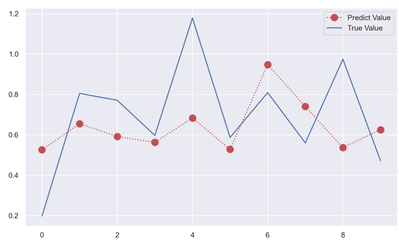
(b) African Prediction based on North American observation



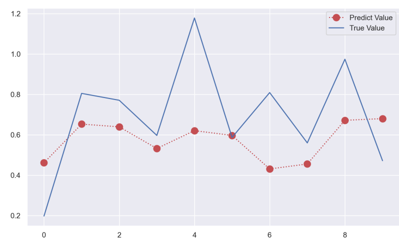
(c) African Prediction based on South American observation



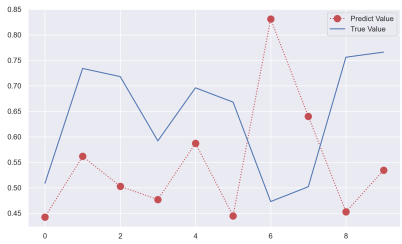
(d) Asian Prediction based on North American observation



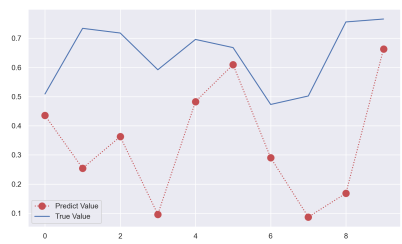
(e) Oceanian Prediction based on Asian observation



(f) Oceanian Prediction based on South American observation



(g) South American Prediction based on Asian observation



(h) South American Prediction based on North American observation

FIGURE 5.6: Continents' Predicted versus Observed Values

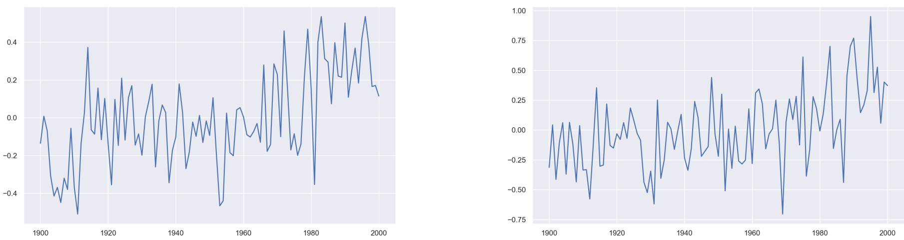
The Mean Squared Errors (MSEs) of these predictions were also calculated and are presented in Table 5.7.

TABLE 5.7: Mean squared Errors of Continents' Temperature Predictions

No.	Region 1	Region 2	MSE Score
1	Africa	Asia	0.1679
2	South America	Asia	0.0448
3	Oceania	Asia	0.0678
4	Africa	North America	0.2521
5	Oceania	South America	0.0713
6	South America	North America	0.1219
7	Africa	South America	0.1466
8	Asia	North America	0.5525

Table 5.7 shows that the majority of the predictions are relatively accurate with MSE

values around 0.1. However, the forecast results for Africa and North America, as well as Asia and North America, are relatively poor when compared to the other sets. Figure 5.7 depicts the residual series of these two sets.



(a) Residual Series of Linear Combination between Africa and North America

(b) Residual Series of Linear Combination between Asia and North America

FIGURE 5.7: Residual Series for the sets of Africa and North America as well as Asia and North America

Figure 5.7 shows that the residual series from the sets of Africa and North America, as well as Asia and North America, have an increasing trend. This demonstrates that the warming trends in Africa and Asia are more prominent than those in North America. This linear growth trend in residuals, however, cannot be confirmed due to a insufficient amount of data. That is why when using the ADF test on the residual series, it still yields a stationary result. As the number of observations grows, the ADF test on the residuals may reveal that the residuals are no longer stationary, and those two sets of long-term equilibrium relationships may not hold.

5.1.4 Individual Cities' Annual Average Temperature Series

Each city's data set includes 111 observations. To test the stationarity of these series, we continue to use the ADF and KPSS methods. It can be demonstrated that 9 cities' annual average temperature series are non-stationary with unit roots, 58 cities' average temperature series are trend stationary, and 31 cities' ADF and KPSS test results differ from each other.

TABLE 5.8: Test Results for Cities Where Both ADF and KPSS Indicate the Existence of Unit Roots

No.	City	P-Value (ADF)	ADF result	P-Value (KPSS)	KPSS result
1	Luanda	0.4090	Unit Root	0.0292	Unit Root
2	Karachi	0.9602	Unit Root	0.0465	Unit Root
3	Dar es SalaamDar	0.2456	Unit Root	0.0179	Unit Root
4	Chengdu	0.1711	Unit Root	0.01	Unit Root
5	Abidjan	0.2292	Unit Root	0.0110	Unit Root
6	Mexico	0.1086	Unit Root	0.0134	Unit Root
7	Kinshasa	0.5015	Unit Root	0.0449	Unit Root
8	Addis Abeba	0.9071	Unit Root	0.01	Unit Root
9	Lagos	0.6670	Unit Root	0.0337	Unit Root

