The Pennsylvania State University

The Graduate School

Eberly College of Science

THREE ESSAYS ON NON-STATIONARY TIME SERIES

A Dissertation in
Statistics
by
Xiaoye Li

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

December 2013

The thesis of Xiaoye Li was reviewed and approved* by the following:

Zhibiao Zhao
Associate Professor of Statistics
Dissertation Adviser, Co-Chair of Committee

Runze Li
Professor of Statistics
Co-Chair of Committee

Fuqing Zhang
Professor of Meteorology
Joint Appointment with the Department of Statistics

Jingzhi Huang
Professor of Finance

Aleksandra Slavkovic
Associate Professor of Statistics
Graduate Program Chair

*Signatures are on file in the Graduate School.

Abstract

We study statistical inference for a class of non-stationary time series with time-dependent variances. Due to non-stationarity and the large number of unknown parameters, existing methods that are developed for stationary or locally stationary time series are not applicable. Based on a self-normalization technique, we address several inference problems, including self-normalized Central Limit Theorem, self-normalized cumulative sum test for change-point problem, long-run variance estimation through blockwise self-normalization, and self-normalization based wild bootstrap for non-stationary time series. Monte Carlo simulation studies show that the proposed self-normalization based methods outperform stationarity based alternatives. We demonstrate the proposed methodology using two real data sets: annual mean precipitation rates in Seoul during 1771–2000, and quarterly U.S. Gross National Product growth rates during 1947–2002. In the literature on change-point analysis, much attention has been paid to detecting changes in certain marginal characteristics, such as mean, variance, and marginal distribution. For time series data with nonparametric time trend, we study the change-point problem for the autocovariance structure of the unobservable error process. To derive the asymptotic distribution of the cumulative sum test statistic, we develop substantial theory for uniform convergence of weighted partial sums and weighted quadratic forms. Our asymptotic results improve upon existing works in several important aspects. The performance of the test statistic is examined through simulations and an application to interest rates data. To model the frequently observed nonstationarity phenomena in social and scientific fields, we propose a class of time-varying exogenous autoregressive models. While the model exhibits nonparametric time-varying dependence structure over a long time span, the model dynamics possess local stationarity within each small time interval. Furthermore, the model can incorporate important external nonstationary inputs to model the main time series of interest. These features allow for theoretical tractability as well as flexible applications. For nonparametric estimation of the coefficient functions, it is shown that the local linear estimation can adapt to the unknown nonstationarity, whereas the local constant estimation is strongly affected by the local stationarity. Some practically important inference and hypothesis testing problems are investigated. To better take into account the nonstationarity and dependence, we propose a sieve-wild bootstrap by combining the ideas from both the sieve and the wild bootstrap. The methodology is illustrated through both Monte Carlo simulations and an application to the stock return-inflation puzzle using the S&P 500 index and Consumer Price Index data during 1982–2012.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my deepest gratitudes to my adviser Dr. Zhibiao Zhao for his guidance and encouragement throughout the semesters. His brilliant insights have made enormous contributions to this thesis. I also appreciate all the members on my thesis committee, Dr. Runze Li, Dr. Fuqing Zhang and Dr. Jingzhi Huang, for their suggestions on improving the content of this dissertation and their time attending my presentation. Finally, I am very grateful to the entire faculty at the department of Statistics at Pennsylvania State University for providing all kinds of learning opportunities and the extraordinary academic environment where I was able to make the most out of my time, and to my dear classmates for our deep friendship and the countless pleasant moments we spent together throughout the years.

Chapter 1

Introduction

In time series analysis, stationarity requires that dependence structure is sustained over time, and thus we can borrow information from one time period to study model dynamics over another period. In practice, however, many climatic, economic, and financial time series are non-stationary and more challenging to deal with. First, since dependence structure varies over time, information is more localized. Second, non-stationary processes often require extra parameters to account for time-varying structure. In this dissertation, we are interested in developing new statistical tools to solve the inference problem for a class of non-stationary time series.

1.1 Inference for Modulated Stationary Processes

Consider a class of modulated stationary processes

$$X_i = \mu + \sigma_i e_i, \quad i = 1, \dots, n, \tag{1.1}$$

where e_i are stationary time series with zero mean, and $\sigma_i > 0$ are unknown constants adjusting for time-dependent variances. Then X_i oscillates around the constant mean μ whereas its variance changes over time in an unknown manner. In the special case of $\sigma_i \equiv 1$, (1.1) reduces to stationary case. If $\sigma_i = s(i/n)$ for a Lipschitz continuous function $s(t)$ on $[0, 1]$, then (1.1) is locally stationary. For the general non-stationary case (1.1), the number of unknown parameters is larger than the number of observations, and thus it is infeasible to estimate σ_i. Due to the non-stationarity structure and large number of unknown parameters, existing methods that are developed for (locally) stationary processes are not applicable, and our main purpose is to develop new statistical inference techniques.

First, we establish a uniform strong approximation result which can be used to derive self-normalized Central Limit Theorem (CLT) for the sample mean $\bar{X}$ of (1.1). For stationary case $\sigma_i \equiv 1$, by Fan and Yao (2003), under mild mixing conditions,

$$\sqrt{n}(\bar{X} - \mu) \Rightarrow N(0, \tau^2), \quad \text{where} \quad \tau^2 = \gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k \quad \text{and} \quad \gamma_k = \mathrm{Cov}(e_i, e_{i+k}). \qquad (1.2)$$

For modulated stationary case (1.1), it is non-trivial whether $\sqrt{n}(\bar{X} - \mu)$ has a CLT without imposing further assumptions on σ_i and the dependence structure of e_i. Moreover, even when the latter CLT exists, it is difficult to estimate the limiting variance due to the large number of unknown parameters. Under a strong invariance principle assumption, we establish a self-normalized CLT with the self-normalizing constant adjusting for time-dependent non-stationarity. The obtained CLT is an extension of the classical CLT for independent and identically distributed (IID) data or stationary time series to non-stationary time series. Furthermore, we extend the idea to linear combinations of means over different time periods, which allows us to address inference regarding mean levels over multiple time periods.

Second, we study wild bootstrap for non-stationary time series. Since the seminal work of Efron (1979), there has been a large literature on bootstrap under various settings, ranging from bootstrap for IID data in Efron (1979), wild bootstrap for independent observations with possibly non-constant variances in Wu (1986) and Liu (1988), to various block bootstrap and resampling methods for stationary time series in Künsch (1989), Politis and Romano (1994), Bühlmann (2002), and the monograph Larihi (2003). With the established self-normalized CLT, we propose a wild bootstrap procedure that is tailored to deal with non-stationary time series: the dependence is removed through a scaling factor, and the non-stationarity structure of the original data is preserved in the wild bootstrap data-generating mechanism. Our simulation study shows that the proposed wild bootstrap method outperforms the widely used stationarity-based block bootstrap.

Third, we address change-point analysis. Change-point problem has been an active area of research; see Pettitt (1980) for proportion changes in binary data, Horváth (1993) for mean and variance changes in Gaussian observations, Bai and Perron (1998) for coefficient changes in linear models, Horváth et al. (2008) for mean change in time series, Shao and Zhang (2010) for change-points for stationary time series, and the monograph by Csörgő and Horváth (1997) for more discussion. Most of the aforementioned works deal with (locally) stationary and/or independent data. Hansen (2000) studied tests for constancy of parameters in linear regression models with non-stationary regressors by assuming the errors form martingale differences and are conditionally homoscedastic. Here we consider

$$H_0: X_i = \mu_i + \sigma_i e_i, \ \mu_1 = \cdots = \mu_n, \text{ versus } H_a: \mu_1 = \cdots = \mu_J \neq \mu_{J+1} = \cdots = \mu_n, \quad (1.3)$$

where J is an unknown change-point. To our knowledge, there has been no attempt on change-point analysis under this non-stationary time series framework. The popular cumulative sum (CUSUM) test is developed for stationary time series and does not take into account the time-dependent variances. Using the self-normalization idea, we propose a self-normalized CUSUM test and a wild bootstrap method to obtain its critical value. Our empirical studies show that the usual CUSUM tests tend to over-reject the null hypothesis in the presence of non-constant variances. By contrast, the self-normalized CUSUM test yields size close to the nominal level.

Fourth, we estimate the long-run variance τ^2 in (1.2). Long-run variance plays an essential role for statistical inference involving time series data. In the literature of long-run variance estimation, most existing works deal with stationary process through various block bootstrap and subsampling approaches; see Carlstein (1986), Künsch (1989), Politis and Romano (1994), Götze and Künsch (1996), Larihi (1999), Bühlmann (2002) and the monograph Larihi (2003). Recently, Müller (2007) studied robust long-run variance estimation for locally stationary process. For non-stationary model (1.1), the error process $\{e_i\}$ is contaminated with unknown standard deviations $\{\sigma_i\}$, and thus the aforementioned methods are not applicable here. To attenuate this problem, we apply blockwise self-normalization to remove non-stationarity, resulting in asymptotically stationary blocks.

Fifth, the proposed self-normalization-based methods can be extended to linear regression model

$$X_i = U_i\beta + \sigma_i e_i, \tag{1.4}$$

where $U_i = (u_{i,1}, \ldots, u_{i,p})$ are deterministic covariates, $\beta = (\beta_1, \ldots, \beta_p)'$ is the unknown column vector of parameters. For example, if $u_{i,r} = i^{r-1}, r = 1, \ldots, p$, then we have polynomial time trend. Phillips et al. (2007) considered the case when the error is a stationary process and their approach is not applicable here due to the unknown non-constant variances σ_i^2.

1.2 Testing for Changes in Autocovariances of Nonparametric Time Series Models

In the large literature on change-point analysis, much effort has been devoted to detecting changes in certain marginal characteristics. Some representative works include Pettitt (1980) for proportion changes in binary data, Horváth (1993) for mean and variance changes in Gaussian observations, Inclán and Tiao (1994) for changes in variance of independent observations, Bai and Perron (1998) for coefficient changes in linear regression models, Lavielle (1999) and Mei (2006) for changes in marginal distribution, Kokoszka and Leipus (2000) and Wied (2011) for variance changes in time series models, Aue et al. (2008b) for mean change in time series with stationary errors, Zhao and Li (2011) for mean changes in time series data with non-stationary errors, and the monograph Csörgő and Horváth (1997) for related references.

While working well for independent observations, marginal information, such as mean, variance, and marginal distribution, is usually less informative for time series data. In fact, one of the main goals of time series analysis is to study how observations are dependent on each other, and marginal information is of little use in such contexts. For such purpose, autocovariance (or autocorrelation) is the most widely used measure of dependence among time series observations. In fact, autocorrelation function plays a key role in identifying potential candidate models in the classical ARMA modeling (Brockwell and Davis, 1991). Motivated by the above discussion, we study the change-point problem for autocovariances $\mathrm{cov}(e_i, e_{i+k})$ at a lag $k \geq 0$ of a time series $\{e_i\}_{i=1}^n$. Specifically, we consider testing

$$H_0^{(k)} : \mathrm{cov}(e_1, e_{1+k}) = \cdots = \mathrm{cov}(e_{n-k}, e_n), \tag{1.5}$$

against the alternative that there exists one J, called the change-point, such that

$$H_A^{(k)} : \mathrm{cov}(e_1, e_{1+k}) = \cdots = \mathrm{cov}(e_J, e_{J+1}) \neq \mathrm{cov}(e_{J+1}, e_{J+1+k}) = \cdots = \mathrm{cov}(e_{n-k}, e_n).$$

Most existing works on the change-point analysis (4.2) deal with parametric time series and the process of interest is directly observable. For example, Picard (1985) studied linear autoregressive models, and Lee, Ha and Na (2003) and Berkes, Gombay and Horváth (2009) considered general linear processes. The change-point analysis on the autoregressive parameters in Davis, Huang and Yao (1995) and Gombay (2008) can be used to detect autocovariance change-points in linear autoregressive models; also, see Ling (2007) for related work on nonlinear parametric autoregressive models. Galeano and Peña (2007) studied the change-point problem for the variance and correlations of vector autoregressive moving average models. On the one hand, a parametric model can provide a parsimonious description of the underlying model dynamics. On the other hand, as a modeling issue, it is hard to believe that we have better information about the explicit form of the underlying model than about the possibility of change-points in such models. Recently, Aue et al. (2009) and Shao and Zhang (2010) proposed tests for change-points in autocovariances of nonparametric time series models, and Wied, Krämer and Dehling (2012) considered testing for change-points in correlations between two processes. In all the aforementioned works, a restrictive assumption is that the process of interest is directly observable.

In practice, however, the process $\{e_i\}$ of interest is often not observable and the actual observations, denoted by $\{X_i\}$, may be non-stationary and contain unknown functions.

Consider the important case of the fixed-design nonparametric regression model:

$$X_i = f(i/n) + e_i, \quad i = 1, \ldots, n, \tag{1.6}$$

where $\{e_i\}$ is a process with $\mathbb{E}(e_i) = 0$, and $f(\cdot)$ is an unknown time trend. Model (4.1) and its variants have been extensively studied under different focuses, mainly on inferences for the (marginal information) mean trend $f(\cdot)$. For example, Altman (1990) considered estimation of $f(\cdot)$, and Wu and Zhao (2007) studied the change-point problem and simultaneous confidence band construction for $f(\cdot)$. Assuming that $f(\cdot)$ is a polynomial function and $\{e_i\}$ are uncorrelated stationary process, Aue et al. (2008a) studied coefficient changes.

Despite the rich literature on the change-point problem and the importance of autocovariances in time series analysis, the change-point analysis for autocovariances of $\{e_i\}$ in (4.1) seems unexplored, and our goal is to address the latter problem. Our results improve upon earlier studies in several important aspects. First, with the introduction of the nonparametric time trend $f(\cdot)$, model (4.1) offers a flexible framework for potential applications in climatic, economic, and financial time series fields. In such applications, the data often exhibit complicated time-varying trend that can hardly be captured by parametric forms. Second, our theory is developed under a general dependence structure covering many linear and nonlinear processes, whereas many previous works (e.g., Lee, Ha and Na, 2003; Berkes, Gombay and Horváth, 2009) depend heavily on the linear process assumption. Third, the presence of the unknown function $f(\cdot)$ makes the problem significantly more challenging, and we develop substantial technical tools, such as moment inequalities and uniform convergence for weighted partial sums and quadratic forms, which may be of independent interest.

1.3 On Time-varying Exogenous Autoregressive Models

By allowing the model dynamics to change smoothly in time, locally stationary models can provide flexible and tractable alternatives over stationary models. On one hand, locally stationary models allow model dynamics to vary over time and thus can be used to model nonstationary time-varying dependence over a long time span. On the other hand, "smooth change in time" implies approximate stationarity within a small time window and thus technical tractability. Thanks to these nice features, there has been growing interest in studying locally stationary models since the seminal work of Dahlhaus (1997). For example, Davis et al. (2006) studied piecewise stationary autoregressive (AR) models, Dahlhaus and Subba Rao (2006) considered locally stationary AR conditional heteroscedastic (ARCH) models, Nason et al. (2000) studied locally stationary wavelet processes, and Wu and Zhou (2011) proposed a general class of locally stationary models using their physical dependence measure. Some earlier contributions include the time-varying AR models in Subba Rao (1970). We refer the reader to Dahlhaus (2012) for a survey.

Compared to the classical autoregressive (AR) models, AR models with exogenous or external inputs (ARX) can provide a more flexible modeling framework. One advantage of ARX models is that they allow the time series of interest to depend on both the lagged observations of the sequence and some external variables outside of the process. This feature allows researchers to incorporate important external information into the modeling of the main time series. For example, in the economic scheme, an exogenous variable can be a shift in the oil supply affecting prices [Kilian (2008)] or a change in consumer preferences for manufactured products affecting prices. However, the classical ARX models assume stationarity and thus may not be appropriate for data with time-varying characteristics.

Motivated by the above discussions, we propose a class of time-varying exogenous AR (TV-ARX) model that possesses several appealing features. First, the functional-coefficients vary over time and thus the model can be used to describe nonstationary time-varying dependence structure. Second, we do not impose any parametric assumptions on the functional-coefficients; instead we adopt the nonparametric approach and let the data speak for themselves. This feature is especially useful when the data exhibit complicated time-varying pattern that can hardly be described using any simple parametric (such as constant or linear) curves. Third, the functional-coefficients are smooth functions of time and therefore the model enjoys local stationarity, i.e., the model dynamics are approximately the same over a small time interval while the model exhibits clear time-varying pattern over a long time span. Fourth, the model can incorporate information from external inputs which are from another nonstationary process.

We develop a comprehensive account of estimation and inference methodology for the proposed model. For nonparametric estimation of the coefficient functions, we show an interesting phenomenon that the local linear estimation can adapt to the unknown local stationarity of the model, whereas the local constant estimation is strongly affected by the local stationarity structure. Some practically important inference and hypothesis testing problems are investigated, including constructing simultaneous confidence band, testing whether the functional-coefficients are time-varying or constant, and testing the significance of exogenous inputs. To better take into account the nonstationarity and dependence, we propose a sieve-wild bootstrap that combines the ideas from both the sieve bootstrap [Bühlmann (1997)] and the wild bootstrap [Liu (1988); Davidson and Flachaire (2008)]. Monte Carlo studies in Section 5.3.1 show good finite sample performance.

The proposed model and methodology can be potentially applied to many data sets with time-varying feature. In Section 5.3.2, we consider an application to the stock return-inflation puzzle and our TV-ARX modeling sheds some new insight.

1.4 The Structure of This Dissertation

The rest of this dissertation is organized as follows. In Chapter 2, we provided a detailed literature review on Central Limit Theorem for the sample mean of (local) stationary time series, the bootstrap methods for independent data, and the pioneer work for the change-point detection.

Chapter 3 demonstrates the statistical inference for modulated stationary time series. We first establish a uniform approximation result which can be used as a self-normalization technique for the specific class of modulated stationary time series model. Based on that, we address several inference problems, including self-normalized Central Limit Theorem, self-normalized cumulative sum test for change-point problem, long-run variance estimation through blockwise self-normalization, and selfnormalization based wild bootstrap for non-stationary time series. Monte Carlo simulation studies are conducted to compare the performance of the proposed self-normalization based methods with the stationarity based alternatives. We also apply the proposed methodology into two real data sets in this chapter.

Chapter 4 is focused on testing for changes in autocovariances of nonparametric time series models. We introduce the dependence structure along with example and establish a functional central limit theorem for the autocovariance process and study the cumulative sum test. Section 4.3 contains some useful results on weighted partial sum and weighted quadratic forms. Simulation studies and an application are provided.

In Chapter 5, we study the time-varying exogenous non-staionary auto-regressive model. We first illustrate the estimation of the time-varying coefficients. Then the asymptotic distribution of the non-parametric estimate is derived under certain constraints of error term. Several inference problems are discussed accordingly. Numerical examples of the proposed model applied to Monte Carlo simulation and real data set are displayed in the end.

Chapter 2

Literature Review

This chapter provides a brief review of the pioneered work related to our study topics, which are Central Limit Theorem (CLT) for the Sample Mean, bootstrap methods, and change-point detection. All these statistical inference problems described in this chapter are on the (local) stationary processes base.

2.1 Central Limit Theorem (CLT) for the Sample Mean

2.1.1 CLT for the Sample Mean of Stationary Data

In probability theory, the central limit theorem (CLT) states conditions under which the mean of a sufficiently large number of independent random variables, each with finite mean and variance, will be approximately normally distributed. The classical central limit theorem (in its common form) requires the random variables to be identically distributed.

Theorem 1. *Let $X_1, \ldots, X_n$ be from an i.i.d data set, with $E(X_1) = \mu$, $var(X_1) = \sigma^2$, and $E(X_1^2) < \infty$, then we have*

$$\sqrt{n}(\bar{X}_n - \mu) \Rightarrow \mathrm{N}(0, \sigma^2) \tag{2.1}$$

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

Although the classical CLT is a breakthrough of the asymptotic theorem, the assumption under which the data is independent identical distributed, is too strong. Fan and Yao (2003) relaxed this assumption to stationary data and proposed the CLT for the sample mean of the stationary m-dependent process.

Theorem 2. *Let $\{X_t\}$ be a strictly m-dependent stationary process, with mean $E(X_t) = 0$, and autocovariance function $\gamma(k) = cov(X_0, X_k)$. Assume that*

$$\sigma_m^2 = \gamma(0) + 2\sum_{k=1}^{m} \gamma(k) \neq 0. \tag{2.2}$$

Let $\bar{X} = (X_1 + \cdots + X_n)/n$ be the sample average. Then

$$\lim_{n\to\infty} n\,Var(\bar{X}) = \sigma_m^2 \quad \text{and} \quad \sqrt{n}\bar{X} \to^{D} N(0, \sigma_m^2). \tag{2.3}$$

They also discussed the extension of theorem 2 to more general time series setting. Assume that ε_i, $i \in$ Z are independent and identically distributed random variables. We can view ε_i as the signal or shock at time t. Let f be an appropriate function. Define

$$X_i = f(\varepsilon_i, \varepsilon_{i-1}, \ldots). \tag{2.4}$$

So basically, X_i is a function of the past shocks ε_j, $j \leq i$. Assume that $\mathrm{E}(X_i) = 0$ and $\mathrm{Var}(X_i) < \infty$. One special example of (2.4) is the causal linear process

$$X_i = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{i-j}. \tag{2.5}$$

A general CLT for stationary process is as followed.

Theorem 3. *Let $X_1, \ldots, X_n$ be from (2.4). Denote by $\{\varepsilon_i'\}$ an independent copy of $\{\varepsilon_i\}$. Assume that for some $\rho \in (0, 1)$*

$$\|X_0 - X_0(k)\|_2 = [E|X_0 - X_0(i)|^2]^{1/2} = O(\rho^k), \tag{2.6}$$

where

$$X_0(k) = f(\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{-k+1}, \varepsilon'_{-k}, \varepsilon'_{-k-1}, \ldots), \tag{2.7}$$

Then

$$Z_n := \frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma} \rightarrow^D N(0,1). \tag{2.8}$$

where

$$\sigma^2 = \gamma(0) + 2\sum_{k=1}^{\infty} \gamma(k), \quad \textit{where} \quad \gamma(k) = Cov(X_i, X_{i+k}). \tag{2.9}$$

2.1.2 Asymptotic Distribution for the Sample Mean of Non-stationary Data

Although stationarity has played an important role in the theoretical treatment of time series procedures, many series (i.e., in economics or sound analysis) show a nonstationary behavior. Locally stationary processes were introduced in Dahlhaus (1997) by using a time-varying spectral representation. The work is a generalization of Whittle's method for stationary processes. Whittle's method is based on minimization of the function

$$L_T(\theta) = \frac{1}{4\pi} \int_{-\pi}^{\pi} (\log f_\theta(\lambda) + \frac{I_T(\lambda)}{f_\theta(\lambda)}) d\lambda \tag{2.10}$$

where $f_\theta(\lambda)$ is the model spectral density and $I_T(\lambda)$ is the periodogram. The Whittle estimate is asymptotically efficient and $L_T(\theta)$ is an approximation to the Gaussian likelihood function. In Dahlhaus (1997), they generalized the method of Whittle to processes that only show locally a stationary behavior and replaced the periodogram $I_T(\lambda)$ in $L_T(\theta)$ by a local version and integrate over time. The resulting estimate was proven asymptotic normality.

In contrast to Dahlhaus's work in 1997, Polonik (2009) used a time-varying $MA(\infty)$ representation and formulate the assumptions in the time domain. The assumptions of his theorem are more general than, for example, in Dahlhaus (1997) since the parameter curves are allowed to have jumps. They first introduced the empirical spectral process indexed by classes of functions, derived its convergence (including a functional central limit theorem) and proved a maximal exponential inequality and a GlivenkoCCantelli-type convergence result. They also discussed parametric quasi-likelihood estimation, nonparametric quasi-likelihood estimation, inference under model misspecification by stationary models and local estimates.

Since data from nonstationary stochastic time series have often been modeled as piecewise stationary processes with abrupt changes. A general class of piecewise locally stationary processes was introduced by Adak, Sudeshna (1998) with the help of spectrograms.

Definition 1. *A sequence of zero-mean stochastic process* $X_{t,N}(t = 1, \ldots, N)$*is called locally stationary at time* u*,*$0 \le u \le 1$*, if there exists a representation*

$$X_{t,N} = \int_{-1/2}^{1/2} A^0_{t,N}(\lambda) e^{i2\pi\lambda t} dZ(\lambda) \tag{2.11}$$

where (a) $Z(\lambda)$ *is an orthogonal increment, zero-mean process, and (b) there exist constants* $K(K \ge 0)$*,* $c(c > 0)$ *and a constant* $\alpha(1/2 < \alpha < 1)$ *and a function* $A : [0, 1] \times [-1/2, 1/2] \to \mathcal{C}$ *with* $A(u, \lambda) = \overline{A(u, -\lambda)}$ *such that for all* N*,*

$$\max_{t:(t/N)\in\varepsilon_N(u)} \sup_{\lambda} |A^0_{t,N}(\lambda) - A(u, \lambda)| \le KN^{-\lambda}, \tag{2.12}$$

where $\varepsilon_N(u) = [u - cN^{\alpha}, u - cN^{\alpha}]$ *is a small interval centered at* u*.*

Definition 2. *For a sequence of processes* $X_{t,N}$*, locally stationary at time* u*, the time dependent spectrum at time* u *is*

$$f(u, \lambda) = |A(u, \lambda)|^2. \tag{2.13}$$

Definition 3. *A sequence of zero-mean stochastic processes is said to be piecewise locally stationary if it is locally stationary (according to Definition 1) at all time points* $u \in [0, 1]$*,* *except possibly at finitely many jump points.*

Theorem 4 below shows that such piecewise locally stationary processes can be well approximated by piecewise stationary processes under some mild regularity conditions.

Theorem 4. *Consider the class of sequences of piecewise stationary process* $\mathcal{M} = \{\widetilde{X}_{t,N}\}$, $t = 1, \ldots, N$,

$$\widetilde{X}_{t,N} = \sum_{j=0}^{J-1} \widetilde{X}_t^{(j)} \mathcal{I}(u_j \le t/N < u_{j+1}), \quad with \quad \tfrac{J}{N} \to 0 \quad as \quad N \to \infty, \tag{2.14}$$

where $0 = u_0 < u_1 < \cdots < u_J = 1$ *form a partition of* $[0, 1]$ *that depends on* N *and* $\widetilde{X}_t^{(j)}$ *are stationary processes with spectra* $f^{(j)}(\lambda)$*. Then, for any sequence of a piecewise locally stationary process* $X_{t,N}(t = 1, \cdots, N)$ *there exists a sequence* $\widetilde{X}_{t,N} \in \mathcal{M}$ *such that for all* N,

$$\frac{1}{N} \sum_{t=1}^{N} \mathrm{E}(X_{t,N} - \widetilde{X}_{t,N})^2 = O(N^{-2\alpha}), \tag{2.15}$$

where $0 < \alpha < 1$.

So a piecewise locally stationary time series can be partitioned into approximately stationary intervals within which it is close to being a stationary process. This suggests that in estimating the time-dependent spectrum of such a non- stationary process, a segmentation procedure that partitions the time series into approximately stationary intervals can be used. The purpose of the segmentation algorithm is to detect changes in the spectra of the process over time and to estimate these change points.

A tree-based adaptive segmented spectrogram algorithm (TASS) is proposed here to estimate the time-dependent spectrum, which partitions the time-domain as follows:

Initially, a complete is grown that recursively halves each segment of the data into very short segment.

The spectrum in each segment of the tree is estimated and an optimal pruning algorithm used to recombine adjacent segments for which the spectra are the same.

Due to the stochastic nature of the data and the variability in the estimated spectra can result in an inaccurate segmentation with a few spurious segments. The general solution to this problem is to prune the tree, but with a penalization on the number of terminal nodes of the tree.

An optimal choice of the penalty parameter is required to trade off between frequency resolution and nonstationarity bias. The optimal choice is usually made by some resampling strategy and cross-validation. However, in the situation of time series, naive resampling is not advisable, due to the correlations in the observed data. It is well known that spectral estimates are approximately uncorrelated (for stationary time series), suggesting that one should resample from the spectra, rather than from the observed data.

2.2 Bootstrap Methods for Independent Observations

The bootstrap resampling procedure is known to be a good general procedure for estimating a sampling distribution under i.i.d. models. In practical situations the i.i.d. setup is often violated, hence many scholars have been focused on the study of bootstrap procedure under the non-i.i.d. models. In his paper in 1986, Wu proposed a wild bootstrap procedure on the regression problem

$$Y_i = x_i\beta + e_i, \tag{2.16}$$

where $x_i's$ are nonzero real numbers, β is the parameter, $\mathrm{E}(e_i) = 0$, $\mathrm{var}(e_i) = \sigma_i^2$ and $e_i's$ are independent. The proposed procedure modified the classical bootstrap procedure providing consistent results regardless of the nonidentical error variances. In Liu (1988), deeper exploration was conducted with a conclusion that under certain constrains of external population, Wu's bootstrap shared the usual second order asymptotic properties of the classical bootstrap.

2.2.1 The Classical Bootstrap

Consider now the simple linear model (2.16). The least square estimate of β is $\hat{\beta} = (\sum_{i=1}^n x_i Y_i)/\sum_{j=1}^n x_j^2$. Clearly $\text{var}(\hat{\beta}) = \sum_{i=1}^n x_i^2 \sigma_i^2/(\sum_{j=1}^n x_j^2)^2$. Let $r_i = Y_i - x_i\hat{\beta}$ be the residuals. The classical bootstrap sample is $Y_i^* = x_i\hat{\beta} + r_i^*$ for $i = 1, \ldots, n$, where $r_1^*, \ldots, r_n^*$ is a random sample from the empirical d.f. based on $(r_1 - \bar{r}_n), \ldots, (r_n - \bar{r}_n)$, where $\bar{r}_n = \sum_{i=1}^n r_i/n$. If $\hat{\beta}_b$ denotes the least square estimate based on Y_i^*s, then the bootstrap expectation of $\hat{\beta}_b$, $\text{E}(\hat{\beta}_b) = \hat{\beta}$, and the bootstrap variance of $\hat{\beta}_b$, $\text{var}^*(\hat{\beta}_b)$ is $n^{-1}\sum_{i=1}^n (r_i - \bar{r}_n)^2/\sum_{j=1}^n x_j^2$ which is equivalent to $n^{-1}\sum_{i=1}^n \sigma_i^2/\sum_{j=1}^n x_j^2$. Therefore, for homoscedastic errors $\sigma_i^2 = \sigma^2$, $\text{var}^*(\hat{\beta}_b)$ is unbiased. But for heteroscedastic errors (unequal σ_i^2 in (2.16)), $\text{var}^*(\hat{\beta}_b)$ is in general biased and inconsistent.

This difficulty with $\text{var}^*(\hat{\beta}_b)$ is due to its very nature. The drawing of i.i.d. samples from $(r_1 - \bar{r}_n), \ldots, (r_n - \bar{r}_n)$ depends on the assumption that the residuals r_i are fairly exchangeable. Any inherent heterogeneity among r_i is lost in the process of i.i.d. sampling and will be be reflected in $\text{var}^*(\hat{\beta}_b)$.

However, it is not difficult to modify this bootstrap to achieve the consistency. Let us form the empirical d.f. on $(x_i/\sqrt{\bar{x_i^2}})(r_i - \bar{r}_n)$ instead of just $r_i - \bar{r}_n$, and let $\tilde{\beta}_b$ denote the resulting bootstrap least square estimate of $\hat{\beta}$. Then under the conditions that $x_i's$ are bounded and $\liminf_{n\to\infty} n^{-1}\sum_{i=1}^n x_i^2 > 0$ (i.e., $\sum_{i=1}^n x_i^2$ grows roughly like n),

$$\text{var}^{**}(\hat{\beta}_b) = \frac{\sum_{i=1}^n x_i^2 r_i^2}{(\sum_{j=1}^n x_j^2)^2} - \frac{n^{-1}(\sum_{i=1}^n x_i r_i)^2}{(\sum_{j=1}^n x_j^2)^2}. \tag{2.17}$$

(** stands for the bootstrap probability under the weighted empirical d.f.), which equals to

$$\frac{\sum_{i=1}^n x_i^2 \sigma_i^2}{(\sum_{j=1}^n x_j^2)^2} + O_p(n^{-3/2}). \tag{2.18}$$

2.2.2 Wu's Bootstrap

As shown in 2.2.1, the method of classical bootstrapping residuals does not adapt well to the possibility of error variance heteroscedasticity. A different method for resampling residuals was proposed in Wu(1986). Define the bootstrap sample

$$Y_1^*, \ldots, Y_n^*, \quad Y_i^* = x_i\hat{\beta} + r_i t_i, \tag{2.19}$$

where $\{t_i\}_i^n$ are i.i.d. with $\mathrm{E}t_i = 0$ and $\mathrm{var}t_i = 1$, for all i.(i.e., r_i is multiplied with an adjusting factor which equals to 1 asymptotically.) Let $\check{\beta}_b$ indicate the least square estimate of $\hat{\beta}$ under the bootstrap (2.19). Then $\mathrm{E}(\check{\beta}_b) = \hat{\beta}$, and

$$\mathrm{var}_t(\check{\beta}_b) = \frac{\sum_{i=1}^n x_i^2 r_i^2 \mathrm{E}t_i^2}{(\sum_{j=1}^n x_j^2)^2} = \frac{\sum_{i=1}^n x_i^2 \sigma_i^2}{(\sum_{j=1}^n x_j^2)^2} + O_p(n^{-3/2}). \tag{2.20}$$

Hence $\mathrm{E}t_i = 0$ and $\mathrm{var}t_i = 1$ for all i are sufficient for proving the consistency of the bootstrap.

2.2.3 The Second Order Asymptotic of Wu's Bootstrap

To study the second order asymptotic of Wu's (1986) bootstrap procedure, Liu (1988) first began by using the idea of the weighted bootstrap in Wu's regression to construct a bootstrap procedure in the context of estimating a common mean from possibly different distribution. Let

$$Y_i = \bar{X}_i + (X_i - \bar{X}_n)t_i, \tag{2.21}$$

where $t_1, \ldots, t_n$ are i.i.d. random variables with mean 0 and variance 1 and are chosen completely independent of data $X_i's$. The distribution of X_i has mean μ_i and variance σ_i^2. The $\mu_i's$ are "roughly" the same while the $\sigma_i^2 s$ are possibly different. The random variables $Y_1, \ldots, Y_n$ from a bootstrap sample. For each i, Y_i has the same mean $\bar{X}_n$ and the variance $var(Y_i) = (X_i - \bar{X}_n)^2$, which reflects the heteroscedasticity of the original data. It is therefore natural to expect that this bootstrap will provide a consistent procedure for estimating the sampling distribution of $\bar{X}_n$ and it will also possess the second order property (one-term Edgeworth correction) of the classical bootstrap, under possibly some further condition on t_i. Liu proposed theorem 5 to establish the consistency of this bootstrap procedure and examined the $n^{-1/2}$-term correction phenomenon afterwards.

Theorem 5. *If* $E|t_1|^3 < \infty$, $E|X_i|^{2+\delta} \le K < \infty$, *for some* $\delta > 0$ *and* $i = 1, \ldots, n$, $\liminf_{n\to\infty}(1/n)\sum_{i=1}^{n}(\mu_i - \bar{\mu}_n)^2 = 0$, *then*

$$\lim_{n\to\infty} \|P^*(\sqrt{n}(\bar{Y}_n - \bar{X}_n) \le x) - P(\sqrt{n}(\bar{X}_n - \bar{\mu}_n) \le x)\|_\infty = 0 \quad a.s. \tag{2.22}$$

For examining the second order asymptotic of this particular bootstrap, Liu considered $\mathrm{E}_t(\sqrt{n}(\bar{Y_n} - \bar{X_n}))^3$, where E_t stands for the expectation w.r.t. $t_i's$ treating $X_i's$ as fixed numbers (i.e., the conditional expectation given $X_1, \ldots, X_n$). Note that

$$\mathrm{E}_t[\sum_{i=1}^{n}(Y_i - \bar{X}_n)]^3 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^3 \mathrm{E}t_i^3. \tag{2.23}$$

Thus, if $\mathrm{E}t_i^3 = 1$,

$$\mathrm{E}_t(\sqrt{n}(\bar{Y}_n - \bar{X}_n))^3 = n^{-1/2}[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^3], \tag{2.24}$$

which is $n^{-1/2}n^{-1}\sum_{i=1}^{n}\mu_{3,i} + o(n^{-1/2})$ assuming that $\mathrm{E}|X_i|^{3+\delta} \le K < \infty$ for some $\delta > 0$ and for all i. This last expression is equal to $\mathrm{E}[n^{1/2}(\bar{X}_n - \bar{\mu}_n)]^3 + o(n^{-1/2})$ if $n^{-1}\sum_{i=1}^{n}|\mu_i - m\bar{u}_n|^3 \to 0$. Therefore, the skewness term in the formal Edgeworth expansions will match if the $t_i's$ have the third central moment equal to 1, besides having mean 0 and variance 1. Moreover, similar calculations show that if $\mathrm{E}t_i^3 = 1$, then the first three cumulants of the studentized statistics $\sqrt{n}(\bar{X}_n - \bar{\mu}_n)/V_n$ and $\sqrt{n}(\bar{Y}_n - \bar{X}_n)/V_n^*$, where $V_n^* = [n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2]^{1/2}$ match up to $o(n^{-1/2})$. Consequently, there is a total $n^{-1/2}$-term corrections by the bootstrap in this case.

Similarly, as for the Wu's bootstrap (2.19) on the simple regression, Liu singled out that under the assumption $\mathrm{E}t_i^3 = 1$, the third moment of $\sqrt{n}(\hat{\beta} - \beta)$ will be estimated correctly up to $O(n^{-1})$ by this bootstrap.

2.3 Pioneer work of the Change-point Detection

Many economic factors may cause a parametric model to be unstable over a period of time. Changes in taste, technical progress, and changes in policies and regulations all are such examples. A change in the economic agent's expectation can induce a change in the reduced-form relationship among economic variables, even though no change in the parameters of the structural relationship is present. As a result, model stability has always been an important concern in econometric and statistical modeling. Most of the aforementioned works deal with (locally) stationary and/or independent data.

2.3.1 The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations

In Lajos Horbath (1993), he proposed a Maximum Likelihood Method to check whether the parameters of normal observations have changed at an unknown point.

Let $X_1, \ldots, X_n$ be independent identical normal random variables with parameters $\mu_i = \mathrm{E}X_i$ and $\sigma^2 = \mathrm{var}X_i$, $i = 1, \ldots, n$. The hypothesis test is

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_n \tag{2.25}$$

against the alternative

$$H_A : \mu_1 = \mu_2 = \ldots = \mu_{[n\tau]} \neq \mu_{[n\tau]+1} = \ldots = \mu_n \tag{2.26}$$

The maximum likelihood method can be used when it is wanted to test H_0 against H_A. It is easy to show that the likelihood ratio is

$$\Lambda_n = \max_{1<k<n-1} \frac{\hat{\sigma}_n^n}{\hat{\sigma}_k^k \breve{\sigma}_{n-k}^{n-k}}, \tag{2.27}$$

where

$$\hat{\sigma}_k^2 = \frac{1}{k} \sum_{1 \le i \le k} (X_i - \hat{X}_k)^2, \quad \breve{\sigma}_{n-k}^2 = \frac{1}{n-k} \sum_{k < i \le n} (X_i - \breve{X}_{nj-k})^2, \tag{2.28}$$

$$\hat{X}_k = \frac{1}{k} \sum 1 \le i \le k X_i, \quad \breve{X}_{n-k} = \frac{1}{n-k} \sum_{k < i \le n} X_i. \tag{2.29}$$

The aim of Horvath's paper was the computation of the asymptotic distribution of

$$\lambda_n = (\max_{1 < k < n-1} (n \log \hat{\sigma}_n^2 - k \log \hat{\sigma}_k^2 - (n-k) \log \breve{\sigma}_{n-k}^2))^{1/2}. \tag{2.30}$$

By introducing the functions $a(x) = (2 \log x)^{1/2}$, $b(x) = 2 \log x + \log \log x$, and the random variable Y satisfying

$$\mathrm{P}(Y \le x) = \exp(-2e^{-x}), \tag{2.31}$$

for all x, Horvath proofed the following based on the limit distribution of the largest deviation between a d-dimensional Ornstein-Uhlenbeck process and the origin.

Theorem 6. *If H_0 holds, then as $n \to \infty$,*

$$a(\log n)\lambda_n - b(\log n) \to_D Y. \tag{2.32}$$

2.3.2 Testing for Changes in Polynomial Regression

In Aue, Alexander and Horvath (2008), they focused on the (nonlinear) polynomial regression model

$$y_i = x_i^T \beta_i + \varepsilon_i, i = 1, \ldots, n, \tag{2.33}$$

where $\{\varepsilon_i\}$ is a sequence of independent, identically distributed normal variables with same variance σ^2. $\{\beta_i\}$ are $(p+1)$-dimensional deterministic vectors and $x_i = (1, i/n, \ldots, (i/n)^p)^T$.

In this setting, they are interested in testing the null hypothesis of structural stability against the alternative of a regime switch at an unknown time, that is,

$$H_0 : \beta_i = \beta_0, i = 1, \ldots, n. \tag{2.34}$$

v.s.

$$H_A : \beta_i = \beta_0, i = 1, \ldots, k^*, \beta_i = \beta_A, i = k^* + 1, \ldots, n, \quad \text{with} \quad \beta_0 \neq \beta_A. \tag{2.35}$$

If the variance σ^2 is known, twice the logarithm of the likelihood ratio for this two-sample problem is given by

$$\ell_n(k) = n\sigma^{-2}(\hat{\sigma}_n^2 - [\hat{\sigma}_{k,1}^2 + \hat{\sigma}_{k,2}^2]) \tag{2.36}$$

where $n\hat{\sigma}_{k,1}^2 = \sum_{i=1}^{k}(y_i - x_i^T\hat{\beta}_k)^2$ and $n\hat{\sigma}_{k,2}^2 = \sum_{i=k+1}^{n}(y_i - x_i^T\hat{\beta}_k^*)^2$ denote the sum of the squared residuals for the first, respectively, second sample, and $n\hat{\sigma}_n^2 = \sum_{i=1}^{n}(y_i - x_i^T\hat{\beta}_n)^2$ the sum of the squared residuals for the whole sample. Therein, $\hat{\beta}_k$ and $\hat{\beta}_k^*$ are the the least squares estimators for β based on the first k and the last $n - k$ observations. Elementary algebra implies that, under H_0, we have

$$\hat{\sigma}_n^2 - [\hat{\sigma}_{k,1}^2 + \hat{\sigma}_{k,2}^2] = \frac{1}{n}S_k^T C_k^{-1} C_n \widetilde{C}_k^{-1} S_k, \tag{2.37}$$

where

$$C_k = \sum_{i=1}^{k} x_i x_i^T, \quad \widetilde{C}_k = \sum_{i=k+1}^{n} x_i x_i^T \tag{2.38}$$

and

$$S_k = \sum_{i=1}^{k} x_i y_i - C_k C_n^{-1} \sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{k} x_i (y_i - x_i^T \hat{\beta}_n). \quad (2.39)$$

Since, in general, the time of change k^* is unknown, we reject the null hypothesis H_0 for large values of

$$T_n = \frac{1}{\sigma^2} \max_{p<k<n-p} S_k^T C_k^{-1} C_n \widetilde{C}_k^{-1} S_k. \quad (2.40)$$

If the common variance σ^2 is unknown, the resulting likelihood ratio is

$$\hat{\ell}_n(k) = (\hat{\sigma}_n^2 - [\hat{\sigma}_{k,1}^2 + \hat{\sigma}_{k,2}^2])^{n/2} \quad (2.41)$$

in the case where the time of change is exactly $k = k^*$. If k^* is unknown, the stability of the regression coefficients is rejected for large values of

$$\hat{T}_n = \max_{p<k<n-p} [-2 \log \hat{\ell}_n(k)]. \quad (2.42)$$

Hansen (2000) studied the asymptotic distribution of the trimmed version

$$T_{n,\delta} = \frac{1}{\hat{\sigma}^2} \max_{\lfloor n\delta \rfloor < k < n - \lfloor n\delta \rfloor} S_k^T C_k^{-1} C_n \widetilde{C}_k^{-1} S_k. \quad (2.43)$$

with some $0 < \delta < 1$.

The limit distribution of $T_{n,\delta}$ is the supreme of quadratic forms of $(p+1)$-dimensional Gaussian processes with a complicated covariance structure.

In contrast, the definition of T_n allows, at least in principle, to detect changes anywhere in the sample, and asymptotic critical values can be easily computed using Theorem 7 below.

Theorem 7. *Assume that the errors have constant variance and are uncorrelated, and that there are two independent standard Brownian motions* $W_{1,n}(s) : s \geq 0$ *and* $W_{2,n}(s) : s \geq 0$*, such that*

$$\max_{1\le k\le n/2} \frac{1}{k^{1/2-\Delta}} |\sum_{i=1}^{k} \varepsilon_i - \sigma W_{1,n}(k)| = O_P(1) \quad (n \to \infty) \tag{2.44}$$

and

$$\max_{n/2\le k\le n} \frac{1}{k^{1/2-\Delta}} |\sum_{i=k+1}^{n} \varepsilon_i - \sigma W_{2,n}(n-k)| = O_P(1) \quad (n \to \infty), \tag{2.45}$$

then, under H_0, the statistic T_n satisfies, for all x

$$\begin{aligned} & \lim_{n\to\infty} P(T_n \le x + 2\log\log n + (p+1)\log\log n - 2\log(\frac{2^{(p+1)/2}\Gamma((p+1)/2)}{p+1})) \\ = \; & \exp(-2e^{-x/2}). \end{aligned} \tag{2.46}$$

The resulting test in finite sample is easy to apply and has good size and power, even in small samples.

2.3.3 Monitoring Shifts in Mean: Asymptotic Normality of Stopping Times

In Horvath (2008), they considered a sequential procedure designed to detect a possible change in the mean of a time series, which is defined as

$$X_i = \begin{cases} \varepsilon_i + \mu, & 1 \le i < m + k^*, \\ \varepsilon_i + \mu + \Delta_m, & m + k^* \le i < \infty. \end{cases} \tag{2.47}$$

where $k^* \ge 1$ and $\{\varepsilon_i\}$ is an error sequence to be specified below. If $\Delta_m \ne 0$, then a change in mean occurs at time $m+k^*$.The means before and after the possible change are unknown and k^*, the time of the possible change, is also not given. They were interested in on-line monitoring for a change point k^*. In their monitoring procedure, the first m observations are used as a training sample and asymptotic are established as m tends to infinity. They showed that the detection time is asymptotically normal as $m \to \infty$, based on the CUSUM (cumulative sum) statistic.

The stopping time in Horvath (2004) was defined as

$$\tau_m = \inf\{k \geq 1 : \Gamma(m,k) \geq g(m,k)\}, \tag{2.48}$$

with the understanding $\tau_m = \infty$ if $\Gamma(m,k) < g(m,k)$ for all $k = 1, 2, \ldots$. The dectector $\Gamma(m,k)$ and the boundary function $g(m,k)$ were chosen so that under the null hypothesis ($\Delta = 0$),

$$\lim_{m\to\infty} \mathrm{P}(\tau_m < \infty) = \alpha, \tag{2.49}$$

where $0 < \alpha < 1$ is a prescribed number, and under the alternative ($\Delta \neq 0$),

$$\lim_{m\to\infty} \mathrm{P}(\tau_m < \infty) = 1, \tag{2.50}$$

By (2.49), as in the Neyman-Pearson test, the probability of a false alarm is asymptotically α.

The detector used was defined as

$$\Gamma(m,k) = \hat{Q}(m,k) = \frac{1}{\hat{\sigma}_m}|\sum_{m<i\leq m+k}(X_i - \bar{X}_m)|, \tag{2.51}$$

where

$$\bar{X}_m = \frac{1}{m}\sum_{1\leq i\leq m} X_i \tag{2.52}$$

and where $\hat{\sigma}_m^2$ is an asymptotically consistent estimator for

$$\sigma^2 = \lim_{n\to\infty} \mathrm{Var}(\sum_{1\leq i\leq n} \varepsilon_i)/n. \tag{2.53}$$

The boundary function was chosen as

$$g(m,k) = cm^{1/2}(1+\frac{k}{m})(\frac{k}{m+k})^{\gamma}, \quad 0 \le \gamma < 1/2. \tag{2.54}$$

Under regularity conditions and different assumptions on the errors ε_i, Horvanth et al.(2004) and Aue et al.(2006b) proved that under H_0 ($\Delta_m = 0$),

$$\lim_{m\to\infty} \mathrm{P}(\tau_m < \infty) = \mathrm{P}(\sum_{0<t\le 1} \frac{|W(t)|}{t^{\gamma}} \ge c), \tag{2.55}$$

where $\{W(t), 0 \le t < \infty\}$ is a standard Wiener process (Brownian motion). Aue and Horvath (2004) showed that

$$\lim_{m\to\infty} \mathrm{P}(\frac{\tau_m - a_m}{b_m} \le x) = \Phi(x), \tag{2.56}$$

where Φ denotes the standard normal distribution function, $a_m = (\frac{cm^{\frac{1}{2}-\gamma}}{|\Delta_m|})^{\frac{1}{1-\gamma}}$ and $b_m = \frac{\sigma}{1-\gamma}\frac{1}{|\Delta_m|}a_m^{1/2}$.