Table 5.9: Test Results for Cities Where Both ADF and KPSS Indicate that the Series are Trend Stationary

No.	City	P-Value (ADF)	ADF result	P-Value (KPSS)	KPSS result
1	Chicago	1.56×10^{-13}	Trend Stationary	0.1	Trend Stationary
2	Berlin	8.80×10^{-8}	Trend Stationary	0.1	Trend Stationary
3	Kiev	0.0019	Trend Stationary	0.1	Trend Stationary
4	Rome	0.0005	Trend Stationary	0.1	Trend Stationary
5	Jiddah	7.55×10^{-10}	Trend Stationary	0.1	Trend Stationary
6	Kanpur	2.79×10^{-6}	Trend Stationary	0.1	Trend Stationary
7	Guangzhou	6.63×10^{-13}	Trend Stationary	0.1	Trend Stationary
8	Moscow	1.95×10^{-7}	Trend Stationary	0.1	Trend Stationary
9	Mashhad	7.36×10^{-6}	Trend Stationary	0.0831	Trend Stationary
10	Casablanca	0.0365	Trend Stationary	0.1	Trend Stationary
11	Nagoya	2.52×10^{-12}	Trend Stationary	0.1	Trend Stationary
12	Seoul	2.57×10^{-14}	Trend Stationary	0.1	Trend Stationary
13	New York	8.32×10^{-15}	Trend Stationary	0.1	Trend Stationary
14	Nagpur	1.69×10^{-8}	Trend Stationary	0.1	Trend Stationary
15	Jakarta	4.99×10^{-12}	Trend Stationary	0.1	Trend Stationary
16	Kabul	3.66×10^{-12}	Trend Stationary	0.1	Trend Stationary
17	Umm Durman	5.41×10^{-6}	Trend Stationary	0.0810	Trend Stationary
18	Paris	0.0046	Trend Stationary	0.1	Trend Stationary
19	Santiago	8.41×10^{-9}	Trend Stationary	0.0820	Trend Stationary
20	Manila	1.79×10^{-12}	Trend Stationary	0.0744	Trend Stationary
21	Montreal	3.22×10^{-15}	Trend Stationary	0.1	Trend Stationary
22	Baghdad	6.84×10^{-13}	Trend Stationary	0.1	Trend Stationary
23	Calcutta	2.95×10^{-5}	Trend Stationary	0.1	Trend Stationary
24	Lima	3.95×10^{-11}	Trend Stationary	0.1	Trend Stationary
25	Tianjin	7.78×10^{-11}	Trend Stationary	0.1	Trend Stationary

No.	City	P-Value (ADF)	ADF result	P-Value (KPSS)	KPSS result
26	Tangshan	1.43×10^{-10}	Trend Stationary	0.0976	Trend Stationary
27	Aleppo	6.41×10^{-11}	Trend Stationary	0.1	Trend Stationary
28	Harare	1.65×10^{-11}	Trend Stationary	0.1	Trend Stationary
29	Santo Domingo	1.30×10^{-9}	Trend Stationary	0.1	Trend Stationary
30	Saint Petersburg	6.45×10^{-121}	Trend Stationary	0.1	Trend Stationary
31	Beijing	7.78×10^{-11}	Trend Stationary	0.1	Trend Stationary
32	Riyadh	2.45×10^{-11}	Trend Stationary	0.1	Trend Stationary
33	Lahore	2.73×10^{-7}	Trend Stationary	0.1	Trend Stationary
34	London	7.88×10^{-11}	Trend Stationary	0.1	Trend Stationary
35	Toronto	2.00×10^{-14}	Trend Stationary	0.1	Trend Stationary
36	Madrid	0.0001	Trend Stationary	0.0696	Trend Stationary
37	Lakhnau	2.79×10^{-6}	Trend Stationary	0.1	Trend Stationary
38	Hyderabad	6.92×10^{-11}	Trend Stationary	0.0708	Trend Stationary
39	Singapore	3.10×10^{-8}	Trend Stationary	0.1	Trend Stationary
40	Rangoon	2.38×10^{-12}	Trend Stationary	0.1	Trend Stationary
41	Istanbul	3.31×10^{-10}	Trend Stationary	0.1	Trend Stationary
42	Delhi	1.32×10^{-9}	Trend Stationary	0.1	Trend Stationary
43	Faisalabad	2.73×10^{-7}	Trend Stationary	0.1	Trend Stationary
44	Taipei	1.46×10^{-11}	Trend Stationary	0.1	Trend Stationary
45	Dhaka	2.20×10^{-11}	Trend Stationary	0.1	Trend Stationary
46	Cape Town	1.84×10^{-7}	Trend Stationary	0.0668	Trend Stationary
47	Jinan	1.75×10^{-11}	Trend Stationary	0.0987	Trend Stationary
48	Ankara	2.20×10^{-10}	Trend Stationary	0.1	Trend Stationary
49	Cali	1.69×10^{-11}	Trend Stationary	0.1	Trend Stationary
50	New Delhi	1.32×10^{-9}	Trend Stationary	0.1	Trend Stationary
51	Jaipur	9.07×10^{-5}	Trend Stationary	0.1	Trend Stationary
52	Tokyo	1.58×10^{-11}	Trend Stationary	0.1	Trend Stationary