In Horvath (2007), they proposed the modification of τ_m as

$$\tau_m^* = \inf\{k : 1 \le k \le N, \hat{Q}(m,k) \ge c(m;t)g^*(m,k)\}, \tag{2.57}$$

with $\tau^* = N$ if $\hat{Q}(m,k) < c(m;t)g^*(m,k)$ for all $1 \le k \le N$, where $N = N(m)$ depends on the training sample size. The functions $g^*(m,k)$ and $c(m;t)$ are defined by

$$g^*(m,k) = m^{1/2}(1+\frac{k}{m})(\frac{k}{m+k})^{1/2} \tag{2.58}$$

and

$$c(m;t) = \frac{t + D(\log m)}{A(\log m)}, \tag{2.59}$$

with $A(x) = (2\log x)^{1/2}$, $D(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log\pi$.

The main result of Horvath et al.(2007) is

Theorem 8. *If $\{\varepsilon_i, 1 \le i < \infty\}$ are independent, identically distributed random variables with $E\varepsilon_i = 0$, $E\varepsilon_i^2 = \sigma^2 > 0$, $E|\varepsilon_i|^\nu < \infty$ with some $\nu > 2$, $\Delta_m = 0$ and*

$$c_1 m \le N \le c_2 m^\lambda \tag{2.60}$$

with some $c_1, c_2 > 0$ and $1 \le \lambda < \infty$, then

$$\lim_{m\to\infty} P(\hat{Q}(m,k) < c(m;t)g^*(m,k) \quad for \quad all 1 \le k \le N) = \exp(-e^{-t}). \tag{2.61}$$

The limit result in theorem (8) is a Darling-Erdos type extreme value limit theorem and it is related to the distribution of $W(t)/t^{1/2}$. To apply it, they chose t in (2.59) such that $\exp(-e^{-t}) = 1 - \alpha$. This implies that under the null hypothesis

$$\lim_{m\to\infty} \mathrm{P}(\tau_m^* < \infty) = \lim_{m\to\infty} (\text{"}stopping \quad in \quad N \quad steps\text{"}) = \alpha. \tag{2.62}$$

In order to accommodate applications, Horvath (2008) required much weaker conditions than the indepence of the innovations $\varepsilon_1, \varepsilon_2, \ldots$. Namely, they assumed that $\varepsilon_1, \varepsilon_2, \ldots$ from a strictly stationary sequence with $\mathrm{E}\varepsilon_1 = 0$ and $\mathrm{E}\varepsilon_1^2 < \infty$,

$$\max_{1\le k\le n} \frac{1}{k^{1/2}} |\sum_{i=1} k\varepsilon_i| = O_p((\log\log n)^{1/2}) \quad (n \to \infty), \tag{2.63}$$

$$\frac{1}{\sigma n^{1/2}} \sum_{1\le i\le nt} \varepsilon_i \to^{D[0,1]} W(t) \quad (n \to \infty), \tag{2.64}$$

where $\{W(t), 0 \le t \le \infty\}$ denotes a standard Wiener process.

Thus, an extension of theorem (8) is as following.

Theorem 9. *Suppose $\varepsilon_1, \varepsilon_2, \ldots$ satisfy assumptions (2.63) and (2.64). If $\Delta_m = 0$ and the condition (2.60) is satisfied then the relation (2.61) holds.*

Assuming $c_3(\log m)^{-\beta} \leq |\Delta_m| \leq c_4$ for some $c_3, c_4 > 0$ and $\beta > 0$ and

$$k^* = o(\frac{(\log\log m)^{1/2}}{\Delta_m^2}), \tag{2.65}$$

Horvath proposed the theorem (10) below for the asymptotic normality of τ_m^*

Theorem 10. *If (2.63-2.65) are satisfied, then for all x,*

$$\lim_{m\to\infty} P(\frac{\Delta_m^2}{2(2\log\log m)^{1/2}}(\frac{\tau_m^*}{\sigma^2} - \frac{2\log\log m}{\Delta_m^2}) \leq x) = \Phi(x). \tag{2.66}$$

If in addition, $\hat{\sigma} - \sigma = o_p(1)$ as $m \to \infty$, , then

$$\lim_{m\to\infty} P(\frac{\Delta_m^2}{2(2\log\log m)^{1/2}}(\frac{\tau_m^*}{\hat{\sigma}_m^2} - \frac{2\log\log m}{\Delta_m^2}) \leq x) = \Phi(x). \tag{2.67}$$

2.3.4 Testing for Changes in the Covariance Structure of Linear Processes

In the literature researchers have realized the importance of non-constant variances. In Berkes, Istvan and Gombay (2009), they considered several procedures to detect changes in the mean or the covariance structure of a linear process. The tests are based on the weighted CUSUM process, which are defined as

$$M_n(t) = \begin{cases} n^{-1/2}\sum_{1\leq i\leq (n+1)t}(X_i - \bar{X}_n), & \text{if} \quad 0 \leq t < 1 \\ 0, & \text{if} \quad t = 1 \end{cases} \tag{2.68}$$

for testing the change point of sample mean, and

$$M_n^{(r)}(t) = \begin{cases} n^{-1/2}\sum_{1\leq i\leq (n+1)t}((X_i - \bar{X}_n)(X_{i-r} - \bar{X}_n) - \bar{X}_n^{(r)}), & \text{if} \quad 0 \leq t < 1 \\ 0, & \text{if} \quad t = 1, \end{cases} \tag{2.69}$$

for testing the possible change in the correlation structure.

Consider the asymptotics for $M_n(t)$ when the observations under the null hypothesis are given by a linear process

$$X_k = \mu + \sum_{0 \le i < \infty} a_i \varepsilon_{k-i}. \tag{2.70}$$

Throughout the paper, they assumed that

$$\sum_{0 \le j < \infty} a_j \neq 0, \tag{2.71}$$

and that ε_i are i.i.d.random variables with $E\varepsilon_i = 0$ and $0 < \sigma^2 = E\varepsilon_i < \infty$, $E|\varepsilon_0|^k < \infty$ with some $k > 2$.

It is well established in the literature that using a weight function $w(t)$, one will increase the power of testing procedures against certain alternatives.

Let w be a function on $(0,1)$, increasing in a neighborhood of 0, decreasing in a neighborhood of 1, and

$$\inf_{c \le t \le 1-c} w(t) > 0, \quad 0 < c < 1/2. \tag{2.72}$$

Let

$$A(x) = (2\log x)^{1/2} \quad \text{and} \quad D(x) = 2\log x + \frac{1}{2}\log\log x - \frac{1}{2}\log \pi \tag{2.73}$$

and

$$I(w,c) = \int_0^1 \frac{1}{t(1-t)} \exp{(-\frac{cw^2(t)}{t(1-t)})} dt \tag{2.74}$$

They developed new strong and weak approximations for the sample mean as well as the sample correlations of linear processes as follow.

Theorem 11. *We assume that* $\sum_{1 \le j < \infty} j|a_j| < \infty$.

(i) If $I(w,c) < \infty$ with some $c > 0$, then

$$\frac{1}{\tau}\sup_{0<t<1}\frac{M_n(t)}{w(t)} \to^D \sup_{0<t<1}\frac{B(t)}{w(t)} \tag{2.75}$$

and

$$\frac{1}{\tau}\sup_{0<t<1}\frac{|M_n(t)|}{w(t)} \to^D \sup_{0<t<1}\frac{|B(t)|}{w(t)} \tag{2.76}$$

where $B(t), 0 \le t \le 1$ denotes a Brownian bridge.

(ii) We have

$$\lim_{n\to\infty} P\{A(\log n)\frac{1}{\tau}\sup_{0<t<1}\frac{|M_n(t)|}{(t(1-t))^{1/2}} \le x + D(\log n)\} = \exp\left(-2e^{-x}\right) \tag{2.77}$$

for all x, where $A()$ and $D()$ are defined above.

Theorem 12. *(i) If $I(w,c) < \infty$ with some $c > 0$, then*

$$\frac{1}{\gamma(r)}\sup_{0<t<1}\frac{M_n^r(t)}{w(t)} \to^D \sup_{0<t<1}\frac{B(t)}{w(t)} \tag{2.78}$$

and

$$\frac{1}{\gamma(r)}\sup_{0<t<1}\frac{|M_n(t)^r|}{w(t)} \to^D \sup_{0<t<1}\frac{|B(t)|}{w(t)} \tag{2.79}$$

where $B(t), 0 \le t \le 1$ denotes a Brownian bridge and

$$0 < \gamma^2(r) = \lim_{n\to\infty}\frac{1}{n}var(\sum_{1\le i\le n}(X_i-\mu)(X_{i-r}-\mu)) < \infty. \tag{2.80}$$

(ii) We have

$$\lim_{n\to\infty} P\{A(\log n)\frac{1}{\gamma(r)}\sup_{0<t<1}\frac{|M_n^{(r)}(t)|}{(t(1-t))^{1/2}} \le x + D(\log n)\} = \exp\left(-2e^{-x}\right) \tag{2.81}$$

for all x, where $A()$ and $D()$ are defined above.

2.3.5 Wu and Zhao's Work on Change-point for Stationary Data

An important problem in time series analysis is the estimation of trends. Assume that the data $X_1, \ldots, X_n$ are observed from the model

$$X_k = \mu(k/n) + e_k, \quad k = 1, \ldots, n. \tag{2.82}$$

where μ is an unknown regression function defined on $[0, 1]$ and e_k is a mean 0 stationary process. The process X_k is mean non-stationary and can be interpreted as a signal μ plus noise e_k model. Wu and Zhao (2007) developed the statistical procedures to test whether the trend μ in model (2.82) has structural breaks or jumps. They also constructed the simultaneous confidence bands (SCBs) for μ if the curve μ is smooth.

The Error Structure

In their paper, Wu and Zhao assumed that the error process e_i in model (2.82) is stationary and causal. Let $\varepsilon_i, i \in \mathrm{Z}$, be independent and identically distributed (IID) random variables and $\mathbf{G}$ a measurable function such that

$$e_i = \mathbf{G}(\ldots, \varepsilon_{i-1}, \varepsilon_i) \tag{2.83}$$

is a proper random variable with mean 0 and finite variance. Let ε'_j be an IID copy of ε_j and $e_i^* = \mathbf{G}(\ldots, \varepsilon_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_i)$. Assume that $\mathrm{E}(|e_i|^p) < \infty, p > 2$ and

$$\sum_{i=1}^{\infty} i \|e_i - e_i^*\|_p < \infty. \tag{2.84}$$

Wu (2006) established the following strong approximation or strong invariance principle. Under inequality (2.84),there is a standard Brownian motion B such that, on a richer probablitity space, S_i can be uniformly approximated by $\sigma \mathrm{B}(i)$:

$$\max |S_i - \sigma \mathrm{B}(i)| = o_{\mathrm{AS}}\{n^{1/p'} \log(\log n)\}, \quad p' = \min(4, p), \tag{2.85}$$

where $\sigma^2 = \sum_{k \in \mathrm{Z}} \mathrm{E}(e_0 e_k)$ is the long run variance.

Inference of Structural Breaks

For model(2.82), the classical changepoint analysis concerns testing the null hypothesis $\mu_1 = \ldots = \mu_n$ against the alternative of one or multiple changepoints,

$$\mu_1 = \ldots = \mu_{k_1} \neq \mu_{k_1+1} = \ldots = \mu_{k_2} \neq \mu_{k_2+1} = \ldots \neq \mu_{k_J+1} = \ldots = \mu_n, \tag{2.86}$$

where $\mu_k = \mu(k/n)$ and $k_1, \ldots, k_J$ are called changepoints. The alternative hypothesis says that μ is piecewise constant. To identify the changepoint, they proposed to compare the local averages of X_j over $nt < j < nt + k_n$ and over $nt - k_n < j < nt$, where k_n is the block length satisfying $k_n \to \infty$ and $k_n/n \to 0$. If the two averages are close, then t is unlikely to be a discontinuous point. A global measure of the discrepancy is

$$D_n^* = 1/k_n \max_{k_n \leq i \leq n-k_n} |\sum_{j=i+1}^{k_n+1} X_j - \sum_{j=i-k_n+1}^{i} X_j|. \tag{2.87}$$

A non-overlapping version of D_n^* is given by

$$D_n = \max_{1 \leq i \leq m-1} |A_i - A_{i-1}| \quad A_k = A_{i,n} = 1/k_n \sum_{j=1}^{k_n} X_{j+ik_n}. \tag{2.88}$$

Here $m = \lfloor n/k_n \rfloor$ is the largest integer not exceeding n/k_n. Let $\omega_k = \mathrm{E}(A_i)$. For $1 \leq i \leq m$ let the interval $I_i = (ik_n/n, (i+1)k_n/n]$. If there is changepoint of μ, we expect either $|A_i - A_{i-1}|$ or $|A_{i+1} - A_i|$ would take large values. So D_n can also be used to detect the changepoint of μ.

The distributions of D_n and D_n^* are shown in the following theorem.

Theorem 13. *Assume that* $\mu \in L[0,1]$ *and*

$$k_n^{-1} n^{1/2} \log n^3 + n^{-2/3} \log n^{1/3} k_n \to 0. \tag{2.89}$$

Let $\gamma_m = [4\log(m) - 2\log(\log(m))]^{1/2}$. *Then we have*

$$\sqrt{\log(m)}(k_n^{1/2}\sigma^{-1}D_n - \gamma_m) \Rightarrow V \tag{2.90}$$

and

$$\sqrt{\log(m)k_n^{1/2}\sigma^{-1}D_n^*} - [2\log(m) + 1/2\log\log(m)] - \log(3) \Rightarrow V. \tag{2.91}$$

where V *has the extreme value distribution* $P(V \leq x) = \exp\left(-\pi^{-1/2}\exp(-x)\right)$.

Theorem 19 is not yet directly applicable since the long run variance σ^2 is typically unknown and it needs to be estimated. Several estimates of σ^2 will be proposed in the next section.

Estimating σ

To apply theorem 19, we should deal with the crucial issue of estimating the long run variance σ^2. Recall the definition of A_i. For a real sequence $a_1, \ldots, a_k$, denote its median by median $(a_1, \ldots, a_n)$, three asymptotically consistent estimates were considered:

$$\widehat{\sigma}_1 = \frac{\sqrt{\pi k_n}}{2(m-1)} \sum_{i=1}^{m-1} |A_i - A_{i-1}|, \tag{2.92}$$

$$\widehat{\sigma}_2 = \frac{\sqrt{k_n}}{\sqrt{(2u_{1/4})}} median(|A_i - A_{i-1}|, 1 \leq i \leq m-1) \tag{2.93}$$

$$\widehat{\sigma}_3 = \frac{\sqrt{k_n}}{\sqrt{2(m-1)}} (\sum_{i=1}^{m-1} |A_i - A_{i-1}|^2)^{1/2}. \tag{2.94}$$

In $\widehat{\sigma}_2, u_{1/4} = 0.674$ is the third quartile of the standard normal distribution. Carlstein (1986) considered strong mixing processes by using non-overlapping blocks. $\widehat{\sigma}_3$ is closely related to Carlsteins subseries variance estimate.

Theorem 14. *Assume that* $\mu \in L[0,1]$.

(a) Let $k_n \asymp n^{5/8}$. *Then* $\widehat{\sigma}_1, \widehat{\sigma}_2 = \sigma + O_P\{n^{-1/16}\log(n)\}$.

(b) Let $k_n \asymp n^{1/3}$. *Then* $E(|\widehat{\sigma}_3^2 - \sigma^2|^2) = O(n^{-2/3})$.

Chapter 3

Inference for Modulated Stationary Processes

[1] In this chapter, we study several statistical inference problems for the non-stationary time series

$$X_i = \mu + \sigma_i e_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where e_i are stationary time series with zero mean, and $\sigma_i > 0$ are unknown constants adjusting for time-dependent variances.

For two sequences $\{a_n\}$ and $\{b_n\}$, write $a_n = O(b_n)$, $a_n = o(b_n)$, and $a_n \asymp b_n$, respectively, if $|a_n/b_n| < c_1$, $a_n/b_n \to 0$, and $c_2 < |a_n/b_n| < c_3$, for some constants $0 < c_1, c_2, c_3 < \infty$. For a random variable e and a real number $q > 0$, write $e \in \mathcal{L}^q$ if $\|e\|_q := \{E(|e|^q)\}^{1/q} < \infty$.

[1] This chapter is based on the article of Zhao and Li (2013).

3.1 Uniform Approximations for Non-stationary Process

In (1.1), assume without loss of generality that $E(e_i) = 0$ and $E(e_i^2) = 1$ so that $\{e_i\}$ and $\{e_i^2 - 1\}$ are centered stationary processes. With the convention $S_0 = S_0^* = 0$, define

$$S_i = \sum_{j=1}^{i} e_j \qquad \text{and} \qquad S_i^* = \sum_{j=1}^{i} (e_j^2 - 1), \qquad i = 1, 2, \ldots. \tag{3.2}$$

Assumption 1. Assume there exist standard Brownian motions $\{B_t\}$ and $\{B_t^*\}$ such that

$$\max_{1 \le i \le n} |S_i - \tau B_i| = o_{\text{a.s.}}(\Delta_n) \quad \text{and} \quad \max_{1 \le i \le n} |S_i^* - \tau^* B_i^*| = o_{\text{a.s.}}(\Delta_n), \tag{3.3}$$

where Δ_n is the approximation error, τ^2 and τ^{*2} are the long-run variances of $\{e_i\}$ and $\{e_i^2 - 1\}$, respectively. Further assume $\tau^2 > 0$ to avoid the degenerate case $\tau^2 = 0$.

The uniform approximations in (4.13) are generally called strong invariance principle. For independent random variables e_i, the celebrated "Hungarian embedding" asserts that (4.13) holds with the optimal rate $\Delta_n = n^{1/4}$. To see how to use (4.13), under H_0 in (4.2) consider

$$F_j = j(\underline{X}_j - \mu) \quad \text{and} \quad \underline{V}_j^2 = \sum_{i=1}^{j} (X_i - \underline{X}_j)^2, \quad \text{where} \quad \underline{X}_j = j^{-1} \sum_{i=1}^{j} X_i. \tag{3.4}$$

Theorem 15 below presents uniform approximations for F_j and V_j^2. For $\{\sigma_i\}$ in (1.1), define

$$r_n = |\sigma_n| + \sum_{i=2}^{n} |\sigma_i - \sigma_{i-1}| \qquad \text{and} \qquad r_n^* = |\sigma_n^2| + \sum_{i=2}^{n} |\sigma_i^2 - \sigma_{i-1}^2|, \tag{3.5}$$

$$\Sigma_j^2 = \sum_{i=1}^{j} \sigma_i^2 \qquad \text{and} \qquad \Sigma_j^{*2} = \Big(\sum_{i=1}^{j} \sigma_i^4 \Big)^{1/2}. \tag{3.6}$$

Theorem 15. *Assume (4.13) holds. For any fixed $c \in (0, 1]$, the uniform approximations* *hold*

$$\max_{cn \le j \le n} \Big| F_j - \tau \sum_{i=1}^{j} \sigma_i (B_i - B_{i-1}) \Big| = O_{\text{a.s.}}(r_n \Delta_n), \tag{3.7}$$

$$\max_{cn \le j \le n} \Big| \underline{V}_j^2 - \Sigma_j^2 \Big| = O_{\text{p}}\{(r_n^2 \Delta_n^2 + \Sigma_n^2)/n + \Sigma_n^{*2} + r_n^* \Delta_n\}. \tag{3.8}$$

Theorem 15 provides quite general results under the mild condition (4.13). We now discuss sufficient conditions for (4.13). Shao (1993) obtained sufficient mixing conditions for (4.13). In this article, we briefly introduce the framework in Wu (2007). Assume that e_i has the causal representation $e_i = G(\ldots, \varepsilon_{i-1}, \varepsilon_i)$, where ε_i are independent and identically distributed, and G is a measurable function such that e_i is well-defined. Proposition 1 below follows from Corollary 4 in Wu (2007).

Proposition 1. *Let $\{\varepsilon_i'\}_{i \in \mathbb{Z}}$ be an independent copy of $\{\varepsilon_i\}_{i \in \mathbb{Z}}$. Define the coupling process $e_i^* = G(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_{i-1}, \varepsilon_i)$. Assume that $\sum_{i=1}^{\infty} i \|e_i - e_i^*\|_8 < \infty$. Then (4.13) holds with $\Delta_n = n^{1/4} \log(n)$, the optimal rate up to a logarithm factor.*

For linear process $e_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}$ with $\varepsilon_i \in \mathcal{L}^8$ and $E(\varepsilon_i) = 0$, $\|e_i - e_i^*\|_8 = \|\varepsilon_0 - \varepsilon_0^*\|_8 |a_i|$. If $\sum_{i=1}^{\infty} i|a_i| < \infty$, then (4.13) holds with $\Delta_n = n^{1/4} \log(n)$. For many nonlinear time series, $\|e_i - e_i^*\|_8$ decays exponentially fast and hence (4.13) holds with $\Delta_n = n^{1/4} \log(n)$; see Section 3.1 of Wu (2007). From now on we assume (4.13) holds with $\Delta_n = n^{1/4} \log(n)$.

As shown in Examples 1–3 below, r_n and r_n^* in (3.5) often have tractable bounds.

Example 1. If σ_i is non-decreasing in i, then $\sigma_n \le r_n \le 2\sigma_n$ and $\sigma_n^2 \le r_n^* \le 2\sigma_n^2$. If σ_i is non-increasing in i, then $r_n = \sigma_1$ and $r_n^* = \sigma_1^2$. If σ_i are piecewise constants, $r_n, r_n^* = O(1)$.

Example 2. Let $\sigma_i = s(i/n^\gamma)$ for $\gamma \in [0, 1]$ and a Lipschitz continuous function $s(t), t \in [0, \infty), \sup_{t \in [0,\infty)} s(t) < \infty$. Then $r_n, r_n^* = O(n^{1-\gamma})$. If $\gamma = 1$, we obtain locally stationary case with time window $i/n \in [0, 1]$; if $\gamma \in [0, 1)$, we have infinite time window $[0, \infty)$ as $n/n^\gamma \to \infty$, which may be more reasonable for data with a long time horizon.

Example 3. If $\sigma_i = i^\beta L(i)$ for a slowly varying function $L(\cdot)$ such that $\frac{L(cx)}{L(x)} \to 1$ as $x \to \infty$ for all $c > 0$. Then we can show that $r_n = O\{n^\beta L(n)\}$, $O\{\log(n)L(n)\}$, $O(1)$ and $r_n^* = O\{n^{2\beta} L^2(n)\}$, $O\{\log(n)L^2(n)\}$, $O(1)$, depending on whether $\beta > 0, \beta = 0, \beta < 0$.

3.2 Self-normalized Central Limit Theorem

Recall the sample average $\bar{X}$ in (3.4). In this section we establish a self-normalized CLT for $\bar{X}$ and use it to construct asymptotic confidence interval for μ. To understand how non-stationarity makes this problem difficult, elementary calculation shows

$$\mathrm{Var}\{\sqrt{n}(\bar{X} - \mu)\} = \frac{\gamma_0}{n}\sum_{i=1}^n \sigma_i^2 + \frac{2}{n}\sum_{1 \le i < j \le n} \sigma_i \sigma_j \gamma_{j-i} := \tau_n^2, \quad \text{where} \quad \gamma_k = \mathrm{Cov}(e_0, e_k).$$

In the stationary case $\sigma_i \equiv 1$, under condition $\sum_{k=0}^\infty |\gamma_k| < \infty$, $\tau_n^2 \to \tau^2$, the long-run variance in (1.2). For non-stationary case, it is difficult to deal with τ_n^2 directly, due to the large number of unknown parameters and complicated structure.

To attenuate the aforementioned problem, we apply the uniform approximations in Theorem 15. Assume that (3.9) below holds. Note that the increments $B_i - B_{i-1}$ of standard Brownian motions are independent standard normal random variables. By (3.7), $n(\bar{X} - \mu)$ is equivalent to $N(0, \Sigma_n^2)$ in distribution. By (3.8), $\underline{V}_n/\Sigma_n \to 1$ in probability. By Slutsky's theorem, Proposition 2 follows.

Proposition 2. *Let (4.13) hold with $\Delta_n = n^{1/4}\log(n)$. For $r_n, r_n^*, \Sigma_n^2, \Sigma_n^{*2}$ in (3.5)– (3.6), assume*

$$\delta_n = r_n \Delta_n / \Sigma_n + (r_n^* \Delta_n + \Sigma_n^{*2})/\Sigma_n^2 \to 0. \tag{3.9}$$

Recall $\underline{V}_n^2$ in (3.4). Then as $n \to \infty$, $n(\bar{X} - \mu)/\underline{V}_n \Rightarrow N(0, \tau^2)$. Consequently, a $(1-\alpha)$ asymptotic confidence interval for μ is $\bar{X} \pm z_{\alpha/2}\hat{\tau}\underline{V}_n/n$, where $\hat{\tau}$ is a consistent estimate of τ (Section 3.5 below), and $z_{\alpha/2}$ is $(1-\alpha/2)$ standard normal quantile.