No.	City	P-Value (ADF)	ADF result	P-Value (KPSS)	KPSS result
53	Wuhan	6.95×10^{-11}	Trend Stationary	0.0545	Trend Stationary
54	Surabaya	2.96×10^{-14}	Trend Stationary	0.1	Trend Stationary
55	Taiyuan	4.60×10^{-10}	Trend Stationary	0.1	Trend Stationary
56	Bogota	1.67×10^{-10}	Trend Stationary	0.1	Trend Stationary
57	Los Angeles	6.84×10^{-12}	Trend Stationary	0.1	Trend Stationary
58	Durban	1.27×10^{-11}	Trend Stationary	0.0837	Trend Stationary

TABLE 5.10: Test Results for Cities Where ADF and KPSS Test Results are Different

No.	City	P-Value (ADF)	ADF result	P-Value (KPSS)	KPSS result
1	Ibadan	1.37×10^{-6}	Trend Stationary	0.0252	Unit Root
2	Alexandria	2.22×10^{-10}	Trend Stationary	0.0259	Unit Root
3	Madras	2.34×10^{-11}	Trend Stationary	0.0231	Unit Root
4	Dalian	1.10×10^{-12}	Trend Stationary	0.0498	Unit Root
5	Melbourne	1.43×10^{-11}	Trend Stationary	0.0195	Unit Root
6	Belo Horizonte	0.0052	Trend Stationary	0.01	Unit Root
7	Harbin	6.71×10^{-13}	Trend Stationary	0.0148	Unit Root
8	Bangalore	0.0001	Trend Stationary	0.01	Unit Root
9	Izmir	7.75×10^{-9}	Trend Stationary	0.0382	Unit Root
10	Chongqing	6.08×10^{-10}	Trend Stationary	0.01	Unit Root
11	Shenyang	9.54×10^{-13}	Trend Stationary	0.0464	Unit Root
12	Kano	1.61×10^{-8}	Trend Stationary	0.0374	Unit Root
13	Shanghai	3.79×10^{-12}	Trend Stationary	0.03423	Unit Root
14	Nanjing	3.79×10^{-11}	Trend Stationary	0.0368	Unit Root
15	Gizeh	1.89×10^{-11}	Trend Stationary	0.0302	Unit Root
16	Bangkok	3.72×10^{-10}	Trend Stationary	0.0374	Unit Root

No.	City	P-Value (ADF)	ADF result	P-Value (KPSS)	KPSS result
17	Sao Paulo	6.04×10^{-17}	Trend Stationary	0.0228	Unit Root
18	Xian	4.72×10^{-8}	Trend Stationary	0.0464	Unit Root
19	Surat	1.42×10^{-12}	Trend Stationary	0.0426	Unit Root
20	Ho Chi Minh City	4.05×10^{-8}	Trend Stationary	0.0178	Unit Root
21	Salvador	1.73×10^{-10}	Trend Stationary	0.01	Unit Root
22	Sydney	7.02×10^{-12}	Trend Stationary	0.0137	Unit Root
23	Fortaleza	2.03×10^{-10}	Trend Stationary	0.0196	Unit Root
24	Cairo	1.89×10^{-11}	Trend Stationary	0.0302	Unit Root
25	Dakar	5.39×10^{-10}	Trend Stationary	0.01	Unit Root
26	Rio de Janeiro	0.0001	Trend Stationary	0.0244	Unit Root
27	Bombay	6.14×10^{-12}	Trend Stationary	0.0317	Unit Root
28	Changchun	9.73×10^{-13}	Trend Stationary	0.0209	Unit Root
29	Brasilia	1.22×10^{-10}	Trend Stationary	0.01	Unit Root
30	Mogadishu	1.30×10^{-7}	Trend Stationary	0.0242	Unit Root
31	Ahmadabad	4.38×10^{-7}	Trend Stationary	0.04359	Unit Root

We used the same approach to examine the stationarity characteristics of the series for the 31 cities in Table 5.10. We can confirm that they all adhere to the ARMA model. Due to the large number of cities in Table 5.10 and the more complex AR components of some time series, we only selected 8 cities from China and present the regression results in Table 5.11.