Proposition 2 is an extension of the classical CLT for IID data or stationary process to non-stationary process. If X_i are IID, then $n(\bar{X} - \mu)/\underline{V}_n \Rightarrow N(0, 1)$. In Proposition 2, τ^2

can be viewed as the variance inflation factor due to the dependence of $\{e_i\}$. For stationary data, the sample variance $\underline{V}_n^2/n$ is a consistent estimate of the population variance. Here, for non-stationary case (1.1), by (3.8) in Theorem 15, $\underline{V}_n^2/n$ can be viewed as an estimate of the time-average "population variance" Σ_n^2/n. So, we can interpret the CLT in Proposition 2 as a self-normalized CLT for non-stationary time series with the self-normalizing term $\underline{V}_n$ adjusting for non-stationarity due to $\sigma_1, \ldots, \sigma_n$ and τ^2 accounting for dependence of $\{e_i\}$. Clearly, parameters $\sigma_1, \ldots, \sigma_n$ are canceled out through self-normalization. Finally, condition is satisfied for Example 2 with $\gamma > 3/4$ and Example 3 with $\beta > -1/4$.

In classical statistics, width of confidence intervals usually shrinks as sample size increases. By Proposition 2 and Theorem 15, the width of the constructed confidence interval for μ is proportional to $\underline{V}_n/n$ or equivalently Σ_n/n. Thus, a necessary and sufficient condition for shrinking confidence interval is $\sum_{i=1}^n \sigma_i^2/n^2 \to 0$, which is satisfied if $\sigma_i = o(\sqrt{i})$. An intuitive explanation is as follows. For IID data, sample mean converges at rate $O(\sqrt{n})$. For non-stationary case, if σ_i grows faster than $O(\sqrt{i})$, the contribution of a new observation is negligible relative to its noise level.

Example 4. If $\sigma_i \asymp i^\beta$ with $\beta \in [0, 1/2)$, the length of confidence interval is proportional to $\Sigma_n/n \asymp n^{\beta-1/2}$. In particular, if $c_1 < \sigma_i < c_2$ for some positive constants c_1 and c_2, then Σ_n/n achieves the optimal rate $O(n^{-1/2})$. If $\sigma_i \asymp \log(i)$, then $\Sigma_n/n \asymp \log(n)/\sqrt{n}$.

The same idea can be extended to linear combinations of means over different time periods. Suppose that we have observations from k consecutive time periods $\mathcal{T}_1, \ldots, \mathcal{T}_k$, each of the form (1.1) with different means, denoted by $\mu_1, \ldots, \mu_k$, and time-dependent variances. Let $\nu = \beta_1\mu_1 + \cdots + \beta_k\mu_k$ for given coefficients $\beta_1, \ldots, \beta_k$. For example, if we are interested in mean change from $\mathcal{T}_1$ to $\mathcal{T}_2$, we can take $\nu = \mu_2 - \mu_1$; if we are interested in whether the increase from $\mathcal{T}_3$ to $\mathcal{T}_4$ is larger than that from $\mathcal{T}_1$ to $\mathcal{T}_2$, we can let $\nu = (\mu_4 - \mu_3) - (\mu_2 - \mu_1)$. Proposition 3 below extends Proposition 2 to multiple means.

Proposition 3. *Let $\nu = \beta_1\mu_1 + \cdots + \beta_k\mu_k$. For $\mathcal{T}_j$, denote by n_j its sample size and $\bar{X}(j)$ its sample average. Assume that (3.9) holds for each individual time periods $\mathcal{T}_j$ and for* *simplicity that* $n_1, \ldots, n_k$ *are of the same order. Then*

$$\frac{\sum_{j=1}^{k} \beta_j \bar{X}(j) - \nu}{\Lambda_n} \Rightarrow N(0, \tau^2), \quad \textit{where} \quad \Lambda_n^2 = \sum_{j=1}^{k} \left\{ \frac{\beta_j^2}{n_j^2} \sum_{i \in \mathcal{T}_j} [X_i - \bar{X}(j)]^2 \right\}.$$

3.3 Wild Bootstrap for Self-normalized Statistic

Recall $\sigma_i e_i$ in (1.1).Suppose we are interested in the self-normalized statistic

$$H_n = \frac{\sum_{i=1}^{n} \sigma_i e_i}{\sqrt{\sum_{i=1}^{n} \sigma_i^2 e_i^2}}.$$

For problems with small sample sizes, it is more natural to use bootstrap distribution instead of the convergence $H_n \Rightarrow N(0, \tau^2)$ in Proposition 2. Wu (1986) and Liu (1988) have pioneered the work on wild bootstrap for independent data with non-identical distributions. We shall extend their wild bootstrap procedure to non-stationary time series case (1.1).

Let $\{\alpha_i\}$ be IID random variables independent of $\{e_i\}$ satisfying $\alpha_i \in \mathcal{L}^3, E(\alpha_i) = 0, E(\alpha_i^2) = 1$. Define self-normalized statistic based on new data

$$H_n^* = \frac{\sum_{i=1}^{n} \xi_i}{\sqrt{\sum_{i=1}^{n} (\xi_i - \bar{\xi})^2}}, \quad \text{where} \quad \xi_i = \sigma_i e_i \alpha_i \quad \text{and} \quad \bar{\xi} = \frac{\xi_1 + \cdots + \xi_n}{n}.$$

Clearly, ξ_i inherits the non-stationarity structure of $\sigma_i e_i$ by writing $\xi_i = \sigma_i e_i^*$ with $e_i^* = e_i \alpha_i$. On the other hand, for the new error process $\{e_i^*\}$, $E(e_i^{*2}) = E(e_i^2) = 1$ and $\text{Cov}(e_i^*, e_j^*) = 0$ for $i \neq j$. Thus, $\{e_i^*\}$ is a white noise sequence with long-run variance one. By Proposition 2, the scaled version $H_n / \tau \Rightarrow N(0, 1)$ is robust against the dependence structure of $\{e_i\}$, and hence it is reasonable to expect that H_n^* should be close to H_n / τ in distribution.

Theorem 16. *Let the conditions in Proposition 2 hold. Further assume*

$$\left(\sum_{i=1}^{n} \sigma_i^3 \right)^2 \left(\sum_{i=1}^{n} \sigma_i^2 \right)^{-3} \to 0. \tag{3.10}$$

Let $\hat{\tau}$ be a consistent estimate of τ. Denote by $\mathbb{P}^*$ the conditional law given $\{e_i\}$. Then

$$\sup_{x\in\mathbb{R}} |\mathbb{P}^*(H_n^* \le x) - \mathbb{P}(H_n/\hat{\tau} \le x)| \to 0, \quad \textit{in probability.} \tag{3.11}$$

Theorem 16 asserts that, H_n^* behaves like the scaled version $H_n/\hat{\tau}$, with the scaling factor $\hat{\tau}$ coming from the dependence of $\{e_i\}$. Here we use the sample mean $\bar{X}$ in (1.1) to illustrate a wild bootstrap procedure to obtain the distribution of $n(\bar{X} - \mu)/(\tau \underline{V}_n)$ in Proposition 2.

(i) Apply method in Section 3.5 to $X_1, \ldots, X_n$ to obtain a consistent estimate $\hat{\tau}$ of τ.

(ii) Subtract the sample mean $\bar{X}$ from data to obtain $\epsilon_i = X_i - \bar{X}, i = 1, \ldots, n$.

(iii) Generate IID random variables $\alpha_1, \ldots, \alpha_n$ satisfying $E(\alpha_i) = 0, E(\alpha_i^2) = 1$.

(iv) Based on ϵ_i in (ii) and α_i in (iii), generate bootstrap data $\xi_i^b = \epsilon_i \alpha_i$, and compute

$$H_n^b = \frac{\sum_{i=1}^n \xi_i^b}{\hat{\tau}^b \sqrt{\sum_{i=1}^n (\xi_i^b - \bar{\xi}^b)^2}},$$

where $\hat{\tau}^b$ is a long-run variance estimate (see Section 3.5) for bootstrap data ξ_i^b.

(vi) Repeat (ii)-(iv) a large number of times and use the empirical distribution of those realizations of H_n^b as the distribution of $n(\bar{X} - \mu)/(\tau \underline{V}_n)$.

The proposed wild bootstrap is an extension of that in Liu (1988) for independent data to non-stationary time series, and it has two appealing features. First, the scaling factor $\hat{\tau}$ makes the statistic independent of the dependence structure. Second, the bootstrap data-generating mechanism is adaptive to unknown time-dependent variances $\{\sigma_i^2\}$. For the distribution of α_i in step (iii), we use $\mathbb{P}(\alpha_i = -1) = \mathbb{P}(\alpha_i = 1) = 1/2$, which enjoys certain nice properties as shown by Davidson and Flachaire (2008).

For the purpose of comparison, we briefly introduce the widely used block bootstrap for a stationary time series $\{X_i\}$ with mean μ. By (1.2), $\sqrt{n}(\bar{X} - \mu) \Rightarrow N(0, \tau^2)$. Suppose that we want to bootstrap the distribution of $\sqrt{n}(\bar{X} - \mu)$. Let $k_n, \ell_n, \mathcal{I}_1, \ldots, \mathcal{I}_{\ell_n}$ be defined as in Section 3.5 below. The non-overlapping block bootstrap works as follows

(i) Take a simple random sample of size ℓ_n with replacement from the blocks $\mathcal{I}_1, \ldots, \mathcal{I}_{\ell_n}$, and form the bootstrap data $X_1^b, \ldots, X_{n'}^b, n' = k_n \ell_n$, by pooling together X_i's for which the index i is within those selected blocks.

(ii) Let $\bar{X}^b$ be the sample average of $X_1^b, \ldots, X_{n'}^b$. Compute $\Xi_n = \sqrt{n'}\{\bar{X}^b - E^*(\bar{X}^b)\}$, where $E^*(\bar{X}^b) = \sum_{i=1}^{n'} X_i/n'$ is the conditional expectation of $\bar{X}^b$ given $\{X_i\}$.

(iii) Repeat (i)-(ii) a large number of times and use the empirical distribution of Ξ_n's as the distribution of $\sqrt{n}(\bar{X} - \mu)$.

In step (ii), another choice is the studentized version $\tilde{\Xi}_n = \sqrt{n'}\{\bar{X}^b - E^*(\bar{X}^b)\}/\hat{\tau}^b$, where $\hat{\tau}^b$ is a consistent estimate of τ based on bootstrap data. Assuming stationarity and $k_n \to \infty$, the blocks are asymptotically independent and share the same model dynamics as the whole data, which validates the above block bootstrap. For non-stationary process, block bootstrap is no longer valid as individual blocks are not representative of the whole data. By contrast, the proposed wild bootstrap is adaptive to unknown dependence and non-stationarity structure.

3.4 Change-point Analysis: Self-normalized CUSUM Test

To test a change-point in the mean of a process $\{X_i\}$, two popular CUSUM type tests are

$$T_n^1 = \max_{cn \le j \le (1-c)n} \frac{\hat{\tau}^{-1}|S_X(j)|}{\sqrt{j(1-j/n)}} \quad \text{and} \quad T_n^2 = \max_{cn \le j \le (1-c)n} \hat{\tau}^{-1}|S_X(j)|. \tag{3.12}$$

where $\hat{\tau}^2$ is a consistent estimate of the long-run variance τ^2 of $\{X_i\}$, and

$$S_X(j) = \Big(1 - \frac{j}{n}\Big) \sum_{i=1}^{j} X_i - \frac{j}{n} \sum_{i=j+1}^{n} X_i. \tag{3.13}$$

Here $c > 0$ ($c = 0.1$ in our simulation studies) is a small number to avoid the boundary issue. For IID data, $j(1 - j/n)$ is proportional to the variance of $S_X(j)$, thus T_n^1 is a studentized version of T_n^2. For IID Gaussian data, T_n^1 is equivalent to likelihood ratio test; see Csörgő and Horváth (1997). Assume that under null hypothesis the convergence holds

$$\Big\{n^{-1/2}\sum_{i=1}^{\lfloor nt\rfloor}[X_i - E(X_i)]\Big\}_{0\le t\le 1} \Rightarrow \tau\{B_t\}_{0\le t\le 1}, \tag{3.14}$$

for a standard Brownian motion $\{B_t\}_{t\ge 0}$. Then $T_n^1 \Rightarrow \max_{c\le t\le 1-c}|B_t - tB_1|/\sqrt{t(1-t)}$ and $T_n^2/\sqrt{n} \Rightarrow \max_{c\le t\le 1-c}|B_t - tB_1|$ via the continuity theorem.

For non-stationary case (4.2), (3.14) is no longer valid. Moreover, since T_n^1 and T_n^2 do not take into account the time-dependent variances σ_i^2, an abrupt change in variances may lead to a false rejection of H_0 when the mean remains constant. For example, our simulation study in Section 3.6.3 shows that the empirical false rejection probability for T_n^1 and T_n^2 is about 10% for nominal level 5%. To alleviate the issue of non-constant variances, we adopt the self-normalization approach as in previous sections. Recall F_j and V_j in (3.4). For each fixed $cn \le j \le (1-c)n$, by Theorem 15 and Slutsky's theorem, $F_j/\underline{V}_j \Rightarrow N(0,\tau^2)$ in distribution, assuming the negligibility of the approximation errors. Therefore, the self-normalization term $\underline{V}_j$ can remove the time-dependent variances. In light of this, we can simultaneously self-normalize the two terms $\sum_{i=1}^{j} X_i$ and $\sum_{i=j+1}^{n} X_i$ in (3.13) and propose the self-normalized test statistic

$$T_n^{\mathrm{SN}} = \max_{cn\le j\le (1-c)n} \hat{\tau}^{-1}|T_n(j)|, \quad \text{where} \quad T_n(j) = \frac{S_X(j)}{\sqrt{(1-j/n)^2\underline{V}_j^2 + (j/n)^2\overline{V}_j^2}}. \tag{3.15}$$

Here, $\underline{V}_j^2$ is defined as in (3.4), $\overline{V}_j^2 = \sum_{i=j+1}^{n}(X_i - \overline{X}_j)^2$ with $\overline{X}_j = (n-j)^{-1}\sum_{i=j+1}^{n} X_i$.

Theorem 17. *Assume (4.13) holds. Let $\delta_n \to 0$ be as in (3.9). Under H_0, we have the uniform approximation*

$$\max_{cn\le j\le (1-c)n} |T_n(j) - \tau T_n^*(j)| = O_{\mathrm{p}}(\delta_n),$$

where $$T_n^*(j) = \frac{(1-j/n)\sum_{i=1}^{j}\sigma_i(B_i - B_{i-1}) - j/n\sum_{i=j+1}^{n}\sigma_i(B_i - B_{i-1})}{\sqrt{(1-j/n)^2\sum_{i=1}^{j}\sigma_i^2 + (j/n)^2\sum_{i=j+1}^{n}\sigma_i^2}}. \tag{3.16}$$

By Theorem 17, under H_0, T_n^{SN} is asymptotically equivalent to $\max_{cn\le j\le (1-c)n}|T_n^*(j)|$.

Thanks to the self-normalization, for each j, the time-dependent variances are removed and $T_n^*(j) \sim N(0,1)$ has a standard normal distribution. However, $T_n^*(j)$ and $T_n^*(j')$ are correlated for $j \neq j'$. So, $\{T_n^*(j)\}$ is a non-stationary Gaussian process with a standard normal marginal density. Due to the large number of unknown parameters σ_i, it is infeasible to obtain the null distribution directly. On the other fact, Theorem 17 establishes the fact that, asymptotically, the distribution of T_n^{SN} in (3.15) depends only on $\sigma_1, \ldots, \sigma_n$ and is robust against the dependence structure of $\{e_i\}$, which motivates us to use the wild bootstrap method in Section 3.3 to find the critical value of T_n^{SN}.

(i) Compute $T_n(j)$ and find $J^* = \mathrm{argmax}_{cn \leq j \leq (1-c)n} |T_n(j)|$.

(ii) Divide the data into two blocks $X_1, \ldots, X_{J^*}$ and $X_{J^*+1}, \ldots, X_n$. Within each block, subtract its sample mean from the observations therein to obtain centered data. Pool all centered data together and denote them by $\epsilon_1, \ldots, \epsilon_n$.

(iii) Based on $\epsilon_1, \ldots, \epsilon_n$, obtain an estimate $\hat{\tau}$ of τ. See Section 3.5 below.

(iv) Compute the test statistic T_n^{SN} in (3.15).

(v) Based on ϵ_i in (ii), use the wild bootstrap method in Section 3.3 to generate synthetic data $\xi_1, \ldots, \xi_n$, and use (i)-(iv) to compute the bootstrap test statistic T_n^b based on the bootstrap data $\xi_1, \ldots, \xi_n$.

(vi) Repeat (v) a large number of times and find $(1-\alpha)$ quantile of those T_n^b's.

As argued in Section 3.3, the synthetic data-generating scheme (v) inherits the time-varying non-stationarity structure of the original data. Also, the statistic T_n^{SN} is robust against the dependence structure, which justifies the proposed bootstrap method. If H_0 is rejected, the change-point is then estimated by $J^* = \mathrm{argmax}_{cn \leq j \leq (1-c)n} |T_n(j)|$.

If there is no evidence to reject H_0, we briefly discuss how to apply the same methodology to test whether there is a change-point in the variances σ_i^2. By (1.1), we have $(X_i - \mu)^2 = \sigma_i^2 + \sigma_i^2 \zeta_i$, where $\zeta_i = e_i^2 - 1$ has mean zero. Therefore, testing a change-point in the variances σ_i^2 of X_i is equivalent to testing a change-point in the mean of the new data $\tilde{X}_i = (X_i - \bar{X})^2$.

3.5 Long-run Variance Estimation

To apply the results in Sections 3.2-3.4, we need a consistent estimate of the long-run variance τ^2. For long-run variance estimation, most existing works deal with stationary time series through various block bootstrap and subsampling approaches; see Lahiri (2003) and references therein. To attenuate the non-stationarity issue, we extend the idea in Section 3.2 to blockwise self-normalization. Let k_n be the block length. Denote by $\ell_n = \lfloor n/k_n \rfloor$ the largest integer not exceeding n/k_n. Ignore the boundary and divide $1, \ldots, n$ into ℓ_n blocks

$$\mathcal{I}_j = \{(j-1)k_n + 1, \ldots, jk_n\}, \quad j = 1, \ldots, \ell_n. \tag{3.17}$$

Recall the overall sample mean $\bar{X}$. For each block j, define the self-normalized statistic

$$D_j = \frac{k_n[\bar{X}(j) - \bar{X}]}{V(j)}, \quad \text{where} \quad \bar{X}(j) = \frac{1}{k_n} \sum_{i \in \mathcal{I}_j} X_i, \quad V^2(j) = \sum_{i \in \mathcal{I}_j} [X_i - \bar{X}(j)]^2. \tag{3.18}$$

By Proposition 2, the self-normalized statistics $D_1, \ldots, D_{\ell_n}$ are asymptotically independent and identically distributed as $N(0, \tau^2)$. Thus, we propose estimating τ^2 by

$$\hat{\tau}^2 = \frac{1}{\ell_n} \sum_{j=1}^{\ell_n} D_j^2. \tag{3.19}$$

As in (3.5)–(3.6), we define the quantities on block j

$$r(j) = |\sigma_{jk_n}| + \sum_{i \in \mathcal{I}_j} |\sigma_i - \sigma_{i-1}| \quad \text{and} \quad r^*(j) = |\sigma_{jk_n}^2| + \sum_{i \in \mathcal{I}_j} |\sigma_i^2 - \sigma_{i-1}^2|, \tag{3.20}$$

$$\Sigma^2(j) = \sum_{i \in \mathcal{I}_j} \sigma_i^2 \quad \text{and} \quad \Sigma^{*2}(j) = \Big(\sum_{i \in \mathcal{I}_j} \sigma_i^4 \Big)^{1/2}. \tag{3.21}$$

Theorem 18. *Let (4.13) hold with $\Delta_n = n^{1/4} \log(n)$. Recall r_n, Σ_n in (3.5)–(3.6). Define*

$$M_n = \frac{1}{k_n} + \max_{1 \le j \le \ell_n} \frac{\Sigma^{*2}(j) + r^*(j)\Delta_n}{\Sigma^2(j)} + \max_{1 \le j \le \ell_n} \frac{r(j)\Delta_n}{\Sigma(j)}. \tag{3.22}$$

Assume that $r_n \Delta_n / \Sigma_n \to 0$ *and*

$$\chi_n = \ell_n^{-1/2} + \log(n) M_n + \sqrt{\log(n)} \frac{\Sigma_n}{\ell_n^2} \sum_{j=1}^{\ell_n} \frac{1}{\Sigma(j)} + \frac{\Sigma_n^2}{\ell_n^3} \sum_{j=1}^{\ell_n} \frac{1}{\Sigma^2(j)} \to 0. \tag{3.23}$$

Then $\hat{\tau}^2 - \tau^2 = O_{\mathrm{p}}(\chi_n)$. *Consequently,* $\hat{\tau}$ *is a consistent estimate of* τ.

Consider Example 2 with $\gamma \in [0, 1)$. Then $\chi_n \asymp \sqrt{\log(n)/\ell_n} + \log^2(n)(n^{1/4}/\sqrt{k_n} + n^{5/4-\gamma}/k_n + \sqrt{k_n} n^{1/4-\gamma})$. For $\gamma \in (3/4, 1)$, it can be shown that the optimal rate is $\chi_n \asymp n^{-1/8} \log^{5/4}(n)$ when $k_n \asymp n^{3/4} \log^{3/2}(n)$. Now consider Example 3 that $\sigma_i = i^\beta$ for some $\beta \in [0, 1)$. Then elementary but tedious calculations show that, the optimal rate is

$$\chi_n \asymp \begin{cases} n^{-1/8} \log^{5/4}(n), & k_n \asymp n^{3/4} \log^{3/2}(n), \quad \beta \in [0, 3/4], \\ n^{\frac{\beta-1}{5-4\beta}} \{\log(n)\}^{\frac{8(1-\beta)}{5-4\beta}}, & k_n \asymp n^{\frac{4.5-4\beta}{5-4\beta}} \{\log(n)\}^{\frac{4}{5-4\beta}}, \quad \beta \in (3/4, 1). \end{cases}$$

3.6 Numerical Results

3.6.1 Selection of Block Length k_n for $\hat{\tau}$

In (3.19), we need to select the block length parameter k_n. Recall that $D_1, \ldots, D_{\ell_n}$ in (3.19) are asymptotically IID normal random variables. To get a sensible choice of k_n, we propose a simulation based method by minimizing the empirical mean squared error (MSE).

(i) Simulate n IID standard normal random variables $Z_1, \ldots, Z_n$.

(ii) Based on $Z_1, \ldots, Z_n$, obtain $\hat{\tau}$ with block length k.

(iii) Repeat (i)-(ii) a large number of times, compute empirical MSE(k) as the average of realizations of $(\hat{\tau} - 1)^2$, and find the optimal choice $k^* = \mathrm{argmin}_k\, \mathrm{MSE}(k)$.

We find that, the optimal block length k is about 12 for $n = 120$, about 15 for $n = 240$, about 20 for $n = 360, 600$, and about 25 for $n = 1200$.

3.6.2 Empirical Coverage Probabilities

Let sample size $n = 120$. Recall e_i and σ_i in (1.1). For σ_i, consider four choices:

$$\text{A1}: \quad \sigma_i = 0.2\mathbf{1}_{i \le n/2} + 0.6\mathbf{1}_{i > n/2},$$

$$\begin{aligned} \text{A2}: &\quad \sigma_i = 0.2\{1+\cos^2(i/n^{4/5})\}, \\ \text{A3}: &\quad \sigma_i = 0.2 + 0.1\log(1+|i-n/2|), \\ \text{A4}: &\quad \sigma_i = 0.3 + \phi(i/60), \end{aligned}$$

where ϕ is the standard normal density and $\mathbf{1}$ is the indicator function. The sequences A1–A4 exhibit different patterns, with a piecewise constancy for A1, a cosine shape for A2, a sharp change around time $n/2$ for A3, and a gradual downtrend for A4. For the stationary process e_i, let ε_i be IID N(0,1), we consider both linear and nonlinear processes:

$$\begin{aligned} \text{B1}: &\quad e_i = \{\eta_i - E(\eta_i)\}/\sqrt{\text{Var}(\eta_i)}, \quad \text{where} \quad \eta_i = \theta|\eta_{i-1}| + \sqrt{1-\theta^2}\varepsilon_i,\ |\theta|<1. \\ \text{B2}: &\quad e_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}, \quad \text{where} \quad a_j = \frac{(j+1)^{-\beta}}{\sqrt{\sum_{j=0}^{\infty}(j+1)^{-2\beta}}},\ \beta > 1/2. \end{aligned}$$

For B1, by Wu (2007), (4.13) holds with $\Delta_n = n^{1/4}\log(n)$. By Andel et al. (1984), $E(\eta_i) = \theta\sqrt{2/\pi}$ and $\text{Var}(\eta_i) = 1-2\theta^2/\pi$. To examine how the strength of dependence affects the performance, we consider $\theta = 0, 0.4, 0.8$, representing independence, intermediate, and strong dependence, respectively. For B2 with $\beta > 2$, (4.13) holds with $\Delta_n = n^{1/4}\log(n)$, and we consider three cases $\beta = 2.1, 3, 4$. To access the effect of block length k_n, three choices $k_n = 8, 10, 12$ are used. Thus, we consider all 72 combinations of $\{\text{A1}, \text{A2}, \text{A3}, \text{A4}\} \times \{\text{B1}, \theta = 0, 0.4, 0.8; \text{B2}, \beta = 2.1, 3, 4\} \times \{k_n = 8, 10, 12\}$.

Without loss of generality we examine coverage probabilities based on 10^3 realized confidence intervals for $\mu = 0$ in (1.1). We compare our self-normalization based confidence intervals to some stationarity based methods. For (1.1), if we pretend that the error process $\{\tilde{e}_i = \sigma_i e_i\}$ is stationary, then we can use (1.2) to construct asymptotic confidence interval for μ. Under stationarity, the long-run variance τ^2 of $\{\tilde{e}_i\}$ can be similarly estimated through the block method in Section 3.5 by using the non-normalized version $D_j = \sqrt{k_n}[\bar{X}(j) - \bar{X}]$ in (3.19); see Lahiri (2003). Thus, we compare two self-normalization based methods and three stationarity based alternatives: self-normalization based confidence intervals through the asymptotic theory in Proposition 2 (SN) and the wild bootstrap (WB) in Section 3.3, stationarity based confidence intervals through the asymptotic theory

(1.2) (ST), non-overlapping block bootstrap (BB), and studentized non-overlapping block bootstrap (SBB) in Section 3.3. From the results in Tables 4.1–4.2, we see that the coverage probabilities of the proposed self-normalization based methods (columns SN and WB) are close to the nominal level 95% for almost all cases considered. By contrast, the stationarity based methods (columns ST, BB and SBB) suffer from substantial undercoverage, especially when dependence is strong ($\theta = 0.8$ in Table 4.1 and $\beta = 2.1$ in Table 4.2). For the two self-normalization based methods, WB slightly outperforms SN.

Table 3.1: Coverage probabilities (in percentage) for μ in (1.1) with e_i from B1. Nominal level is 95%. SN and WB denote self-normalization based confidence intervals using asymptotic theory in Proposition 2 and the wild bootstrap procedure, respectively; ST, BB, SBB denote stationarity based confidence intervals using asymptotic theory in (1.2), non-overlapping block bootstrap, and studentized non-overlapping block bootstrap, respectively.

θ	k_n	σ_i	SN	WB	ST	BB	SBB	σ_i	SN	WB	ST	BB	SBB
0.0	8		98.0	94.7	93.1	92.2	92.8		96.6	95.2	92.3	92.5	92.5
	10	A1	98.2	95.0	92.6	92.4	92.2	A2	94.6	94.6	90.0	89.5	89.4
	12		98.1	95.6	91.7	91.4	91.1		92.1	93.7	89.7	89.5	89.6
	8		96.4	95.0	92.5	92.3	92.0		96.6	95.6	93.1	92.6	93.0
	10	A3	94.7	94.7	90.8	90.6	90.6	A4	95.1	95.1	91.4	91.3	91.3
	12		93.7	94.8	90.8	90.4	90.5		92.9	93.7	89.8	89.7	89.5
0.4	8		98.7	95.9	92.7	92.6	92.9		96.6	95.3	92.5	92.4	92.0
	10	A1	98.5	95.7	92.8	92.7	92.3	A2	95.4	95.4	91.6	91.1	91.6
	12		98.0	95.0	90.8	90.8	90.2		92.5	94.0	89.4	89.1	89.4
	8		96.6	95.2	91.7	91.7	91.6		95.4	94.1	90.8	90.9	90.6
	10	A3	95.3	95.5	91.5	91.3	91.5	A4	95.0	94.8	91.2	90.7	90.8
	12		93.1	94.6	90.2	89.9	89.9		94.1	95.1	90.3	89.8	90.1
0.8	8		97.9	94.6	87.8	86.8	87.3		96.1	94.7	87.2	87.3	87.0
	10	A1	97.6	95.5	87.3	87.0	86.7	A2	93.3	92.9	86.4	86.8	86.1
	12		97.3	94.0	85.8	85.5	85.1		92.6	93.4	86.5	86.4	86.4
	8		94.8	93.5	85.7	85.7	86.0		95.5	94.7	86.3	86.1	86.1
	10	A3	93.5	93.8	85.7	85.5	85.2	A4	95.3	95.1	88.5	88.3	88.5
	12		92.4	93.3	87.2	86.7	86.9		92.6	94.2	86.3	85.8	85.7

Table 3.2: Coverage probabilities (in percentage) for μ in (1.1) with e_i from B2. Nominal level is 95%. SN and WB denote self-normalization based confidence intervals using asymptotic theory in Proposition 2 and the wild bootstrap procedure, respectively; ST, BB, SBB denote stationarity based confidence intervals using asymptotic theory in (1.2), non-overlapping block bootstrap, and studentized non-overlapping block bootstrap, respectively.

β	k_n	σ_i	SN	WB	ST	BB	SBB	σ_i	SN	WB	ST	BB	SBB
4.0	8		97.6	94.9	91.8	91.4	91.9		95.9	94.2	91.9	92.0	91.1
	10	A1	97.7	93.2	88.9	88.1	88.3	A2	95.7	95.7	92.1	91.8	92.1
	12		97.9	95.5	90.7	90.2	90.0		93.3	94.6	90.0	89.9	89.7
	8		94.6	93.3	89.8	89.5	89.5		95.6	94.7	91.3	91.7	91.0
	10	A3	95.1	95.2	91.6	91.4	91.5	A4	95.4	95.9	92.8	92.2	93.0
	12		93.8	95.4	90.8	90.6	90.2		93.9	94.9	88.9	88.5	88.6
3.0	8		99.1	95.7	91.1	91.0	91.2		95.8	94.6	90.4	89.8	90.1
	10	A1	98.5	96.4	91.6	90.9	91.1	A2	95.6	95.2	92.1	91.9	91.5
	12		97.9	94.6	89.6	89.3	89.0		94.1	95.0	90.5	90.2	90.4
	8		95.9	94.6	92.0	91.9	91.7		96.0	94.5	90.6	90.4	90.3
	10	A3	94.3	94.4	90.0	89.9	89.8	A4	94.3	94.4	89.2	89.3	88.9
	12		93.2	94.5	88.9	88.6	88.7		93.1	94.1	89.6	88.9	88.8
2.1	8		97.1	92.5	86.2	86.2	85.5		95.7	93.8	88.9	89.0	88.7
	10	A1	97.6	94.7	89.2	88.9	88.6	A2	93.5	93.6	88.8	88.8	88.4
	12		97.2	95.1	87.9	87.5	87.7		92.6	93.9	88.0	87.6	87.7
	8		94.0	93.7	88.5	88.4	88.3		95.0	93.1	88.8	88.7	88.6
	10	A3	93.3	93.8	88.1	87.9	87.8	A4	94.1	94.2	89.1	88.8	89.1
	12		92.9	94.7	89.1	88.4	88.4		91.5	92.6	87.7	87.5	87.5

3.6.3 Size and Power Study

In (4.2), we use the same setting for σ_i and e_i as in Section 3.6.2. For mean μ_i, we consider $\mu_i = \lambda \mathbf{1}_{i>40}, \lambda \geq 0$, and compare the test statistics T_n^1, T_n^2 in (3.12) and T_n^{SN} in (3.15). First, we compare their size under the null with $\lambda = 0$. For T_n^{SN}, its critical value is obtained using the wild bootstrap in Section 3.4; for T_n^1 and T_n^2, the critical values are based on the block bootstrap in Section 3.3. In each case, we use 10^3 bootstrap samples, nominal level 5%, and block length $k_n = 10$, and summarize the empirical sizes (under the null $\lambda = 0$) in Table 3.3 based on 10^3 realizations. While T_n^{SN} has size close to 5%, T_n^1 and T_n^2 tend to over-reject the null and the false rejection probabilities can be twice more than the nominal level

5%. Next, we compare the size-adjusted power. Instead of using the bootstrap methods to obtain critical values, we use 95% quantiles of 10^4 realizations of the test statistics when data are simulated directly from the null model so that the empirical size is exactly 5%. Figure 3.1 presents the power curves for combinations {A1–A4} $\times$ {B1 with $\theta = 0.4$; B2 with $\beta = 3.0$} with 10^3 realizations each. For A1, T_n^{SN} outperforms T_n^1 and T_n^2; for A2–A4, there is a moderate loss of power for T_n^{SN}. Overall, T_n^{SN} has a comparable power to other two tests. In practice, however, the null model is unknown, and when one turns to the bootstrap method to obtain the critical values, the usual cumulative sum tests T_n^1 and T_n^2 will likely over-reject the null as shown in in Table 3.3. In summary, with such small sample size and complicated non-stationarity structure, T_n^{SN} along with the wild bootstrap method delivers reasonably good power and the size is close to nominal level.

Table 3.3: Size (in percentage) comparison of T_n^1 and T_n^2 in (3.12) and T_n^{SN} in (3.15), with sample size $n = 120$, nominal level 5%, and block length $k_n = 10$.

	Model B1				Model B2			
σ_i	θ	T_n^{SN}	T_n^1	T_n^2	β	T_n^{SN}	T_n^1	T_n^2
A1	0.0	4.9	9.1	8.4	2.1	7.3	12.2	13.4
	0.4	4.7	9.4	9.6	3.0	4.7	8.6	9.2
	0.8	6.0	15.1	14.7	4.0	5.6	9.9	7.7
A2	0.0	5.7	8.2	6.1	2.1	5.8	9.5	8.6
	0.4	6.1	8.9	6.8	3.0	5.3	9.6	6.8
	0.8	7.3	12.6	9.3	4.0	4.2	7.5	4.2
A3	0.0	5.0	5.7	4.8	2.1	5.5	7.7	6.7
	0.4	5.3	6.9	5.4	3.0	5.8	6.1	4.9
	0.8	7.0	9.8	10.0	4.0	5.0	6.5	4.2
A4	0.0	5.4	8.4	6.0	2.1	6.9	8.8	7.1
	0.4	5.7	7.9	5.2	3.0	4.8	6.6	6.3
	0.8	7.2	11.1	9.2	4.0	5.3	6.2	5.8

3.6.4 Application to Real Data Sets

Application to Annual Mean Precipitation in Seoul during 1771–2000

The data set consists of annual mean precipitation rates in Seoul during 1771–2000; see Figure 3.2 for a time series plot. The mean levels seem to be different for the two time periods

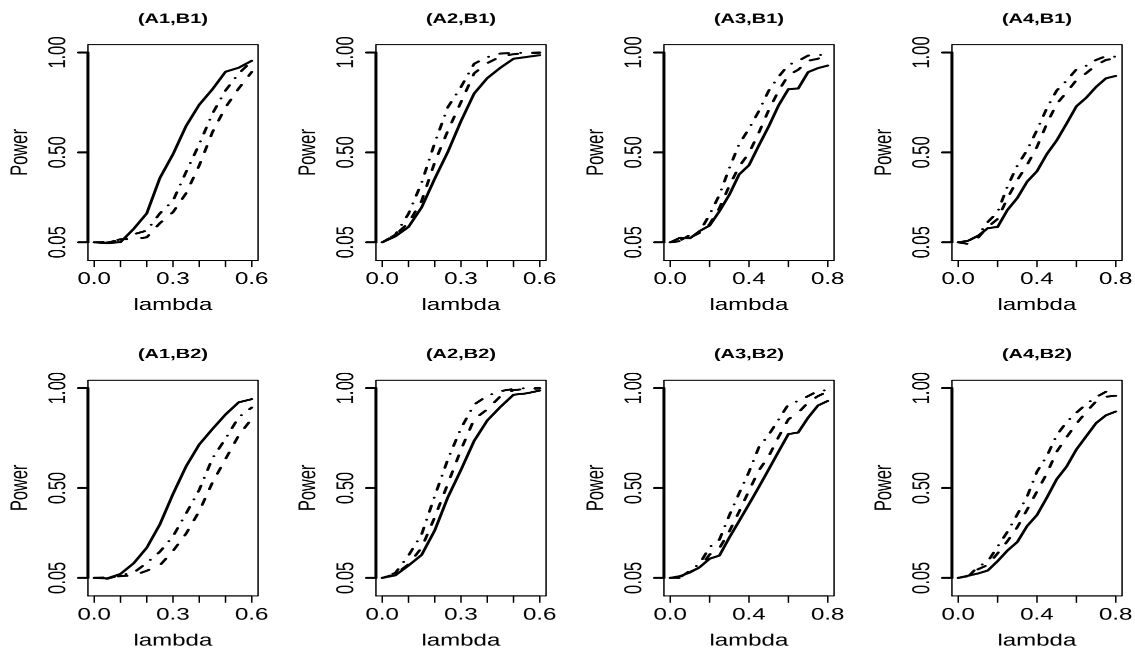

Figure 3.1: Size-adjusted power curves for T_n^1 (dashed curve) and T_n^2 (dotdash curve) in (3.12) and T_n^{SN} (solid curve) in (3.15) as functions of change size λ (horizontal axis) with sample size $n = 120$ and block length $k_n = 10$. For (A1,B1)–(A4,B1), the error process $\{e_i\}$ is from B1 with $\theta = 0.4$; for (A1,B2)–(A4,B2), the error process $\{e_i\}$ is from B2 with $\beta = 3.0$.

1771–1880 and 1881–2000. Ha and Ha (2006) assumed the observations are independent and identically distributed under the null hypothesis. From Figure 3.2, the variations are changing over time. Also, the auto-correlation function plot, not reported here, indicates strong dependence up to lag 18. Therefore, it is more reasonable to apply our self-normalization based test that is tailored to deal with non-stationary time series. With sample size $n = 230$, by the method in Section 3.6.1, the optimal block length is about 15. Based on 10^5 bootstrap samples as described in Section 3.4, we obtain the corresponding p-values 0.016, 0.005, 0.045, 0.007, with block length parameter $k_n = 12, 14, 16, 18$, respectively. For all choices of k_n, there is compelling evidence that a change-point occurred at year 1880. While our result is consistent with that of Ha and Ha (2006), our non-stationary time series framework seems to be more reasonable. Denote by μ_1 and μ_2 the mean levels over pre-change and post-change time periods 1771–1880 and 1881–2000. For the two sub-periods with sample sizes 110 and 120, the optimal block length is about 12. With $k_n = 12$, applying the wild bootstrap in Section 3.3 with 10^5 bootstrap samples, we obtain 95% confidence intervals $[121.7, 161.3]$ for μ_1, $[100.9, 114.3]$ for μ_2. For the difference $\mu_1 - \mu_2$, with optimal block length $k_n = 15$, the 95% wild bootstrap confidence interval is $[19.6, 48.2]$. Note that the latter confidence interval for $\mu_1 - \mu_2$ does not cover zero, which provides further evidence for $\mu_1 \neq \mu_2$ and the existence of the change-point at year 1880.

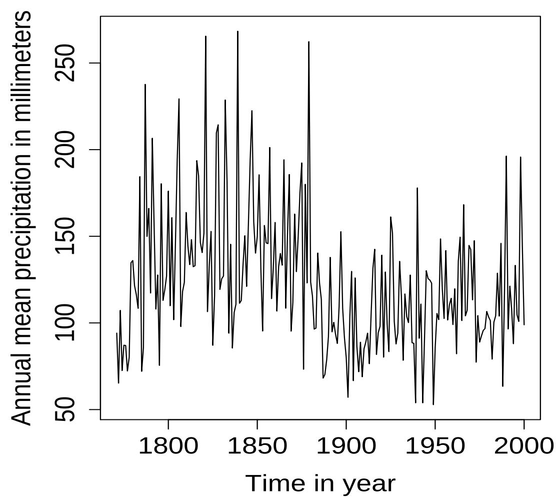

Figure 3.2: Annual mean precipitation in Seoul during 1771–2000.

Quarterly U.S. GNP Growth Rates during 1947–2002

The data set consists of quarterly U.S. Gross National Product (GNP) growth rates from the first quarter of 1947 to the third quarter of 2002; see Section 3.8 in Shumway and Stoffer (2006). Shumway and Stoffer (2006) fitted a stationary autoregressive model. However, the plot in Figure 3.3 suggests a non-stationary pattern: the variation becomes smaller after year 1985 whereas the mean level remains constant. Moreover, the stationarity test in Kwiatkowski et al. (1992) provides fairly strong evidence for non-stationarity with a p-value of 0.088. With the block length $k_n = 12, 14, 16, 18,$ we obtain the corresponding p-values $0.853, 0.922, 0.903, 0.782,$ and hence there is no evidence to reject the null hypothesis of a constant mean μ. Based on $k_n = 15$, the 95% wild bootstrap confidence interval for μ is $[0.66\%, 1.00\%]$. To test whether there is a change-point in the variance, by the discussion in the last paragraph of Section 3.4, we consider $\tilde{X}_i = (X_i - \underline{X}_n)^2$. With $k_n = 12, 14, 16, 18,$ the corresponding p-values are $0.001, 0.006, 0.001, 0.010,$ indicating strong evidence for a change-point in the variance at year 1984. In summary, we conclude there is no change-point in the mean level whereas there is a change-point in the variance at year 1984.

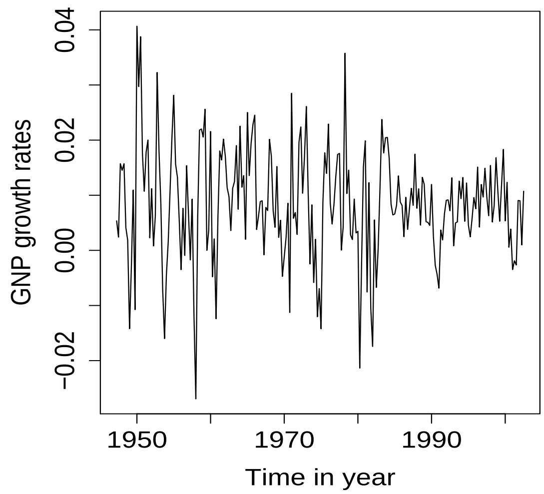

Figure 3.3: Quarterly U.S. GNP growth rates during 1947–2002.

3.7 Proofs

Proof of Theorem 15. Let $r_j = |\sigma_j| + \sum_{i=2}^{j} |\sigma_i - \sigma_{i-1}|$. By the triangle inequality, we have $r_j \le r_n$. For S_i in (4.13), $e_i = S_i - S_{i-1}$. By the summation by parts formula, (3.7) follows via

$$\begin{aligned} F_j = \sum_{i=1}^{j} \sigma_i (S_i - S_{i-1}) &= \sigma_j S_j + \sum_{i=1}^{j-1} (\sigma_i - \sigma_{i+1}) S_i \\ &= \sigma_j \tau B_j + \sum_{i=1}^{j-1} (\sigma_i - \sigma_{i+1}) \tau B_i + O_{\text{a.s.}}(r_n \Delta_n) \\ &= \tau \sum_{i=1}^{j} \sigma_i (B_i - B_{i-1}) + O_{\text{a.s.}}(r_n \Delta_n). \end{aligned} \tag{3.24}$$

By Kolmogorov's maximal inequality for independent random variables, for $\delta > 0$,

$$P\Big\{ \max_{1 \le j \le n} \Big| \sum_{i=1}^{j} \sigma_i (B_i - B_{i-1}) \Big| \ge \delta \Sigma_n \Big\} \le (\delta \Sigma_n)^{-2} E\Big[\Big\{ \sum_{i=1}^{n} \sigma_i (B_i - B_{i-1}) \Big\}^2 \Big] = \delta^{-2}. \tag{3.25}$$

Thus, by (3.24), $\max_{1\le j\le n}|F_j| = O_{\mathrm{p}}(\Sigma_n + r_n\Delta_n)$. Observe that

$$\underline{V}_j^2 - \Sigma_j^2 = W_j - F_j^2/j, \quad \text{where} \quad W_j = \sum_{i=1}^{j} \sigma_i^2(e_i^2 - 1). \tag{3.26}$$

By (4.13), the same argument in (3.24)–(3.25) shows $W_j = O_{\mathrm{p}}(\Sigma_n^{*2} + r_n^*\Delta_n)$, uniformly. The desired result then follows via (3.26). $\diamondsuit$

Proof of Theorem 16. Denote by $\Phi(x)$ the standard normal distribution function. By Proposition 2 and Slutsky's theorem, $H_n/\hat{\tau} \to N(0,1)$ and $\mathbb{P}(H_n/\hat{\tau} \le x) \to \Phi(x)$ for each fixed $x \in \mathbb{R}$. Since $\Phi(x)$ is a continuous distribution, $\sup_{x\in\mathbb{R}} |\mathbb{P}(H_n/\hat{\tau} \le x) - \Phi(x)| = 0$. It remains to prove $\sup_{x\in\mathbb{R}} |\mathbb{P}^*(H_n^* \le x) - \Phi(x)| \to 0$, in probability. Note that, conditioning on $\{e_i\}$, $\{\xi_i\}$ are independent random variables with zero mean. By the Berry-Esséen bound in Bentkus et al. (1996), there exists a finite constant c such that

$$\sup_{x\in\mathbb{R}} |\mathbb{P}^*(H_n^* \le x) - \Phi(x)| \le c\sum_{i=1}^{n} E^*(|\xi_i|^3)\left\{\sum_{i=1}^{n} E^*(|\xi_i|^2)\right\}^{-3/2}, \tag{3.27}$$

where E^* denotes conditional expectations given $\{e_i\}$. Clearly, $E^*(|\xi_i|^2) = \sigma_i^2 e_i^2 E(\alpha_1^2)$ and $E(|\xi_i|^3) = \sigma_i^3|e_i^3|E(|\alpha_1^3|)$. Thus, under the assumption $e_i \in \mathcal{L}^3$, we have $\sum_{i=1}^n E^*(|\xi_i|^3) = O_{\mathbb{P}}(\sum_{i=1}^n \sigma_i^3)$. Meanwhile, by the proof of Theorem 15, $\sum_{i=1}^n E^*(|\xi_i|^2) = \sum_{i=1}^n \sigma_i^2 e_i^2 = \{1 + o_{\mathbb{P}}(1)\}\sum_{i=1}^n \sigma_i^2$. Therefore, the desired result follows from (3.27) in view of (3.10). $\diamondsuit$

Proof of Theorem 17. For $cn \le j \le (1-c)n$, $c \le (1 - j/n), j/n \le 1 - c$. For $S_X(j)$ in (3.13), by (3.7), we have $\max_{cn\le j\le(1-c)n} |S_X(j) - \tau S_X^*(j)| = O_{\mathrm{a.s.}}(r_n\Delta_n)$, where

$$S_X^*(j) = \left(1 - \frac{j}{n}\right)\sum_{i=1}^{j} \sigma_i(B_i - B_{i-1}) - \frac{j}{n}\sum_{i=j+1}^{n} \sigma_i(B_i - B_{i-1}).$$

By (3.8), $\max_{cn\le j\le(1-c)n} |(1-j/n)^2\underline{V}_j^2 + (j/n)^2\overline{V}_j^2 - V_j^2| = O_{\mathrm{p}}(\varpi_n)$, where

$$V_j^2 = (1-j/n)^2\sum_{i=1}^{j}\sigma_i^2 + (j/n)^2\sum_{i=j+1}^{n}\sigma_i^2 \quad \text{and} \quad \varpi_n = (r_n^2\Delta_n^2 + \Sigma_n^2)/n + \Sigma_n^{*2} + r_n^*\Delta_n.$$

For $cn \le j \le (1-c)n$, $V_j^2 \ge c^2\Sigma_n^2$. Thus, condition (3.9) implies $\varpi_n = o(V_j^2)$ and $\{V_j^2 + O_{\mathrm{p}}(\varpi_n)\}^{1/2} = V_j + O_{\mathrm{p}}(\varpi_n/V_j)$. Therefore, uniformly over $cn \le j \le (1-c)n$,

$$T_n(j) - \tau T_n^*(j) = \frac{\tau S_X^*(j) + O_{\mathrm{a.s.}}(r_n\Delta_n)}{V_j + O_{\mathrm{p}}(\varpi_n/V_j)} - \frac{\tau S_X^*(j)}{V_j} = O_{\mathrm{p}}\Big\{\frac{r_n\Delta_n}{V_j} + \frac{\varpi_n S_X^*(j)}{V_j^3}\Big\}.$$

By (3.25), $\max_j |S_X^*(j)| = O_{\mathrm{p}}(\Sigma_n)$. So the desired result follows in view of $V_j \ge c\Sigma_n$. ♢

Proof of Theorem 18. Condition $M_n \to 0$ implies $\max_{1\le j\le \ell_n} r(j)\Delta_n/\Sigma(j) \to 0$. By (3.8),

$$\omega_j = \frac{V^2(j)}{\Sigma^2(j)} - 1 = O_{\mathrm{p}}\Big\{\frac{\Sigma^{*2}(j) + r^*(j)\Delta_n}{\Sigma^2(j)} + \frac{1}{k_n}\Big\} = O_{\mathrm{p}}(M_n) \to 0. \qquad (3.28)$$

Define $U_j = \Sigma^{-1}(j)\sum_{i\in\mathcal{I}_j}\sigma_i(B_i - B_{i-1})$. Clearly, $U_1,\ldots,U_{\ell_n}$ are independent standard normal random variables. Thus, $\max_{1\le j\le\ell_n}|U_j| = O_{\mathrm{p}}\{\sqrt{\log(\ell_n)}\} = O_{\mathrm{p}}\{\sqrt{\log(n)}\}$. By (3.7), $\underline{X}_n - \mu = O_{\mathrm{p}}\{(\Sigma_n + r_n\Delta_n)/n\} = O_{\mathrm{p}}(\Sigma_n/n)$. Recall the definition of D_j in (3.18). By the same argument in (3.7), using $\sqrt{1+x} = 1 + O(x)$ as $x \to 0$, we have

$$\begin{aligned} D_j &= \frac{k_n\{\bar{X}(j) - \mu\}}{\Sigma(j)}\frac{1}{\sqrt{1+\omega_j}} + \frac{k_n(\mu - \underline{X}_n)}{\Sigma(j)}\frac{1}{\sqrt{1+\omega_j}} \\ &= \Big[\tau U_j + O_{\mathrm{a.s.}}\Big\{\frac{r(j)\Delta_n}{\Sigma(j)}\Big\}\Big]\{1 + O(\omega_j)\} + O_{\mathrm{p}}\Big\{\frac{k_n\Sigma_n}{n\Sigma(j)}\Big\} \\ &= \tau U_j + O_{\mathrm{p}}\Big\{\sqrt{\log(n)}M_n + \frac{\Sigma_n}{\ell_n\Sigma(j)}\Big\}. \end{aligned}$$

By the latter expression and $\log(n)M_n \to 0$, we can easily verify $\hat{\tau}^2 - \tau^2 = O_{\mathbb{P}}(\chi_n)$. ♢

Chapter 4

Testing for Changes in Autocovariances of Nonparametric Time Series Models

[1] In this chapter, we are interested in testing the dependency structure of a process in the form of

$$X_i = f(i/n) + e_i, \quad i = 1, \dots, n, \tag{4.1}$$

where $\{e_i\}$ is a process with $\mathbb{E}(e_i) = 0$, and $f(\cdot)$ is an unknown time trend.