Table 5.11: Chinese Cities' Temperature Series Estimation Results

No.	City	Time Period	Model	Estimated Coefficient of the AR component
1	Dalian	1900-2000	ARMA(1,1)	0.9809
2	Dalian	1900-2010	ARMA(1,1)	0.9825
3	Harbin	1900-2000	ARMA(1,1)	0.9786
4	Harbin	1900-2010	ARMA(1,1)	0.9844
5	Chongqing	1900-2000	ARMA(1,1)	0.9497
6	Chongqing	1900-2010	ARMA(1,1)	0.9667
7	Shenyang	1900-2000	ARMA(1,1)	0.9805
8	Shenyang	1900-2010	ARMA(1,1)	0.9847
9	Shanghai	1900-2000	ARMA(1,1)	0.9737
10	Shanghai	1900-2010	ARMA(1,1)	0.9846
11	Nanjing	1900-2000	ARMA(1,1)	0.9654
12	Nanjing	1900-2010	ARMA(1,1)	0.9814
13	Xian	1900-2000	ARMA(1,2)	0.9343
14	Xian	1900-2010	ARMA(1,2)	0.9732
15	Changchun	1900-2000	ARMA(1,1)	0.9787
16	Changchun	1900-2010	ARMA(1,1)	0.9849

Table 5.11 shows that the average temperature series for the eight cities are all ARMA series with AR components of order one. If data from 1900 to 2000 are used, the estimated coefficient value of the AR component is already quite close to one. When we extended the time interval so that it ends in 2010 instead of 2000, with the number of observations increased from 101 to 111, the estimated coefficient value was even closer to one. These additional evidences demonstrate that varying coefficient models can

adequately represent all these series. As time intervals increase, the existence of unit roots becomes clearer. As a result, the eight Chinese series should all be nonstationary series with unit roots. Following confirmation of the above time series' stationarity, we used the EG two-step method to test the cointegration relationship between the eight Chinese cities' series, and the results are shown in Figure 5.8.

FIGURE 5.8: Chinese Cities' Cointegration Test results

Figure 5.8 shows that the majority of the cities' annual average temperature series are cointegrated and have long-term equilibrium relationships between them. Table 5.12

summarises the long-term equilibrium relationships derived from linear regression analysis.

Table 5.12: Long-Term Equilibrium Relationships between Chinese Cities' Temperatures

No.	Region	Relationship
1	Changchun and Harbin	$\text{Changchun}_t = -0.0325 + 0.9270 \times \text{Harbin}_t$
2	Chongqing and Xian	$\text{Xian}_t = 0.0573 + 0.9520 \times \text{Chongqing}_t$
3	Shanghai and Nanjing	$\text{Nanjing}_t = 0.0091 + 0.9840 \times \text{Shanghai}_t$
4	Dalian and Harbin	$\text{Harbin}_t = 0.0908 + 1.1316 \times \text{Dalian}_t$
5	Nanjing and Xian	$\text{Xian}_t = 0.0070 + 0.7641 \times \text{Nanjing}_t$
6	Shanghai and Xian	$\text{Xian}_t = 0.0120 + 0.7268 \times \text{Shanghai}_t$
7	Chongqing and Nanjing	$\text{Nanjing}_t = 0.0190 + 0.8062 \times \text{Chongqing}_t$
8	Chongqing and Shanghai	$\text{Shanghai}_t = 0.0056 + 0.7765 \times \text{Chongqing}_t$
9	Dalian and Xian	$\text{Xian}_t = -0.0585 + 0.4598 \times \text{Dalian}_t$
10	Shenyang and Shanghai	$\text{Shanghai}_t = -0.0862 + 0.4586 \times \text{Shenyang}_t$
11	Shanghai and Changchun	$\text{Changchun}_t = 0.1663 + 1.0576 \times \text{Shanghai}_t$
12	Shenyang and Nanjing	$\text{Nanjing}_t = -0.0756 + 0.4705 \times \text{Shenyang}_t$
13	Dalian and Shenyang	$\text{Shenyang}_t = -0.0157 + 1.0877 \times \text{Dalian}_t$
14	Shenyang and Xian	$\text{Xian}_t = -0.0508 + 0.3655 \times \text{Shenyang}_t$
15	Nanjing and Changchun	$\text{Changchun}_t = 0.1523 + 1.0116 \times \text{Nanjing}_t$
16	Chongqing and Shenyang	$\text{Shenyang}_t = 0.0819 + 0.5948 \times \text{Chongqing}_t$
17	Shenyang and Changchun	$\text{Changchun}_t = 0.0649 + 1.0701 \times \text{Shenyang}$
18	Dalian and Changchun	$\text{Changchun}_t = 0.0484 + 1.1542 \times \text{Dalian}_t$
19	Chongqing and Changchun	$\text{Changchun}_t = 0.1417 + 0.5348 \times \text{Chongqing}_t$
20	Xian and Changchun	$\text{Changchun}_t = 0.1158 + 0.7041 \times \text{Xian}_t$
21	Harbin and Shenyang	$\text{Shenyang}_t = -0.0751 + 0.7409 \times \text{Harbin}_t$

We also used the obtained long-term equilibrium relationships to forecast each city's temperature from 2001 to 2010 and compared them to the actual observed values. MSEs (mean squared errors) were also computed, and the results are shown in Table 5.13.