We consider testing the change-point of qutocovariances of the error term $cov(e_i, e_{i+k})$ for lag $k \geq 0$. The null hypothesis is

$$H_0^{(k)} : \operatorname{cov}(e_1, e_{1+k}) = \cdots = \operatorname{cov}(e_{n-k}, e_n), \tag{4.2}$$

against the alternative that there exists one J, called the change-point, such that

$$H_A^{(k)} : \operatorname{cov}(e_1, e_{1+k}) = \cdots = \operatorname{cov}(e_J, e_{J+1}) \neq \operatorname{cov}(e_{J+1}, e_{J+1+k}) = \cdots = \operatorname{cov}(e_{n-k}, e_n).$$

[1] This chapter is based on the article of Li and Zhao (2013).

4.1 Dependence Structure

We adopt Wu (2007)'s framework for the dependence structure on $\{e_i\}$. Under the null hypothesis of no change-points, we assume that e_i has the causal representation

$$e_i = G(\ldots, \varepsilon_{i-1}, \varepsilon_i), \tag{4.3}$$

for a measurable function G and independent and identically distributed (IID) innovations ε_i. Let $\{\varepsilon'_i\}_{i\in\mathbb{Z}}$ be an IID copy of $\{\varepsilon_i\}_{i\in\mathbb{Z}}$. Define the coupling process of $\{e_i\}$:

$$e'_i = G(\ldots, \varepsilon'_{-1}, \varepsilon'_0, \varepsilon_1, \ldots, \varepsilon_i). \tag{4.4}$$

For a random variable Z, we write $Z \in \mathcal{L}^q, q > 0$, if $\|Z\|_q = [\mathbb{E}(|Z|^q)]^{1/q} < \infty$. Following Shao and Wu (2007), we say that $\{e_i\}$ satisfies the *geometric moment contraction* condition GMC(q, ρ) for some $q > 0$ and $\rho \in (0, 1)$ if $e_i \in \mathcal{L}^q$ and

$$\mathrm{GMC}(q, \rho): \qquad \|e_i - e'_i\|_q \le C\rho^i, \quad i \in \mathbb{N}, \tag{4.5}$$

for some constant $C < \infty$. If we view (4.3) as an input-output system with $(\ldots, \varepsilon_{i-1}, \varepsilon_i), G$, and e_i being, respectively, the input, filter, and output, then (4.5) asserts that the impact of replacing the distant input $\{\varepsilon_j\}_{j\le 0}$ by an IID copy $\{\varepsilon'_j\}_{j\le 0}$ decays exponentially.

Example 5. Consider the linear process $e_i = \sum_{j=1}^{\infty} a_j \varepsilon_{i-j}$. If $\varepsilon_i \in \mathcal{L}^q$ and $a_i = O(\rho^i)$ for some $\rho \in (0, 1)$, then (4.5) holds. In particular, it is well-known that causal ARMA models (all roots of the autoregressive polynomial lie outside of the unit disc) satisfy the latter condition; see Brockwell and Davis (1991).

Example 6. Consider the nonlinear autoregressive model of order $p \in \mathbb{N}$:

$$e_i = \mu(e_{i-1}, \ldots, e_{i-p}) + s(e_{i-1}, \ldots, e_{i-p})\varepsilon_i, \tag{4.6}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are two measurable functions and $\{\varepsilon_i\}$ are IID innovations. Write

$x = (x_1, \ldots, x_p), x' = (x'_1, \ldots, x'_p)$. If $\varepsilon_0 \in \mathcal{L}^q$ and there exists $a_1, \ldots, a_p$ such that

$$|\mu(x) - \mu(x')| + |s(x) - s(x')| \|\varepsilon_0\|_q \leq \sum_{j=1}^{p} a_j |x_j - x'_j| \quad \text{and} \quad \sum_{j=1}^{p} a_j < 1, \tag{4.7}$$

then the causal representation (4.5) holds for some $\rho \in (0, 1)$; see Theorem 5.1 in Shao and Wu (2007). Many widely used parametric models are special cases of (4.6), including linear autoregressive model, threshold autoregressive model, exponential autoregressive model, random coefficient model, and autoregressive conditional heteroscedastic model.

Proposition 4 below is adopted from Zhao, Wei and Lin (2012). It shows that, if $\{e_i\}$ satisfies (4.5), then its properly transformed process also satisfies (4.5).

Proposition 4. *For $0 < \varsigma \leq 1$ and $\upsilon \geq 0$, define the collection of functions h*

$$\mathcal{H}(\varsigma, \upsilon) = \{h : |h(x) - h(x')| \leq c|x - x'|^{\varsigma}(1 + |x| + |x'|)^{\upsilon}, \quad x, x' \in \mathbb{R}\}, \tag{4.8}$$

where c is a constant. Suppose $\{e_i\}$ satisfies the condition GMC(q, ρ) in (4.5). Then the transformed process $e_i^* = h(e_i)$ satisfies GMC$(q/(\varsigma + \upsilon), \rho^{\varsigma})$.

In (4.8), $\mathcal{H}(\varsigma, 0)$ is the class of uniformly Hölder-continuous functions with index ς. If $h(x) = |x|^b, b > 1$, then $h \in \mathcal{H}(1, b - 1)$. Proposition 4 provides a way to generate many nonlinear processes satisfying (4.5). In particular, by Proposition 4 and Examples 5–6, our subsequent results have much wider applications compared to the linear process framework in Lee, Ha and Na (2003) and Berkes, Gombay and Horváth (2009).

Proposition 5 below gives some useful moment inequalities.

Proposition 5. *Suppose $\{e_i\}$ satisfies (4.5) for some $q \geq 2$.*

(i) For any $i \leq j$, $\mathbb{E}(e_i e_j) = O(\rho^{j-i})$ holds uniformly.

(ii) If $q \geq 4$, then $\mathbb{E}(e_i e_j e_r e_s) = O[\sqrt{\rho}^{(j-i)+(s-r)}]$ for all integers $i \leq j \leq r \leq s$.

4.2 Autocovariance Process and Cumulative Sum Test

For $\{e_i\}$ in (4.1), denote by $\gamma_k = \mathrm{cov}(e_i, e_{i+k})$ the autocovariance function under the null. Since $e_1, \ldots, e_n$ are unknown, we need to estimate them first in order to study their autocovariances. For each $t \in (0, 1)$, we can estimate $f(t)$ by the Priestley-Chao estimator (Priestley and Chao, 1972):

$$\hat{f}(t) = \frac{1}{nb}\sum_{i=1}^{n} X_i K\Big(\frac{i/n - t}{b}\Big), \tag{4.9}$$

where $b > 0$ is a bandwidth, and $K(\cdot) \geq 0$ is a kernel function. Based on the residuals $\hat{e}_i = X_i - \hat{f}(i/n)$, we define the sample autocovariance process:

$$\hat{\gamma}_k(t) = \frac{\sum_{i=\lfloor n\delta \rfloor}^{\lfloor nt \rfloor} \hat{e}_i \hat{e}_{i+k}}{\lfloor nt \rfloor - \lfloor n\delta \rfloor + 1}, \quad t \in (\delta, 1-\delta), \quad \text{where} \quad \hat{e}_i = X_i - \hat{f}(i/n). \tag{4.10}$$

Here $\delta > 0$ is a small constant to avoid the boundary issue since the estimate $\hat{f}(t)$ is usually not consistent at the boundaries.

Assumption 2. *(i) $K(\cdot)$ is bounded, symmetric, integrates to one, and has bounded support and bounded derivative; (ii) $\sup_{t\in[0,1]} |f''(t)| < \infty$; (iii) $nb^2/\log^2 n \to \infty$ and $nb^8 \to 0$.*

Theorem 19 below establishes the convergence of the normalized autocovariance process to a time-shifted Brownian motion, where the shift δ is due to the boundary truncation.

Theorem 19. *Suppose (4.5) holds for some $q > 4$ and Assumption 2 holds. Under (4.3), we have the functional central limit theorem in the Skorokhod space (Billingsley, 1968):*

$$\Big\{\sqrt{n}(t-\delta)[\hat{\gamma}_k(t) - \gamma_k]\Big\}_{t\in(\delta,1-\delta)} \Rightarrow \tau(k)\Big\{\mathbb{B}_{t-\delta}\Big\}_{t\in(\delta,1-\delta)},$$

where $\{\mathbb{B}_t\}$ is a standard Brownian Motion, and

$$\tau^2(k) = \mathbb{E}(\eta_0^2) + 2\sum_{i=1}^{\infty} \mathbb{E}(\eta_0 \eta_i), \quad \textit{with} \quad \eta_i = e_i e_{i+k} - \gamma_k. \tag{4.11}$$

To address the change-point problem (4.2), consider the cumulative sum test:

$$T_n^{(k)} = \sup_{t\in(\delta,1-\delta)} |T_n^{(k)}(t)|, \quad \text{where} \quad T_n^{(k)}(t) = n^{-1/2} \sum_{i=\lfloor n\delta \rfloor}^{\lfloor nt \rfloor} [\hat{e}_i \hat{e}_{i+k} - \hat{\gamma}_k(1-\delta)]. \tag{4.12}$$

By writing $T_n^{(k)}(t) = n^{-1/2}(\lfloor nt \rfloor - \lfloor n\delta \rfloor + 1)[\hat{\gamma}_k(t) - \hat{\gamma}_k(1-\delta)]$, we see that $T_n^{(k)}(t)$ measures the deviation between the two estimates of γ_k: $\hat{\gamma}_k(t)$ based on the residuals up to time t, and $\hat{\gamma}_k(1-\delta)$ based on all residuals (excluding the boundaries). Under $H_0^{(k)}$, we expect that such deviation should be small. Thus, $T_n^{(k)}$ measures the overall maximal deviation over $t \in (\delta, 1-\delta)$, with a large value indicating the rejection of $H_0^{(k)}$.

Theorem 20. *Assume the same conditions in Theorem 19. Under (4.3), we have*

$$T_n^{(k)} \Rightarrow \tau(k) \sup_{t\in(\delta,1-\delta)} \left| \mathbb{B}_{t-\delta} - \frac{t-\delta}{1-2\delta} \mathbb{B}_{1-\delta} \right|,$$

where $\tau(k)$ and $\{\mathbb{B}_t\}$ are defined as in Theorem 19.

For $\delta = 0.05$, by simulations, the 90%, 95%, and 99% quantiles of $\sup_{t\in(\delta,1-\delta)} |\mathbb{B}_{t-\delta} - (t-\delta)/(1-2\delta)\mathbb{B}_{1-\delta}|$ are 1.14, 1.26, and 1.52, respectively. Therefore, at significance level 0.05, we reject $H_0^{(k)}$ if $T_n^{(k)} \geq 1.26\hat{\tau}(k)$, where $\hat{\tau}^2(k)$ is a consistent estimator of $\tau^2(k)$.

Remark 1. The causal representation (4.3) implies the strict stationarity of $\{e_i\}$. Thus, the asymptotic null distribution in Theorem 20 is based on the strict stationarity of $\{e_i\}$, which implies but is stronger than the null hypothesis $H_0^{(k)}$ in (4.2). Unfortunately, it is unclear how to derive the asymptotic null distribution under $H_0^{(k)}$ instead of the stronger assumption (4.3). In the absence of strict stationarity, different segments of the data may have completely different behavior and it is technically challenging to study them. For example, consider the following model: the first half of the data are IID random variables with zero mean and unit variance whereas the second half of the data follow the autoregressive conditional heteroscedastic model (Engle, 1982) $e_i = \sqrt{0.5 + 0.5e_{i-1}^2}\varepsilon_i$, where ε_i are IID random variables with $\mathbb{E}(\varepsilon_i) = 0$ and $\mathbb{E}(\varepsilon_i^2) = 1$. Then both segments have the same autocovariances $\text{cov}(e_i, e_j) = 1$ for $i = j$ and $\text{cov}(e_i, e_j) = 0$ for $i \neq j$.

Using the well-known Bartlett kernel estimator, we can estimate $\tau^2(k)$ by

$$\hat{\tau}^2(k) = \widehat{\text{cov}(\eta_0, \eta_0)} + 2\sum_{j=1}^{I}\Big(1 - \frac{j}{I+1}\Big)\widehat{\text{cov}(\eta_0, \eta_j)}, \tag{4.13}$$

where I is a bandwidth parameter, $\widehat{\text{cov}(\eta_0, \eta_j)}$ is the sample version of $\mathbb{E}(\eta_0\eta_j)$ given by

$$\widehat{\text{cov}(\eta_0, \eta_j)} = \frac{1}{\ell_j}\sum_{i=\lfloor n\delta\rfloor}^{\lfloor n(1-\delta)\rfloor - j}[\hat{e}_i\hat{e}_{i+k} - \hat{\gamma}_k(1-\delta)][\hat{e}_{i+j}\hat{e}_{i+j+k} - \hat{\gamma}_k(1-\delta)],$$

and $\ell_j = \lfloor n(1-\delta)\rfloor - \lfloor n\delta\rfloor + 1 - j$.

Theorem 21. *Suppose (4.5) holds for some $q \ge 8$ and Assumption 2 holds. Further assume*

$$I \to \infty \qquad \textit{and} \qquad I\Big[b^2 + \Big(\frac{\log n}{nb}\Big)^{1/2}\Big] \to 0. \tag{4.14}$$

Under (4.3), for each given $k \ge 0$, we have $\hat{\tau}^2(k) = \tau^2(k) + o_p(1)$.

The proof of Theorem 21, given in Section 4.5, relies on the decomposition $\hat{\tau}^2(k) = \tau_I^2 + R_n + O_p(1)\Omega_n$ in (4.45), where τ_I^2 is a truncated version of $\tau^2(k)$ in (4.11) using finite terms, and R_n and Ω_n quantifies the influence from estimating the unknown function $f(\cdot)$. In (4.14), the condition $I \to \infty$ is needed to ensure $\tau_I^2 \to \tau^2(k)$, and the condition $I\{b^2 + [(nb)^{-1/2}\log n]^{1/2}\} \to 0$ guarantees the negligibility of R_n and Ω_n (the effect of estimating f).

We now address the selection of the bandwidth I. It is well-known that the optimal choice of b in (4.9) is $b \asymp n^{-1/5}$. If $I \asymp n^{\beta}$, then (4.14) holds for $\beta \in (0, 2/5)$. For long-run variance estimation with Bartlett kernel, the optimal bandwidth is $I = cn^{1/3}$ for some constant c (Andrews, 1991), which satisfies (4.14) when using $b \asymp n^{-1/5}$. By the proof of Theorem 21, under (4.5), $\mathbb{E}(\eta_0\eta_j) = O(\rho^j)$ in (4.11) decays exponentially fast. Thus, even in the presence of strong dependence, a relatively small truncation lag I can result in a reasonably accurate convergence of τ_I^2 to τ^2. For example, if $\rho = 0.7$, the tail $\sum_{j=11}^{\infty}\rho^j \approx 0.066$ is small. Furthermore, a small I is preferred so that $I\{b^2+[(nb)^{-1/2}\log n]^{1/2}\} \to 0$, and thus the effect of estimating f is negligible. In summary, we recommend using a relatively small I; based on our simulations, taking $c = 1$ in the optimal choice $I = cn^{1/3}$ works quite well.

Remark 2. In (4.13), we can also use other kernels; see, e.g. Andrews (1991) for several other choices. For the class of flat-top kernels developed by Politis and Romano in a series of papers, Politis (2003) proposed some data-driven bandwidth choice.

Now we study the power of the proposed test statistic. Consider the fixed alternative

$$H_a : e_i = \begin{cases} G_1(\ldots, \varepsilon_{i-1}, \varepsilon_i), & i = 1, \ldots, \lfloor nt_0 \rfloor, \\ G_2(\ldots, \varepsilon_{i-1}, \varepsilon_i), & i = \lfloor nt_0 \rfloor + 1, \ldots, n, \end{cases} \tag{4.15}$$

for two measurable functions G_1 and G_2 and a change-point $t_0 \in (0, 1)$. Under H_a, $\{e_i\}$ consists of two stationary processes of the form (4.3).

Theorem 22. *Suppose both G_1 and G_2 satisfy (4.5) for some $q > 4$ and Assumption 2 holds. For the two stationary processes in (4.15), denote their autocovariances by $\gamma_{k,1}$ and $\gamma_{k,2}$, respectively. Further assume $t_0 \in (\delta, 1 - \delta)$, $\gamma_{k,1} \neq \gamma_{k,2}$, and that the bandwidth I in (4.13) satisfies $I/n \to 0$. Then $\mathbb{P}\{T_n^{(k)} > \hat{\tau}(k)\} \to 1$ as $n \to \infty$.*

By Theorem 22, the proposed test statistic has asymptotic power one under the fixed alternative (4.15). Our simulation studies in Section 4.4.1 further demonstrate that the power increases very fast as the before change-point model moves further away from the after change-point model [i.e., as λ in (4.23) increases]. We point out that, it is quite challenging to study the asymptotic power under local alternatives and it deserves future research.

The null hypothesis (4.2) tests the constancy of γ_k for each given $k \geq 0$. Alternatively, we can also test the joint constancy of $\gamma_0, \gamma_1, \ldots, \gamma_T$ for a lag T. Such joint constancy test can address whether there is a change-point in the overall autocovariance structure but does not tell which specific autocovariance has a change-point. By contrast, testing (4.2) is more informative and it tells us whether there is a change-point in the autocovariance at a particular lag k. For example, it is possible that we have a change-point in the variance (γ_0) but not in the autocovariance at lag one (γ_1). Furthermore, it is technically more challenging to test the joint constancy due to the complicated correlations among the sample autocovariance functions. For example, Berkes, Gombay and Horváth (2009) also considered only the test for the constancy of individual γ_k.

4.3 Weighted Partial Sum and Quadratic Forms

The presence of the unknown function $f(\cdot)$ makes it substantially more challenging to prove Theorem 19 and Theorem 21. Intuitively, the residual $\hat{e}_i$ in (4.10) has two components: one is the underlying noise e_i and the other is the estimation error from estimating $f(\cdot)$ by $\hat{f}(\cdot)$. To control the second component, three technical issues arise. First, we need to study certain weighted partial sum of $\{e_i\}$. Second, since the estimation error also contains $\{e_i\}$, we must also deal with some weighted quadratic form in order to study $\hat{\gamma}_k(t)$ in (4.10). Third, to study the process $\{\hat{\gamma}_k(t)\}_t$, it is necessary to obtain uniform convergence of the latter weighted partial sum process and quadratic form. In this section we present some useful results in these directions, which may be of independent interest.

Proposition 6. *Suppose (4.5) holds for some $q \geq 2$. Then, for any sequence $\{a_i\}_{i=1}^n$,*

$$\max_{1\leq k\leq n}\left|\sum_{i=1}^{k} a_i e_i\right| = O_p\left\{\log n\left[\max_{1\leq i\leq n}|a_i|\sum_{i=1}^{n}|a_i|\right]^{1/2}\right\}. \tag{4.16}$$

The bound in (4.16) is optimal up to a logarithm factor. For example, if $a_i \equiv 1$ and $\{e_i\}$ are IID, then $\max_{1\leq k\leq n}|\sum_{i=1}^{k} e_i| = O_p(\sqrt{n})$ via Kolmogorov's maximal inequality. Similarly, the bounds in Proposition 7 below are also optimal up to a logarithm factor.

Proposition 7. *Suppose (4.5) holds for some $q \geq 2$. Then, for any double array $\{\theta_{i,j}\}_{i,j=1}^n$,*

$$(i) \quad \max_{1\leq k\leq n}\left|\sum_{i=1}^{k}\sum_{j=1}^{n}\theta_{i,j}e_j\right| = O_p(\sqrt{\chi_n}n\log n), \tag{4.17}$$

$$\textit{where} \quad \chi_n = \max_{1\leq i,j\leq n}|\theta_{i,j}|\max_{1\leq i\leq n}\sum_{j=1}^{n}|\theta_{i,j}|; \tag{4.18}$$

$$(ii) \quad \max_{1\leq i\leq n}\left|\sum_{j=1}^{n}\theta_{i,j}e_j\right| = O_p\left\{\sqrt{\Theta_n\log n} + L_n n^{1/\min(q,4)}\log n\right\}, \tag{4.19}$$

$$\text{where} \quad \Theta_n = \max_{1\le i\le n} \sum_{j=1}^{n} \theta_{i,j}^2, \quad L_n = \max_{1\le i\le n} \left(|\theta_{i,n}| + \sum_{j=1}^{n-1} |\theta_{i,j} - \theta_{i,j+1}| \right). \tag{4.20}$$

Proposition 8 below presents a uniform bound for a weighted quadratic form of $\{e_i\}$.

Proposition 8. *Suppose (4.5) holds for some $q \ge 4$. Let $\{\theta_{i,j}\}_{i,j=1}^n$ be any double array. Assume that there exists N_n such that*

$$\max_{1\le i\le n} |\{1 \le j \le n : \theta_{i,j} \ne 0\}| \le N_n, \quad \max_{1\le j\le n} |\{1 \le i \le n : \theta_{i,j} \ne 0\}| \le N_n. \tag{4.21}$$

Here $|\mathcal{S}|$ denotes the cardinality of a finite set $\mathcal{S}$. Then

$$\max_{1\le k\le n} \left| \sum_{i=1}^{k} \sum_{j=1}^{n} \theta_{i,j} e_i e_j \right| = O_p\Big(\sqrt{nN_n} \log n \max_{1\le i,j\le n} |\theta_{i,j}|\Big). \tag{4.22}$$

4.4 Numerical Results

Throughout our subsequent analysis we use significance level 0.05 and $\delta = 0.05$ in (4.10).