TABLE 5.13: Mean squared Errors of Individual Cities' Predictions

No.	City 1	City 2	MSE
1	Harbin	Changchun	0.0274
2	Chongqing	Xian	0.1232
3	Shanghai	Nanjing	0.0020
4	Dalian	Harbin	0.2494
5	Nanjing	Xian	0.0535
6	Shanghai	Xian	0.0577
7	Chongqing	Nanjing	0.3765
8	Chongqing	Shanghai	0.3831
9	Dalian	Xian	0.3099
10	Shenyang	Shanghai	0.4341
11	Shanghai	Changchun	0.4105
12	Shenyang	Nanjing	0.4601
13	Dalian	Shenyang	0.0364
14	Shenyang	Xian	0.3871
15	Nanjing	Changchun	0.4306
16	Chongqing	Shenyang	0.6314
17	Shenyang	Changchun	0.1038
18	Dalian	Changchun	0.1868
19	Chongqing	Changchun	1.4077
20	Xian	Changchun	1.0768
21	Harbin	Shenyang	0.0502

Table 5.13 shows that the majority of the predictions are reasonably accurate. However, using the equilibrium relationships, the forecast results for two pairs, Chongqing versus Changchun and Xian versus Changchun, are relatively poor compared to the other pairs. By comparing the residuals of these two pairs of equilibrium relationships, we discovered that the residual series of both pairs have an increasing trend. This demonstrates that Chongqing and Xian's growth trends are more prominent than Changchun's. However, due to the lack of available data, this linear growth trend in residuals cannot be confirmed. That is why when using the ADF test on the residuals, it still yields a stationary result, leading to the final conclusion that the two pairs are cointegrated. As a result, we can expect this long-term equilibrium relationship to break down as the number of observations increases. We can also use the EG two-step method to perform a cointegration test on other cities' average temperature series. We might get results similar to what we have seen for Chinese cities.

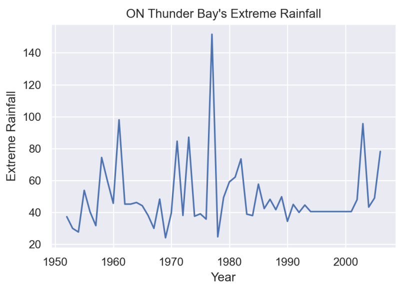
5.1.5 Tests of Cointegration between Temperature Series Averaged at Different Spatial Scales

Previous cointegration tests confirmed the existence of cointegration relationships between the temperature series averaged at the same spatial scales. For example, averaged at the continental scale, we confirmed that Asia's annual average temperature series is cointegrated with Africa's annual average temperature series; averaged at the individual city's scale, we confirmed that the annual average temperature of Dalian, China is cointegrated with that of Xian, China. In addition, we also ran cointegration tests between the annual average temperature series spatially averaged at different scales; for example, between globally averaged series and continentally averaged series, between continentally averaged series and individual cities' series, and so on. Using the same EG two-step method, we confirmed that the globally averaged temperature series is cointegrated with the temperature series averaged over the Northern Hemisphere. But the cointegration between the global annual average temperature series and that of the Southern hemisphere cannot be confirmed because the stochastic trend in the southern hemisphere is weak.

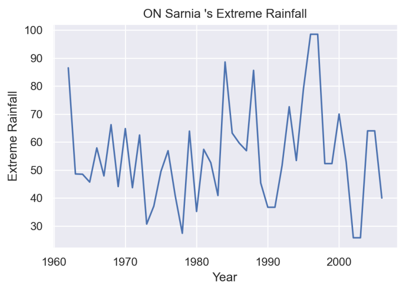
Between the hemispheric temperature series and continental temperature series, such as between the temperature series of Northern Hemisphere and North America, cointegration relationships were also found. Between the continental and individual city's temperature series, such as between Harbin, China and Asia, cointegration relationships were found to exist as well. It was also confirmed that some long-term equilibrium cointegration relationships exist between the globally averaged annual average temperature and continentally averaged annual average temperature series. We also performed cointegration tests on global annual average temperature and individual city's annual average temperature series. It was confirmed that the global average temperature series has cointegration relationships with the annual average temperature series of the eight Chinese cities listed in Table 5.11. All these additional cointegration tests demonstrated that integrated temperature series, whether they are spatially averaged at the global, hemispheric, continental, or individual city scales, are almost all cointegrated. This implies that the stochastic trends contained in all the temperature series may be caused by the same physical processes.

5.2 Extreme Rainfall Time Series

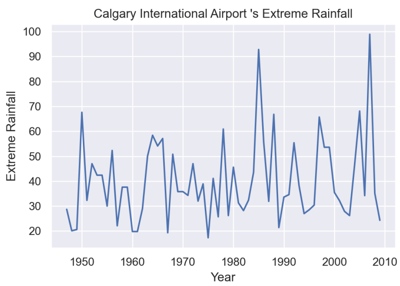
We analyzed extreme rainfall records for several Canadian cities and plotted them as time series to see if they are stationary. We obtained extreme rainfall data for 11 cities over a period of about 50 years from the Government of Canada's website $(https://climate.weather.gc.ca/)$. Figure 5.9 depicts the extreme rainfall series for each cities.