4.4.1 Simulation Studies

In this section we examine the empirical size and power of the test based on 1000 realizations. To reduce the boundary effect, in (4.9) we use the local linear smoothing method implemented by the command `locpoly` in `R` package `KernSmooth`, with the optimal plug-in bandwidth selection using the command `dpill`. In (4.1), we let $f(t) = \cos(2\pi t)$ and sample size $n = 600$. For e_i, we consider two models below.

First, we consider the linear autoregressive model:

$$\text{Model 1}: \quad e_i = \begin{cases} 0.2e_{i-1} + 0.2\varepsilon_i, & i \le 250; \\ (0.2+\lambda)e_{i-1} + 0.2\varepsilon_i, & i \ge 251, \end{cases} \tag{4.23}$$

where ε_i are IID N(0,1) random variables. For $i \le 250$, the autocovariance function of $\{e_i\}$ is $\gamma_k = 0.2^{2+k}/(1-0.2^2), k \ge 0$; for $i \ge 251$, the autocovariance function is $\gamma_k =$

$0.2^2(0.2+\lambda)^k/[1-(0.2+\lambda)^2], k \geq 0$. Thus, $\lambda = 0$ corresponds to the null hypothesis of no change-point, whereas $\lambda \neq 0$ indicates a change-point at time 251. By the discussion following Theorem 21, the optimal I is of order $n^{1/3} \approx 8.4$; three choices $I = 8, 10, 12$ are examined to assess the effect of I. We consider testing change-points in $\gamma_0, \gamma_1, \gamma_2$, and report the empirical size ($\lambda = 0.0$), empirical power and size-corrected power ($\lambda = 0.2, 0.4, 0.6$) in Table 4.1. We see that the empirical size is close to the nominal level 0.05, and the empirical power increases with λ. Also, the three choices of I give similar performance, with $I = 8$ slightly outperforming $I = 10, 12$. Interestingly, it is the easiest to detect change-points in γ_1 whereas it is the most difficult to detect change-points in the variance γ_0. Note that, for $\lambda = 0.2$, the differences between before and after change-points autocovariances are 0.006, 0.011 and 0.006 for γ_0, γ_1 and γ_2, respectively. The largest power of γ_1 is due to its most visible difference. One possible explanation for the relatively better power of γ_2 over γ_0 is: although the differences for γ_2 and γ_0 are the same, the variances of the estimates $\hat{\gamma}_2$ and $\hat{\gamma}_0$ could be very different, especially under the alternative hypothesis. For different models, it may be easier to detect change-points in autocovariances at different lags.

Next, we consider the nonlinear model:

$$\text{Model 2}: \quad e_i = \begin{cases} \lambda e_{i-1} + 0.2\varepsilon_i, & i \leq 250; \\ 0.2\sqrt{1+e_{i-1}^2}\varepsilon_i, & i \geq 251, \end{cases} \tag{4.24}$$

where ε_i are IID N(0,1) random variables. For $i \leq 250$, we have a linear autoregressive model with the autocovariance function $\gamma_k = 0.2^2\lambda^k/(1-\lambda^2), k \geq 0$; for $i \geq 251$, we have the autoregressive conditional heteroscedastic model (Engle, 1982), which is widely used in financial econometrics. For $i \geq 251$, we can show that $\gamma_0 = 0.2^2/(1-0.2^2)$, and by the martingale difference property $\gamma_k = 0, k \geq 1$. Therefore, for $\lambda = 0$, although the model dynamics are quite different before and after the change-point, the autocovariance structures are almost the same (except that the variance changes from $0.2^2 = 0.04$ to $0.2^2/(1-0.2^2) \approx 0.042$). As λ increases, the discrepancy between the autocovariances increases. The empirical power in Table 4.2 shows the same pattern as in Table 4.1.

Table 4.1: Empirical power and size-corrected power (in parenthesis) for Model 1 in (4.23) ($\lambda = 0.0$ corresponds to the size)

		λ			
	k	0.0	0.2	0.4	0.6
	0	0.053	0.113(0.102)	0.347(0.327)	0.697(0.676)
$I = 8$	1	0.048	0.437(0.449)	0.946(0.949)	1.000(1.000)
	2	0.055	0.170(0.161)	0.681(0.665)	0.988(0.986)
	0	0.045	0.095(0.113)	0.355(0.395)	0.702(0.736)
$I = 10$	1	0.050	0.430(0.431)	0.955(0.955)	0.998(0.998)
	2	0.048	0.151(0.164)	0.695(0.715)	0.977(0.979)
	0	0.045	0.078(0.079)	0.300(0.317)	0.694(0.714)
$I = 12$	1	0.046	0.409(0.417)	0.931(0.935)	0.999(1.000)
	2	0.045	0.137(0.159)	0.648(0.685)	0.963(0.973)

Table 4.2: Empirical power and size-corrected power (in parenthesis) for Model 2 in (4.24) ($\lambda = 0.0$ corresponds to the size)

		λ				
	k	0.0	0.2	0.4	0.6	0.8
	0	0.055	0.044(0.041)	0.079(0.069)	0.408(0.393)	0.813(0.798)
$I = 8$	1	0.040	0.423(0.461)	0.967(0.970)	0.999(1.000)	1.000(1.000)
	2	0.044	0.053(0.059)	0.230(0.247)	0.810(0.823)	0.993(0.993)
	0	0.056	0.048(0.041)	0.088(0.074)	0.359(0.322)	0.762(0.744)
$I = 10$	1	0.039	0.437(0.469)	0.947(0.955)	1.000(1.000)	1.000(1.000)
	2	0.038	0.053(0.067)	0.189(0.234)	0.791(0.825)	0.992(0.996)
	0	0.056	0.039(0.038)	0.086(0.081)	0.341(0.334)	0.749(0.743)
$I = 12$	1	0.042	0.406(0.436)	0.945(0.960)	1.000(1.000)	1.000(1.000)
	2	0.048	0.055(0.055)	0.193(0.198)	0.783(0.785)	0.991(0.991)

4.4.2 An Application to Interest Rates

Interest rates play an important role in virtually every aspect of financial markets, including consumer spending, housing markets, banks' borrowing costs, inflation, and stock market performance. Here we consider market yields on U.S. treasury securities at one-year constant maturity. The data, available at *http://www.federalreserve.gov/releases/h15/data.htm*, consist of weekly (Friday) rates during January 5, 1962–June 10, 2011, with a total of $n = 2579$ observations. From the plot in the left panel of Figure 4.1, we see dramatic interest rates hike in the early 1980s with a peak as high as 17% in 1981. Lai (2008) argued that the observed unit-root behavior of interest rates may be due to structural breaks. Here we shall examine whether there are change-points in autocovariances.

The left plot of Figure 4.1 shows a complicated time-varying trend that can hardly be described using any parametric form. Thus, we use the nonparametric model (4.1). The right panel of Figure 4.1 shows the residuals $\hat{e}_i$ in (4.10) after removing the nonparametric time trend f. We consider testing the null hypothesis $H_0^{(k)}$ in (4.2) for $k = 0, 1, 2$. Since $n^{1/3} \approx 13.7$, we use four choices $I = 12, 15, 18, 20$ in (4.13), and all lead to the same conclusion that we can reject $H_0^{(0)}, H_0^{(1)}, H_0^{(2)}$ at the significance level 0.05. For example, with $I = 15$, the corresponding p-values for $H_0^{(0)}, H_0^{(1)}, H_0^{(2)}$ are 0.013, 0.017, 0.023, respectively, indicating strong evidence to reject the null hypothesis.

With $I = 15$, Figure 4.2 plots the cumulative sum process $|T_n^{(k)}(t)|/\hat{\tau}(k)$ for testing the constancy of γ_0 ($k = 0$, solid curve), γ_1 ($k = 1$, dotted curve), and γ_2 ($k = 2$, dashed curve), where $T_n^{(k)}(t)$ is defined in (4.12). It is in December 1981 when the cumulative sum processes exceed the critical level 1.26 (corresponding to the significance level 0.05), and we conclude that the autocovariances $\gamma_0, \gamma_1, \gamma_2$ have changed since January 1982. Also, the sharp increase in $|T_n^{(k)}(t)|/\hat{\tau}(k)$ in the early 1980s roughly corresponds to the "interest rates hike" period (also see the unusual activities of the residuals plot in Figure 4.1).

During the 1970s, U.S. economy experienced rising unemployment and accelerating inflation. According to U.S. Bureau of Labor Statistics (*http://www.bls.gov/data*), the Consumer Price Index doubled from 37.9 in January 1970 to 76.9 in December 1979, compared to a mere 25% increase during the 1960s. At a meeting in October 1979, federal policymakers approved new policy procedures aiming to bring down the inflation by increasing the interest rates. As a result, during the next three years (1980–1982), the interest rates increased sharply and averaged 13%. Such abrupt changes likely resulted in the change-point of autocovariances in December 1981 as detected by our analysis above.

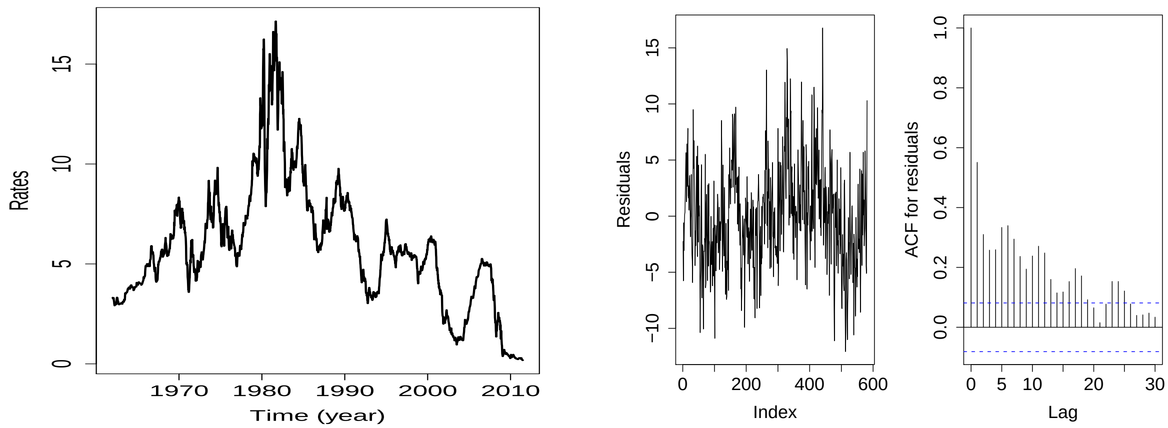

Figure 4.1: Left: Plot of the weekly rates during January 5, 1962–June 10, 2011. Right: Plot of the residuals $\hat{e}_i$ in (4.10) after removing the nonparametric time trend f.

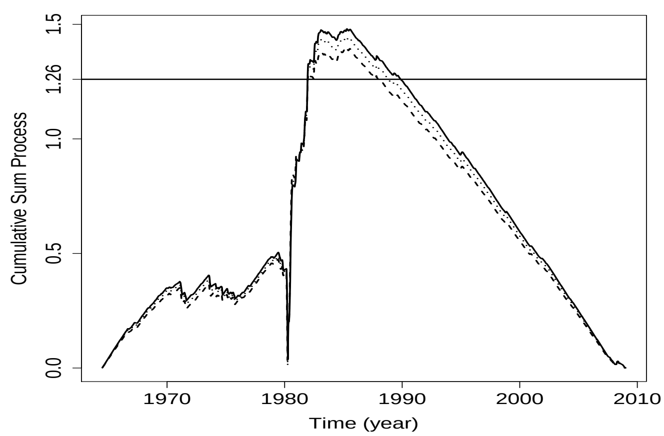

Figure 4.2: Plot of the cumulative sum process $|T_n^{(k)}(t)|/\hat{\tau}(k)$ ($\delta = 0.05$) in (4.12) for $k = 0$ (solid curve), $k = 1$ (dotted curve), and $k = 2$ (dashed curve). The horizontal line is the critical value 1.26 at the significance level 0.05. Here $\hat{\tau}^2(k)$ is the estimator in (4.13) with $I = 15$.

4.5 Proofs

4.5.1 Proof of Propositions 4–5

Proof of Proposition 4. For completeness, we include the proof from Zhao, Wei and Lin (2012). Let $q^* = q/(\varsigma + \upsilon), p_1 = \upsilon/\varsigma + 1$, and $p_2 = \varsigma/\upsilon + 1$ so that $\varsigma q^* p_1 = q, \upsilon q^* p_2 = q$, and

$1/p_1 + 1/p_2 = 1$. By (4.8) and Hölder's inequality [p. 67, Kufner, John and Fučik (1977)] $\mathbb{E}|Z_1 Z_2| \le \|Z_1\|_{p_1} \|Z_2\|_{p_2}$, we have

$$\begin{aligned}\|h(e_i') - h(e_i)\|_{q^*}^{q^*} &\le O(1)\mathbb{E}[|e_i' - e_i|^{\varsigma q^*}(1 + |e_i| + |e_i'|)^{\upsilon q^*}] \\ &\le O(1)\{\mathbb{E}[|e_i - e_i'|^{\varsigma q^* \cdot p_1}]\}^{1/p_1}\{\mathbb{E}[(1 + |e_i| + |e_i'|)^{\upsilon q^* \cdot p_2}]\}^{1/p_2} \\ &= O(1)\|e_i - e_i'\|_q^{q/p_1}\|e_0\|_q^{q/p_2} = O(\rho^{iq/p_1}).\end{aligned}$$

The above expression gives $\|h(e_i') - h(e_i)\|_{q^*} \le O(1)[\rho^{q/(p_1 q^*)}]^i = O(1)(\rho^{\varsigma})^i$. ♢

Proof of Proposition 5. Let $\{\varepsilon_i'\}_{i\in\mathbb{Z}}$ be an IID copy of $\{\varepsilon_i\}_{i\in\mathbb{Z}}$. Define the coupling process

$$e_{\ell,k} = G(\ldots, \varepsilon_{k-1}', \varepsilon_k', \varepsilon_{k+1}, \ldots, \varepsilon_\ell), \quad \ell > k.$$

By condition (4.5), we have $\|e_\ell - e_{\ell,k}\|_q \le C\rho^{\ell-k}$. Recall that $q \ge 2$. Thus, by Hölder's inequality, $\|e_\ell - e_{\ell,k}\|_2 \le \|e_\ell - e_{\ell,k}\|_q = O(\rho^{\ell-k})$.

(i) For $i \le j$, e_i is independent of $e_{j,i}$, which implies $\mathbb{E}(e_i e_{j,i}) = \mathbb{E}(e_i)\mathbb{E}(e_{j,i}) = 0$. Thus, by the Cauchy-Schwarz inequality, $|\mathbb{E}(e_i e_j)| = \mathbb{E}[e_i(e_j - e_{j,i})]| \le \|e_i\|_2\|e_j - e_{j,i}\|_2 = O(\rho^{j-i})$.

(ii) First, we prove that, for any $i < j < r < s$, the following assertions hold:

(iia) $\mathbb{E}(e_i e_j e_r e_s) = O[\sqrt{\rho}^{(j-i)+(s-r)}]$;
(iib) $\mathbb{E}(e_i^2 e_j e_r) = O(\rho^{r-j})$, $\mathbb{E}(e_i e_j e_r^2) = O(\rho^{j-i})$, $\mathbb{E}(e_i e_j^2 e_r) = O(\sqrt{\rho}^{r-i})$;
(iic) $\mathbb{E}(e_i e_j^3) = O(\rho^{j-i})$, $\mathbb{E}(e_i^3 e_j) = O(\rho^{j-i})$.

By assumption, for $q \ge 4$, $\|e_\ell - e_{\ell,k}\|_4 \le \|e_\ell - e_{\ell,k}\|_q \le C\rho^{\ell-k}$.

(iia) By Hölder's inequality,

$$|\mathbb{E}[e_i(e_j - e_{j,i})e_r e_s]| \le \|e_i\|_4 \times \|e_j - e_{j,i}\|_4 \times \|e_r\|_4 \times \|e_s\|_4 = O(\rho^{j-i}).$$

Similarly, $|\mathbb{E}[e_i e_{j,i}(e_r - e_{r,i})e_s]| = O(\rho^{r-i})$ and $|\mathbb{E}[e_i e_{j,i} e_{r,i}(e_s - e_{s,i})]| = O(\rho^{s-i})$. The independence between e_i and $e_{j,i}e_{r,i}e_{s,i}$ implies $\mathbb{E}(e_i e_{j,i} e_{r,i} e_{s,i}) = \mathbb{E}(e_i)\mathbb{E}(e_{j,i}e_{r,i}e_{s,i}) = 0$. Thus,

$$|\mathbb{E}(e_i e_j e_r e_s)| = |\mathbb{E}\{[e_i(e_j - e_{j,i})e_r e_s] + [e_i e_{j,i}(e_r - e_{r,i})e_s] + [e_i e_{j,i} e_{r,i}(e_s - e_{s,i})]\}| = O(\rho^{j-i}).$$

The same argument also shows $|\mathbb{E}(e_i e_j e_r e_s)| = |\mathbb{E}[e_i e_j e_r (e_s - e_{s,r})]| = O(\rho^{s-r})$. Combining the two bounds together and using the inequality $\min(a, b) \le \sqrt{ab}$ for all $a, b \ge 0$, we get

$$|\mathbb{E}(e_i e_j e_r e_s)| = O(1) \min(\rho^{j-i}, \rho^{s-r}) \le O(1)\sqrt{\rho}^{(j-i)+(s-r)}. \tag{4.25}$$

(iib) As in (iia), by Hölder's inequality, we have

$$|\mathbb{E}(e_i^2 e_j e_r)| = |\mathbb{E}[e_i^2 e_j (e_r - e_{r,j})]| \le \|e_i^2\|_2 \times \|e_j\|_4 \times \|e_r - e_{r,j}\|_4 = O(\rho^{r-j}).$$

For $\mathbb{E}(e_i e_j e_r^2)$, using $|\mathbb{E}[e_i e_{j,i}(e_r^2 - e_{r,i}^2)]| = |\mathbb{E}[e_i e_{j,i}(e_r - e_{r,i})(e_r + e_{r,i})]| = O(\rho^{r-i})$, we have

$$|\mathbb{E}(e_i e_j e_r^2)| \le |\mathbb{E}[e_i(e_j - e_{j,i})e_r^2]| + |\mathbb{E}[e_i e_{j,i}(e_r^2 - e_{r,i}^2)]| = O(\rho^{j-i} + \rho^{r-i}) = O(\rho^{j-i}).$$

Similarly, $\mathbb{E}(e_i e_j^2 e_r) = \mathbb{E}[e_i(e_j^2 e_r - e_{j,i}^2 e_{r,i})] = O(\rho^{j-i})$ and $\mathbb{E}(e_i e_j^2 e_r) = \mathbb{E}[e_i e_j^2 (e_r - e_{r,j})] = O(\rho^{r-j})$. Thus, as in (4.25), $|\mathbb{E}(e_i e_j^2 e_r)| = O(1)\min(\rho^{j-i}, \rho^{r-j}) = O(\sqrt{\rho}^{r-i})$.

(iic) We skip the proof since the argument above also applies here.

Notice that $\rho \le \sqrt{\rho}$. We conclude the proof by noting that (iia)–(iic) imply $\mathbb{E}(e_i e_j e_r e_s) = O[\sqrt{\rho}^{(j-i)+(s-r)}]$ for all $i \le j \le r \le s$. For example, if $i < j < r < s$, (iia) gives the latter bound; if $i = j < r < s$, then (iib) gives $\mathbb{E}(e_i e_i e_r e_s) = \mathbb{E}(e_i^2 e_r e_s) = O(\rho^{s-r}) = O[\sqrt{\rho}^{(i-i)+(s-r)}]$; if $i = j < r = s$, then the Cauchy-Schwarz inequality gives $\mathbb{E}(e_i e_i e_r e_r) \le \|e_i^2\|_2 \|e_r^2\|_2 = O(1) = O[\sqrt{\rho}^{(i-i)+(r-r)}]$; if $i = j = r < s$, then (iic) gives $\mathbb{E}(e_i e_i e_i e_s) = \mathbb{E}(e_i^3 e_s) = O(\rho^{s-i}) = O[\sqrt{\rho}^{(i-i)+(s-i)}]$. ♢

4.5.2 Proof of Propositions 6–8

First, we state some elementary inequalities. Let $a_i, i \in \mathbb{N}$, be any sequence of real numbers. Then for any $\rho \in (0, 1)$ and integers $1 \le \ell_1 < \ell_2$,

$$\sum_{i=\ell_1}^{\ell_2} |a_i| \rho^{|j-i|} \le \max_{\ell_1 \le i \le \ell_2} |a_i| \sum_{i=\ell_1}^{\ell_2} \rho^{|j-i|} \le \frac{1+\rho}{1-\rho} \max_{\ell_1 \le i \le \ell_2} |a_i|, \quad \text{for all } j. \tag{4.26}$$

Here the last inequality follows from $\sum_{i=\ell_1}^{\ell_2} \rho^{|j-i|} \le \sum_{k=-\infty}^{\infty} \rho^{|k|} = (1+\rho)/(1-\rho)$. By (4.26),

$$\sum_{i=\ell_1}^{\ell_2}\sum_{j=\ell_1}^{\ell_2} |a_i a_j|\rho^{|j-i|} = \sum_{i=\ell_1}^{\ell_2}\left(|a_i|\sum_{j=\ell_1}^{\ell_2}|a_j|\rho^{|j-i|}\right) \le \frac{1+\rho}{1-\rho}\max_{\ell_1\le i\le\ell_2}|a_i|\sum_{i=\ell_1}^{\ell_2}|a_i|. \tag{4.27}$$

Lemma 1. *Let $N_1,\ldots,N_n$ be any normal random variables satisfying $\mathbb{E}(N_i)=0$ and $\mathbb{E}(N_i^2)\le s_n^2$ for some $s_n>0$. Then $\max_{1\le i\le n}|N_i| = O_p(s_n\sqrt{\log n})$.*

Proof. Let N be a standard normal random variable. The proof is completed via

$$\mathbb{P}\Big(\max_{1\le i\le n}|N_i|\ge 2s_n\sqrt{\log n}\Big) \le \sum_{i=1}^{n}\mathbb{P}(|N_i|\ge 2s_n\sqrt{\log n}) \le n\mathbb{P}(|N|\ge 2\sqrt{\log n})\to 0.$$

Here the last convergence follows from $\mathbb{P}(N>z)=[1+o(1)]\phi(z)/z$ as $z\to\infty$ [see formula 7.1.13 in Abramowitz and Stegun (1964)] and $\phi(2\sqrt{\log n})=O(n^{-2})$, with $\phi(z)$ being the standard normal density. ♢

Lemma 2 below presents a useful maximal inequality from Wu (2007).

Lemma 2 (Proposition 1 in Wu (2007))**.** *Let d be a positive integer, and $Z_i\in\mathcal{L}^2, 1\le i\le 2^d$, be any random variables. Write $S_k=\sum_{i=1}^k Z_i$. Then*

$$\Big\|\max_{1\le k\le 2^d}|S_k|\Big\|_2 \le \sum_{r=0}^{d}\left[\sum_{m=1}^{2^{d-r}}\Big\|\sum_{i=(m-1)2^r+1}^{m2^r} Z_i\Big\|_2^2\right]^{1/2}.$$

Proof of Proposition 6. Let $H_k=\sum_{i=1}^k a_i e_i$. Write $d=\lceil\log_2 n\rceil$, where $\lceil x\rceil$ is the smallest integer that is not less than x. Then $n\le 2^d\le 2n$. For convenience, write $a_i=0$ for $n+1\le i\le 2^d$ so that $H_i=H_n$ for $n+1\le i\le 2^d$. By Proposition 5 (i), $|\mathbb{E}(e_ie_j)|=O(\rho^{|j-i|})$ for all i,j. Thus, by (4.27), for any $1\le\ell_1<\ell_2\le 2^d$,

$$\begin{aligned}\Big\|\sum_{i=\ell_1}^{\ell_2}a_ie_i\Big\|_2^2 &\le \sum_{i=\ell_1}^{\ell_2}\sum_{j=\ell_1}^{\ell_2}|a_ia_j||\mathbb{E}(e_ie_j)| \\ &= O(1)\sum_{i=\ell_1}^{\ell_2}\sum_{j=\ell_1}^{\ell_2}|a_ia_j|\rho^{|j-i|} = O(1)\max_{1\le i\le n}|a_i|\sum_{i=\ell_1}^{\ell_2}|a_i|. \end{aligned}\tag{4.28}$$

Therefore, by (4.28) and Lemma 2,

$$\begin{aligned}\Big\|\max_{1\le k\le n}|H_k|\Big\|_2=\Big\|\max_{1\le k\le 2^d}|H_k|\Big\|_2 &\le \sum_{r=0}^{d}\left[\sum_{m=1}^{2^{d-r}}\Big\|\sum_{i=(m-1)2^r+1}^{m2^r}a_ie_i\Big\|_2^2\right]^{1/2}\\ &= O(1)\sum_{r=0}^{d}\left[\max_{1\le i\le n}|a_i|\sum_{m=1}^{2^{d-r}}\sum_{i=(m-1)2^r+1}^{m2^r}|a_i|\right]^{1/2}\\ &= O\left\{\log n\left[\max_{1\le i\le n}|a_i|\sum_{i=1}^{n}|a_i|\right]^{1/2}\right\},\end{aligned}$$

completing the proof. ♢

Proof of Proposition 7. First, we prove (4.17). Define $W_i=\sum_{j=1}^{n}\theta_{i,j}e_j$. As in (4.28),

$$\|W_i\|_2^2=O(1)\max_{1\le j\le n}|\theta_{i,j}|\sum_{j=1}^{n}|\theta_{i,j}|\le O(1)\max_{1\le i,j\le n}|\theta_{i,j}|\max_{1\le i\le n}\sum_{j=1}^{n}|\theta_{i,j}|=O(\chi_n). \qquad (4.29)$$

Applying the triangle inequality, for any $\ell_1<\ell_2$, we can obtain

$$\Big\|\sum_{i=\ell_1}^{\ell_2}W_i\Big\|_2\le\sum_{i=\ell_1}^{\ell_2}\|W_i\|_2=O[(\ell_2-\ell_1)\sqrt{\chi_n}]. \qquad (4.30)$$

Then (4.17) follows by applying (4.30) and Lemma 2 to $\|\max_{1\le k\le n}|\sum_{i=1}^{k}W_i|\|_2$.