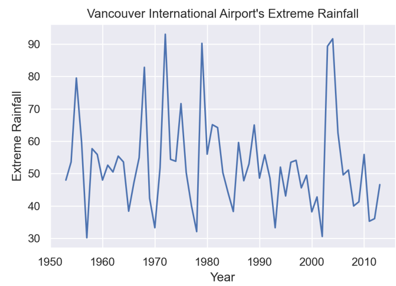
(a) Thunder Bay



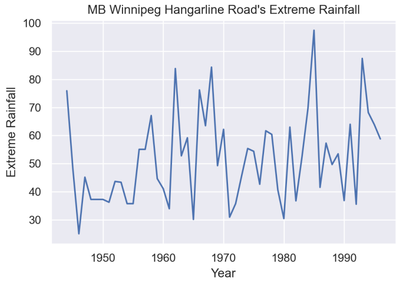
(b) Sarnia



(c) Calgary Int Airport

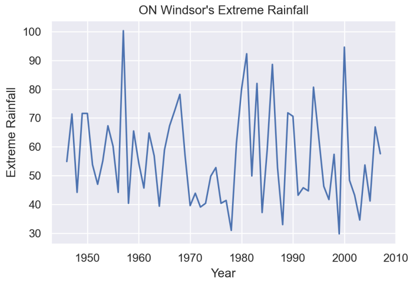


(d) Vancouver Int Airport

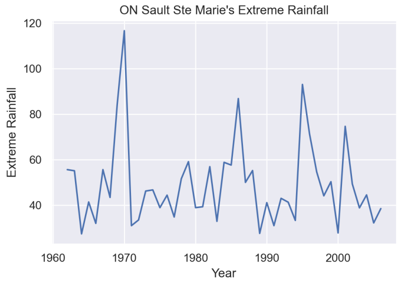


(e) Winnipeg

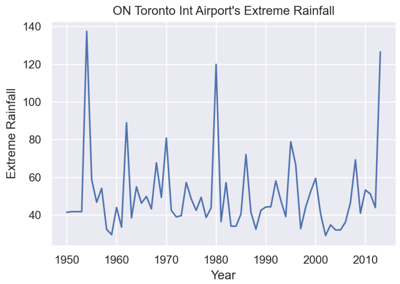
(f) Victoria



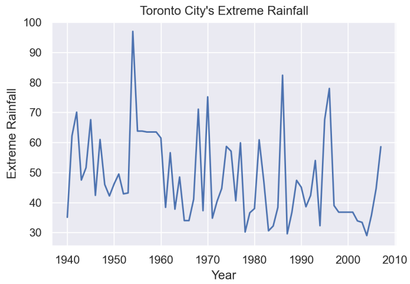
(g) Windsor



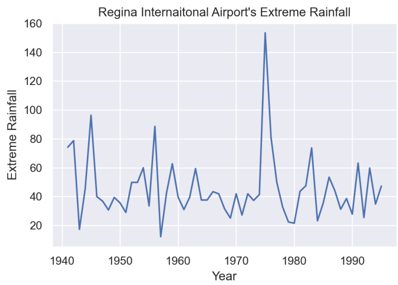
(h) Sault Ste Marie



(i) Toronto Int Airport



(j) Toronto City



(k) Regina Int Airport

FIGURE 5.9: Canadian Extreme Rainfall Series

To begin, in some cities where data was missing in specific years, we used the Forward method to fill in the gaps. The Forward method fills missing values with previous data. We then used the Mann-Kendall test to test for deterministic trend. It was shown that none of them have a deterministic trend. Afterwards, the ADF and KPSS methods were used to test for stochastic trend. According to the KPSS test, all series are stationary. However, the ADF test results for two cities' extreme rainfall time series differ from that of the KPSS test. The detailed test results for the two cities are summarized in Table 5.14.

Table 5.14: Extreme Rainfall Stochastic Trend Test Results

No.	City	P-Value (ADF)	ADF result	P-Value (KPSS)	KPSS result
1	Victoria International Airport	0.2862	Unit Root	0.1	stationary
2	Sarnia	0.3863	Unit Root	0.1	stationary

When we used the AIC criterion to validate their ARMA models, we discovered that their residual series may have ARCH characteristics. That is why we performed additional analysis on these two time series.

To begin, we validated the ARMA models of these two series using the AIC criterion. The results are shown in Table 5.15.

Table 5.15: Extreme Rainfall Regression Results

No.	City	Model
1	Victoria International Airport	white noise
2	Sarnia	AR(1)

Secondly, we used the AR(1) model to regress Sarnia's time series and obtained the residual series. The squared residual series meets the ARCH model's characteristics. This was confirmed using the Ljung-Box test. Thirdly, we used the AIC criterion to confirm that Sarnia's extreme rainfall residual process fits the GARCH(1,1) model shown below.

$$\begin{aligned} r_t &= \sigma_t \varepsilon_t \\ \sigma_t^2 &= 2.3264 + 1.1574 \times 10^{-12} r_{t-1}^2 + \sigma_{t-1}^2 \end{aligned}$$

The above results show that the residuals of Sarnia's time series are not independent, and the coefficient of σ is positive, indicating that the variance of the random error in the time series increases with time. It means that the time series is becoming more unpredictable. We conducted the Ljung-Box test for other locations' extreme rainfall series as well, it revealed that all other locations' extreme rainfall series are white noise.

Chapter 6

Summary and Conclusions

6.1 Summary

This thesis focuses on the application of the large deviation principle to varying coefficient autoregressive models of order one, particularly in the case where $\phi_T = 1 - \frac{1}{T}$. The process will transition from stationary to non-stationary with $T \to \infty$. Stationarity test methods such as ADF and KPSS may not work in these types of models because they are only applicable when the coefficient is fixed. According to the large deviation principle, if the estimator's probability is far from its actual value with constant distance, the probability will follow the rate function described in Chapter 4. This allows us to better understand the asymptotic behaviour of this estimator and thus how it converges when numerically simulated. Many time series in the actual dataset, particularly temperature series, follow the varying coefficient autoregressive model. By comparing the estimated coefficient values of the model with different observation periods, we can see that the estimator approaches one as the number of observations increases. In many cases, we may believe that the time series changes from stationary to non-stationary because it has a sudden change point at a specific moment or period, resulting in the change of state. However, when we examined the regional temperature series, we discovered that it has been increasing since the beginning and has not changed significantly in a long time. The difference in trend identification results is caused by the fact that the time series follow varying coefficient models.

6.2 Future Research

First, we believe that dynamic change models, rather than simply using breakpoints to reflect possible significant state changes in the sequence, would better represent the regional annual average temperature series. We discovered that the annual average temperature in many regions is rising, but the differenced time series is stationary. As a result, we believe that the likelihood of a sudden change in a short period is very low, and that the series itself has been affected at some point. This impact does not have a significant impact in a short period of time, but it has a profound effect over time, similar to what we call long memory time series.

Second, when it comes to actual data, the model we fit for each series is more complex than simple autoregressive models. The ARMA model, rather than the simple AR process, should be used by the majority of them. As a result, the large and moderate deviation principle for ARMA models with varying coefficients may be pursued in the future.

We discovered that most time series are displayed as white noise series when we processed the extreme rainfall series, which is due to an insufficient number of observations.

However, we discovered that the residuals of some of the extreme rainfall series are serially dependent on each other. As a result, more data should be collected to confirm whether there are any deterministic trends in the series and whether they are autoregressive. Furthermore, we know that rainfall and temperature have a specific physical relationship. However, because we have not yet determined whether the time series of extreme precipitation contain unit roots through diagnosis, we are unable to conduct a cointegration analysis between extreme precipitation and regional average temperature. As a result, we may conduct additional research on the rainfall sequence in the future to confirm whether it meets the characteristics of the varying coefficient model and then perform a cointegration analysis.

Appendix A

Chapter 6 Supplement

A1 Stationarity Test Code

```
# Trend Test
def mk_test(data):
    return mk.original_test(data,alpha = 0.05)[0]

# ADF test
def adf_test(data):
    if (mk_test(data) == 'increasing') or (mk_test(data) == 'decreasing'):
        dftest = adfuller(data, regression = "ct", autolag="AIC")
        output = []
        output.append(dftest[1])
        if dftest[1]>0.05:
            output.append("Non-stationary")
        else:
            output.append("Stationary")
        return output
    else:
        dftest = adfuller(data, regression = "c", autolag="AIC")
        output = []
        output.append(dftest[1])
        if dftest[1]>0.05:
            output.append("Non-stationary")
        else:
            output.append("Stationary")
        return output

def adf_test_2(data):
    dftest = adfuller(data, regression = "nc", autolag="AIC")
    output = []
    output.append(dftest[1])
    if dftest[1]>0.05:
        output.append("Non-stationary")
    else:
        output.append("Stationary")
    return output

# KPSS test
def kpss_test(data):
```

```
    if (mk_test(data) == 'increasing') or (mk_test(data) == 'decreasing'):
        kpsstest = kpss(data, regression = 'ct', nlags="auto")
        output = []
        output.append(kpsstest[1])
        if kpsstest[1]>0.05:
            output.append("Stationary")
        else:
            output.append("Non-stationary")
        return output
    else:
        kpsstest = kpss(data, regression = 'c', nlags="auto")
        output = []
        output.append(kpsstest[1])
        if kpsstest[1]>0.05:
            output.append("Stationary")
        else:
            output.append("Non-stationary")
        return output

def kpss_test_2(data):
    kpsstest = kpss(data, regression = 'c', nlags="auto")
    output = []
    output.append(kpsstest[1])
    if kpsstest[1]>0.05:
        output.append("Stationary")
    else:
        output.append("Non-stationary")
    return output

# White Noise test
def white_noise_test(data):
    result = acorr_ljungbox(data.values, lags=[6, 12], return_df=True)
    if result.loc[6, 'lb_pvalue'] < 0.05:
        return 'not white noise'
    else:
        return 'white noise'
```

A2 Cointegration Test Code

```
# Define EG two steps method
def find_cointegrated_pairs(data):
    n = data.shape[1]
    score_matrix = np.zeros((n, n))
    pvalue_matrix = np.ones((n, n))
    keys = data.keys()
    pairs = []

    for i in range(n):
        for j in range(i+1, n):
            S1 = data[keys[i]]
```

```
            S2 = data[keys[j]]
            result = coint(S1, S2, trend = "ct", method = 'aeg')
            score = result[0]
            pvalue = result[1]
            score_matrix[i, j] = score
            pvalue_matrix[i, j] = pvalue
            if pvalue < 0.05:
                #print((keys[i], keys[j]), pvalue)
                pairs.append((keys[i], keys[j], pvalue))
    return score_matrix, pvalue_matrix, pairs

# Return p-values of each pair's test and draw the heat map
scores, pvalues, pairs = find_cointegrated_pairs(DF)

import seaborn
fig, ax = plt.subplots(figsize=(12,10))
heatmap = seaborn.heatmap(pvalues, xticklabels=DF.keys(), yticklabels=DF.keys(),annot=True,cmap="YlGnBu", mask = (pvalues >= 0.05))
heatmap.figure.savefig('heatmap.png', dpi = 600)
```

Bibliography

[1] B. Bercu. On large deviations in the Gaussian autoregressive process: stable, unstable and explosive cases. *Bernoulli* (2001), 299–316.

[2] B. Bercu, F. Gamboa, and M. Lavielle. Sharp large deviations for Gaussian quadratic forms with applications. *ESAIM: Probability and Statistics* 4 (2000), 1–24.

[3] B. Bercu, F. Gamboa, and A. Rouault. Large deviations for quadratic forms of stationary Gaussian processes. *Stochastic Processes and their Applications* 71(1) (1997), 75–90.

[4] N. H. Chan and C.-Z. Wei. Asymptotic inference for nearly nonstationary AR (1) processes. *The Annals of Statistics* (1987), 1050–1063.

[5] Y. Chang, R. K. Kaufmann, C. S. Kim, J. I. Miller, J. Y. Park, and S. Park. Evaluating trends in time series of distributions: A spatial fingerprint of human effects on climate. *Journal of Econometrics* 214(1) (2020), 274–294.

[6] X. Chen. Moderate deviations for m-dependent random variables with Banach space values. *Statistics & probability letters* 35(2) (1997), 123–134.

[7] H. Cramér. Sur un nouveau théoreme-limite de la théorie des probabilités. *Actual. Sci. Ind.* 736 (1938), 5–23.

[8] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Springer, 1998.

[9] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association* 74(366a) (1979), 427–431.

[10] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society* (1982), 987–1007.

[11] R. F. Engle and C. W. Granger. Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society* (1987), 251–276.

[12] S. Fatichi, S. Barbosa, E. Caporali, and M. Silva. Deterministic versus stochastic trends: Detection and challenges. *Journal of Geophysical Research: Atmospheres* 114(D18) (2009).

[13] T. Hastie and R. Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)* 55(4) (1993), 757–779.

[14] Y. Hong and T.-H. Lee. Inference on predictability of foreign exchange rates via generalized spectrum and nonlinear time series models. *Review of Economics and Statistics* 85(4) (2003), 1048–1062.

[15] P. Jones, D. Parker, T. Osborn, and K. Briffa. Global and hemispheric temperature anomalies–land and marine instrumental records. *U.S. Department of Energy Office of Scientific and Technical Information* (2006).

[16] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54(1-3) (1992), 159–178.

[17] A. Mas and L. Menneteau. Large and moderate deviations for infinite-dimensional autoregressive processes. *Journal of Multivariate Analysis* 87(2) (2003), 241–260.

[18] Y. Miao and G. Yang. A moderate deviation principle for m-dependent random variables with unbounded m. *Acta applicandae mathematicae* 104(2) (2008), 191–199.

[19] J. Y. Park. Weak Unit Roots (2003).

[20] P. C. Phillips. Understanding spurious regressions in econometrics. *Journal of Econometrics* 33(3) (1986), 311–340.

[21] P. C. Phillips. Time series regression with a unit root. *Econometrica: Journal of the Econometric Society* (1987), 277–301.

[22] P. C. Phillips. Regression theory for near-integrated time series. *Econometrica: Journal of the Econometric Society* (1988), 1021–1043.

[23] D. I. Stern and R. K. Kaufmann. Detecting a global warming signal in hemispheric temperature series: Astructural time series analysis. *Climatic Change* 47(4) (2000), 411–438.

[24] J. S. White. The limiting distribution of the serial correlation coefficient in the explosive case. *The Annals of Mathematical Statistics* (1958), 1188–1197.

[25] W. A. Woodward and H. L. Gray. Global warming and the problem of testing for trend in time series data. *Journal of Climate* 6(5) (1993), 953–962.

[26] W. A. Woodward and H. Gray. Selecting a model for detecting the presence of a trend. *Journal of Climate* 8(8) (1995), 1929–1937.

[27] J. Worms. Large and moderate deviations upper bounds for the Gaussian autoregressive process. *Statistics & probability letters* 51(3) (2001), 235–243.

[28] M. Xu, J. Li, and Y. Chen. Varying coefficient functional autoregressive model with application to the US treasuries. *Journal of Multivariate Analysis* 159 (2017), 168–183.

[29] M. Yu and S. Si. Moderate deviation principle for autoregressive processes. *Journal of Multivariate Analysis* 100(9) (2009), 1952–1961.