Next, we prove (4.19). By Corollary 4 in Wu (2007), there exists a standard Brownian Motion $\{\mathbb{B}_t\}$ such that

$$\max_{1\le j\le n}\left|\sum_{i=1}^{j}e_i-s\mathbb{B}_j\right|=O_{\text{a.s.}}\{n^{1/\min(q,4)}\log(n)\}. \qquad (4.31)$$

Here $s^2=\mathbb{E}(e_0^2)+2\sum_{i=1}^{\infty}\mathbb{E}(e_0e_i)$ is the long-run variance of $\{e_i\}$. Write $S_j=\sum_{i=1}^{j}e_i$ so that $e_j=S_j-S_{j-1}$. Then by (4.31) and the summation-by-part formula,

$$\sum_{j=1}^{n}\theta_{i,j}e_j = \sum_{j=1}^{n}\theta_{i,j}(S_j-S_{j-1})$$

$$
\begin{aligned}
&= \sum_{j=1}^{n-1}(\theta_{i,j}-\theta_{i,j+1})S_j + \theta_{i,n}S_n \\
&= \sum_{j=1}^{n-1}(\theta_{i,j}-\theta_{i,j+1})s\mathbb{B}_j + \theta_{i,n}s\mathbb{B}_n + O_{\text{a.s.}}[L_n n^{1/\min(q,4)}\log(n)] \\
&= s\sum_{j=1}^{n}\theta_{i,j}(\mathbb{B}_j - \mathbb{B}_{j-1}) + O_{\text{a.s.}}[L_n n^{1/\min(q,4)}\log(n)]. \qquad (4.32)
\end{aligned}
$$

Since $\{\mathbb{B}_j - \mathbb{B}_{j-1}\}_{j=1}^n$ are IID standard normals, $\sum_{j=1}^n \theta_{i,j}(\mathbb{B}_j - \mathbb{B}_{j-1}), i = 1,\ldots,n$, are normally distributed, and the variances $\sum_{j=1}^n \theta_{i,j}^2 \le \Theta_n$. Therefore, by Lemma 1, $\max_{1\le i\le n}|\sum_{j=1}^n \theta_{i,j}(\mathbb{B}_j - \mathbb{B}_{j-1})| = O_p[(\Theta_n \log n)^{1/2}]$. Finally, (4.20) follows from (4.32). $\diamondsuit$

Proof of Proposition 8. Recall N_n in (4.21). Suppose we can prove that, uniformly for all $\ell_1 < \ell_2$,

$$
\left\| \sum_{i=\ell_1}^{\ell_2} U_i \right\|_2^2 = O(1)N_n(\ell_2-\ell_1)\max_{1\le i,j\le n}|\theta_{i,j}|^2, \quad \text{where} \quad U_i = e_i\sum_{j=1}^n \theta_{i,j}e_j. \qquad (4.33)
$$

Then the result easily follows by applying Lemma 2 to $\|\max_{1\le k\le n}|\sum_{i=1}^k U_i|\|_2$.

In order to prove (4.33), we observe that

$$
\left\| \sum_{i=\ell_1}^{\ell_2} U_i \right\|_2^2 = \left\| \sum_{i=\ell_1}^{\ell_2}\sum_{j=1}^{n} \theta_{i,j}e_ie_j \right\|_2^2 \le \sum_{i_1=\ell_1}^{\ell_2}\sum_{i_2=\ell_1}^{\ell_2}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} |\theta_{i_1,j_1}\theta_{i_2,j_2}| \times |\mathbb{E}(e_{i_1}e_{i_2}e_{j_1}e_{j_2})|. \quad (4.34)
$$

By considering the 24 different orderings along with Proposition 5 (ii), we can get

$$
|\mathbb{E}(e_{i_1}e_{i_2}e_{j_1}e_{j_2})| = O\Big(\sqrt{\rho}^{|i_1-i_2|+|j_1-j_2|} + \sqrt{\rho}^{|i_1-j_1|+|i_2-j_2|} + \sqrt{\rho}^{|i_1-j_2|+|i_2-j_1|}\Big), \qquad (4.35)
$$

for all i_1, j_1, i_2, j_2. We consider the three terms separately below.

For convenience, write $\theta_* = \max_{1\le i,j\le n}|\theta_{i,j}|$. Applying (4.26) and (4.27), we have

$$
\sum_{j_1,j_2=1}^{n} |\theta_{i_1,j_1}\theta_{i_2,j_2}|\sqrt{\rho}^{|j_1-j_2|} \le \theta_*\frac{1+\sqrt{\rho}}{1-\sqrt{\rho}}\sum_{j_1=1}^{n}|\theta_{i_1,j_1}| \le \frac{1+\sqrt{\rho}}{1-\sqrt{\rho}}N_n\theta_*^2. \qquad (4.36)
$$

Thus, for the first term in (4.35), by (4.36) and (4.27), we have

$$\sum_{i_1,i_2=\ell_1}^{\ell_2} \sum_{j_1,j_2=1}^{n} |\theta_{i_1,j_1}\theta_{i_2,j_2}| \sqrt{\rho}^{|i_1-i_2|+|j_1-j_2|} = O(N_n\theta_*^2) \sum_{i_1,i_2=\ell_1}^{\ell_2} \sqrt{\rho}^{|i_1-i_2|} = O[N_n(\ell_2-\ell_1)\theta_*^2].$$

Now we consider the second term in (4.35). Observe that, by (4.26),

$$\sum_{i_1=\ell_1}^{\ell_2} \sum_{j_1=1}^{n} |\theta_{i_1,j_1}| \sqrt{\rho}^{|i_1-j_1|} \le \theta_* \sum_{i_1=\ell_1}^{\ell_2} \sum_{j_1=1}^{n} \sqrt{\rho}^{|i_1-j_1|} = O[(\ell_2-\ell_1)\theta_*]. \tag{4.37}$$

On the other hand, using the transformation $k = i_1 - j_1$, we can easily see

$$\sum_{i_1=\ell_1}^{\ell_2} \sum_{j_1=1}^{n} |\theta_{i_1,j_1}| \sqrt{\rho}^{|i_1-j_1|} = \sum_{k} \left(\sqrt{\rho}^{|k|} \sum_{i_1} |\theta_{i_1,i_1+k}| \right) \le N_n\theta_* \sum_{k} \sqrt{\rho}^{|k|} = O(N_n\theta_*). \tag{4.38}$$

Combining the two bounds in (4.37) and (4.38), we get

$$\sum_{i_1=\ell_1}^{\ell_2} \sum_{j_1=1}^{n} |\theta_{i_1,j_1}| \sqrt{\rho}^{|i_1-j_1|} = O(1)\min\{(\ell_2-\ell_1)\theta_*, N_n\theta_*\} = O(1)\theta_*\sqrt{N_n(\ell_2-\ell_1)}.$$

Therefore, we have

$$\sum_{i_1,i_2=\ell_1}^{\ell_2} \sum_{j_1,j_2=1}^{n} |\theta_{i_1,j_1}\theta_{i_2,j_2}| \sqrt{\rho}^{|i_1-j_1|+|i_2-j_2|} = \left(\sum_{i_1,j_1} |\theta_{i_1,j_1}| \sqrt{\rho}^{|i_1-j_1|} \right)^2 = O[N_n(\ell_2-\ell_1)\theta_*^2].$$

Similarly, the bound $O[N_n(\ell_2-\ell_1)\theta_*^2]$ holds for the third term in (4.35). So (4.33) holds. ♢

4.5.3 Proof of Theorems 19–22

Lemma 3. *Suppose the assumptions in Theorem 19 hold. Then*

$$\sup_{\lfloor n\delta\rfloor \le m \le \lfloor n(1-\delta)\rfloor} |D_m| = o_p(\sqrt{n}), \quad \textit{where} \quad D_m = \sum_{i=\lfloor n\delta\rfloor}^{m} \hat{e}_i\hat{e}_{i+k} - \sum_{i=\lfloor n\delta\rfloor}^{m} e_i e_{i+k}.$$

Proof. It is easy to see the decomposition $\hat{e}_i = e_i + f(i/n) - \hat{f}(i/n) = e_i + \Delta_i - \zeta_i$, where

$$\Delta_i = f(i/n) - \frac{1}{nb}\sum_{j=1}^{n} f(j/n)K\Big(\frac{i-j}{nb}\Big) \quad \text{and} \quad \zeta_i = \frac{1}{nb}\sum_{j=1}^{n} e_j K\Big(\frac{i-j}{nb}\Big). \tag{4.39}$$

Thus, we can write

$$D_m = \sum_{i=\lfloor n\delta\rfloor}^{m} (e_i\Delta_{i+k} + \Delta_i e_{i+k} + \Delta_i\Delta_{i+k} + \zeta_i\zeta_{i+k} - e_i\zeta_{i+k} - \zeta_i e_{i+k} - \zeta_i\Delta_{i+k} - \Delta_i\zeta_{i+k}). \tag{4.40}$$

It suffices to show that each of the eight terms above is uniformly bounded by $o_p(\sqrt{n})$.

By Lemma 4 below, $\Delta_i = O[b^2 + (nb)^{-1}]$ uniformly over $\lfloor n\delta\rfloor \le i \le \lfloor n(1-\delta)\rfloor$. Thus, $\sum_{i=\lfloor n\delta\rfloor}^{m} \Delta_i\Delta_{i+k} = O\{n[b^2 + (nb)^{-1}]^2\} = o(\sqrt{n})$ under Assumption 2 (iii).

For $\sum_{i=\lfloor n\delta\rfloor}^{m} e_i\Delta_{i+k}$, by Proposition 6, we have under Assumption 2 (iii),

$$\sum_{i=\lfloor n\delta\rfloor}^{m} e_i\Delta_{i+k} = O_p\{[b^2 + (nb)^{-1}]\sqrt{n}\log n\} = o_p(\sqrt{n}),$$

uniformly over m. The same bound also holds for $\sum_{i=\lfloor n\delta\rfloor}^{m} \Delta_i e_{i+k}$.

For $\sum_{i=\lfloor n\delta\rfloor}^{m} \zeta_i\Delta_{i+k}$, we shall apply (4.17) in Proposition 7 by writing

$$\sum_{i=\lfloor n\delta\rfloor}^{m} \zeta_i\Delta_{i+k} = \frac{1}{nb}\sum_{i=\lfloor n\delta\rfloor}^{m}\sum_{j=1}^{n} \theta_{i,j}e_j, \quad \text{where} \quad \theta_{i,j} = \Delta_{i+k}K\Big(\frac{i-j}{nb}\Big).$$

For χ_n in (4.18), using $\theta_{i,j} = O[b^2 + (nb)^{-1}]K\{(i-j)/(nb)\}$ and $K\{(i-j)/(nb)\} \ne 0$ only when $|i-j| = O(nb)$, it is easy to see $\chi_n = O\{[b^2 + (nb)^{-1}]^2 nb\}$. Thus, by (4.17) in Proposition 7, $\sum_{i=\lfloor n\delta\rfloor}^{m} \zeta_i\Delta_{i+k} = O_p\{[b^2 + (nb)^{-1}]\sqrt{nb}n\log n/(nb)\} = o_p(\sqrt{n})$, uniformly over m. The same bound also holds for $\sum_{i=\lfloor n\delta\rfloor}^{m} \Delta_i\zeta_{i+k}$.

Similarly, by Proposition 8, we can show that both $\sum_{i=\lfloor n\delta\rfloor}^{m} e_i\zeta_{i+k}$ and $\sum_{i=\lfloor n\delta\rfloor}^{m} \zeta_i e_{i+k}$ have the uniform bound $O_p[\sqrt{n^2 b}\log n/(nb)] = o_p(\sqrt{n})$. For $\sum_{i=\lfloor n\delta\rfloor}^{m} \zeta_i\zeta_{i+k}$, by (4.28), $\|\zeta_i\|_2^2 = O[(nb)^{-1}]$, uniformly over i. Thus, by the triangle inequality,

$$\mathbb{E}\left[\max_{\lfloor n\delta\rfloor \le m \le \lfloor n(1-\delta)\rfloor}\left|\sum_{i=\lfloor n\delta\rfloor}^{m} \zeta_i\zeta_{i+k}\right|\right] \le \sum_{i=\lfloor n\delta\rfloor}^{\lfloor n(1-\delta)\rfloor} \mathbb{E}(|\zeta_i\zeta_{i+k}|) = O(1/b) = o(\sqrt{n}).$$

Here we have used the Cauchy-Schwarz inequality $\mathbb{E}(|\zeta_i\zeta_{i+k}|) \le \|\zeta_i\|_2\|\zeta_{i+k}\|_2 = O[(nb)^{-1}]$. This completes the proof. $\diamondsuit$

Lemma 4. *Let Δ_i be defined in (4.39). Assume $b \to 0, nb \to \infty$, and that Assumption 2 (i)–(ii) hold. Then $\Delta_i = O[b^2 + (nb)^{-1}]$, uniformly over $\lfloor n\delta \rfloor \le i \le \lfloor n(1-\delta) \rfloor$.*

Proof. Without loss of generality, we assume that $K(\cdot)$ has support $[-1, 1]$ so that we only need to consider $|(j-i)/n| \le b$ in Δ_i. By Taylor's expansion $f(j/n) = f(i/n) + f'(i/n)(j-i)/n + O(b^2)$, we write $\Delta_i = f(i/n)(1 - I_0) - f'(i/n)I_1 + O(b^2)I_0$, where

$$I_0 = \frac{1}{nb}\sum_{j=1}^{n} K\Big(\frac{i-j}{nb}\Big), \qquad I_1 = \frac{1}{nb}\sum_{j=1}^{n} \frac{j-i}{n} K\Big(\frac{i-j}{nb}\Big).$$

The result then follows if we can prove $I_0 = 1 + O[(nb)^{-1}]$ and $I_1 = O(n^{-1})$. We only prove the first assertion since the second one can be similarly treated. Observe that

$$\begin{aligned} \left| I_0 - \frac{1}{nb}\int_0^n K\Big(\frac{x-i}{nb}\Big)dx \right| &\le \frac{1}{nb}\sum_{j=1}^{n}\int_{j-1}^{j}\Big|K\Big(\frac{j-i}{nb}\Big) - K\Big(\frac{x-i}{nb}\Big)\Big|dx \\ &\le \frac{\sup_u |K'(u)|}{nb}\sum_{|j-i|\le nb+1}\int_{j-1}^{j}\frac{1}{nb}dx = \frac{O(1)}{nb}. \end{aligned} \tag{4.41}$$

Here, the second inequality in (4.41) follows from two facts: (1°) when $|j-i| > nb+1$, we have $|j-i|/(nb) > 1$ and $|(x-i)/(nb)| > 1$ for all $x \in [j-1, j]$, and consequently $K\{(j-i)/(nb)\} = 0$ and $K\{(x-i)/(nb)\} = 0$; (2°) $|K(u) - K(u')| \le |u-u'|\sup_u |K'(u)|$. By the change-of-variable $(x-i)/(nb) = u$, it is easy to see that

$$\frac{1}{nb}\int_0^n K\Big(\frac{x-i}{nb}\Big)dx = \int_{-i/(nb)}^{(n-i)/(nb)} K(u)du = \int_{-1}^{1} K(u)du = 1, \tag{4.42}$$

where the second equality in (4.42) follows since, for $\lfloor n\delta \rfloor \le i \le \lfloor n(1-\delta) \rfloor$ with a fixed constant $\delta \in (0,1)$, we have $-i/(nb) < -1$ and $(n-i)/(nb) > 1$ for sufficiently small $b \to 0$. By (4.41) and (4.42), we obtain $I_0 = 1 + O[(nb)^{-1}]$. $\diamondsuit$

Proof of Theorems 19–20. Let $\eta_i = e_ie_{i+k} - \gamma_k$ be defined in (4.11). Note that $\mathbb{E}(\eta_i) = 0$ and η_i is of the form (4.3). As in (4.4), consider the coupling process $\eta_i' = e_i'e_{i+k}' - \gamma_k$, where

e_i' is defined in (4.4). By (4.5) and the inequality $\|XY\|_{q/2} \le \|X\|_q \times \|Y\|_q$ for $X, Y \in \mathcal{L}^q$,

$$\|\eta_i' - \eta_i\|_{q/2} \le \|e_i'(e_{i+k}' - e_{i+k})\|_{q/2} + \|e_{i+k}(e_i' - e_i)\|_{q/2} = O(\rho^i). \tag{4.43}$$

Thus $\{\eta_i\}$ satisfies GMC$(q/2, \rho)$. For $q > 4$, $n^{-1/2+1/\min(q/2,4)} \log n = o(1)$. By a similar strong approximation in (4.31) for $\sum_{i=\lfloor n\delta\rfloor}^{\lfloor nt\rfloor} \eta_i$, $\{n^{-1/2} \sum_{i=\lfloor n\delta\rfloor}^{\lfloor nt\rfloor} \eta_i\}_t \Rightarrow \tau(k)\{\mathbb{B}_{t-\delta}\}_t$, where $\tau^2(k)$, defined in (4.11), is the long-run variance of $\{\eta_i\}$. By Lemma 3, Theorem 19 follows. Finally, Theorem 20 follows from Theorem 19 and the continuous mapping theorem. $\diamondsuit$

Proof of Theorem 21. Let Δ_i and ζ_i be defined as in the proof of Lemma 3. By the proof of Lemma 3, $\max_i |\Delta_i| = O[b^2 + (nb)^{-1}]$. For ζ_i, by (4.19) in Proposition 7, $\max_i |\zeta_i| = O_p\{(nb)^{-1} n^{1/4} \log n + [\log n/(nb)]^{1/2}\}$. Thus, under the condition $nb^2/(\log^2 n) \to \infty$,

$$\hat{e}_i = e_i + \Delta_i + \zeta_i = e_i + O_p(\epsilon_n), \quad \text{where} \quad \epsilon_n = b^2 + \Big(\frac{\log n}{nb}\Big)^{1/2}, \tag{4.44}$$

uniformly over i. Also, by Theorem 19, $\hat{\gamma}_k(1-\delta) = \gamma_k + O_p(1/\sqrt{n}) = \gamma_k + O_p(\epsilon_n)$. Recall $\eta_i = e_i e_{i+k} - \gamma_k$ in (4.11). Then $\hat{e}_i \hat{e}_{i+k} - \hat{\gamma}_k(1-\delta) = \eta_i + O_p(\epsilon_n)\upsilon_i$, where $\upsilon_i = 1 + |e_i| + |e_{i+k}|$. Therefore,

$$\begin{aligned}\widehat{\operatorname{cov}(\eta_0, \eta_j)} &= \frac{1}{\ell_j} \sum_{i=\lfloor n\delta\rfloor}^{\lfloor n(1-\delta)\rfloor - j} [\eta_i + O_p(\epsilon_n)\upsilon_i][\eta_{i+j} + O_p(\epsilon_n)\upsilon_{i+j}] \\ &= \frac{1}{\ell_j} \sum_{i=\lfloor n\delta\rfloor}^{\lfloor n(1-\delta)\rfloor - j} \eta_i\eta_{i+j} + \frac{O_p(\epsilon_n)}{\ell_j} \sum_{i=\lfloor n\delta\rfloor}^{\lfloor n(1-\delta)\rfloor - j} (|\upsilon_i\eta_{i+j}| + |\eta_i\upsilon_{i+j}| + |\upsilon_i\upsilon_{i+j}|).\end{aligned}$$

Since $I = o(n)$, we have $\ell_j \asymp n(1-2\delta)$. Consequently, we have the decomposition

$$\hat{\tau}^2(k) = \tau_I^2 + R_n + O_p(1)\Omega_n, \tag{4.45}$$

where

$$\tau_I^2 = \mathbb{E}(\eta_0^2) + 2\sum_{j=1}^{I} \Big(1 - \frac{j}{I+1}\Big)\mathbb{E}(\eta_0\eta_j),$$

$$
\begin{aligned}
R_n &= \Gamma_0 + 2\sum_{j=1}^{I}\Big(1-\frac{j}{I+1}\Big)\Gamma_j, \quad \text{with} \quad \Gamma_j = \frac{1}{\ell_j}\sum_{i=\lfloor n\delta\rfloor}^{\lfloor n(1-\delta)\rfloor - j}[\eta_i\eta_{i+j} - \mathbb{E}(\eta_0\eta_j)],\\
\Omega_n &= \frac{\epsilon_n}{n}\sum_{j=0}^{I}\sum_{i=\lfloor n\delta\rfloor}^{\lfloor n(1-\delta)\rfloor - j}(|\upsilon_i\eta_{i+j}| + |\eta_i\upsilon_{i+j}| + |\upsilon_i\upsilon_{i+j}|).
\end{aligned}
$$

In what follows we prove: (i) $\tau_I^2 \to \tau^2(k)$; (ii) $R_n = o_p(1)$; and (iii) $\Omega_n = o_p(1)$.

(i) $\tau_I^2 \to \tau^2(k)$. By the proof of Theorem 19, $\{\eta_i\}$ satisfies GMC$(4,\rho)$ in (4.5). By Proposition 5, $\mathbb{E}(\eta_0\eta_j)$ decays exponentially fast. Thus, by the Dominated Convergence Theorem, $\tau_I^2 \to \tau^2(k)$ as $I \to \infty$.

(ii) $R_n = o_p(1)$. Note that (4.28) is a general result: as long as $\{e_i\}$ satisfies the condition GMC(q,ρ) for some $q \geq 2$, then (4.28) holds. For the summands $\{\tilde{e}_i := \eta_i\eta_{i+j} - \mathbb{E}(\eta_0\eta_j)\}$ in Γ_j, by the assumption GMC(8,ρ) and the same argument in (4.43), we can show that the new process $\{\tilde{e}_i\}$ satisfies GMC$(2,\rho)$. Applying (4.28), we obtain $\|\Gamma_j\|_2^2 = O(1/\ell_j) = O(1/n)$, uniformly in j. Thus, by the triangle inequality, $\|R_n\|_2 \leq \|\Gamma_0\|_2 + 2\sum_{j=1}^{I}\|\Gamma_j\|_2 = O(I/\sqrt{n}) \to 0$.

(iii) $\Omega_n = o_p(1)$. Under the assumption $\mathbb{E}(|e_i|^3) < \infty$, there exists a constant $c < \infty$ such that $\mathbb{E}|\upsilon_i\eta_{i+j}| + \mathbb{E}|\eta_i\upsilon_{i+j}| + \mathbb{E}|\upsilon_i\upsilon_{i+j}| < c$. Therefore, $\mathbb{E}(\Omega_n) = O(\epsilon_n I) = o(1)$. $\diamondsuit$

Proof of Theorem 22. It suffices to show that, with probability approaching one, $T_n^{(k)}/\sqrt{n} > c$ for some constant $c > 0$ and $\hat{\tau}(k)/\sqrt{n} \to 0$.

Recall the decomposition of D_m in (4.40). Since Δ_i depends only on f, Lemma 4 still holds. In the right hand side of (4.40), by breaking the summation into two summations according to the two processes and applying the same argument in Lemma 3 to the two summations separately, we see that Lemma 3 still holds. Thus, by Lemma 3 and (4.10),

$$
\hat{\gamma}_k(1-\delta) = \frac{\sum_{i=\lfloor n\delta\rfloor}^{\lfloor n(1-\delta)\rfloor}\hat{e}_i\hat{e}_{i+k}}{\lfloor n(1-\delta)\rfloor - \lfloor n\delta\rfloor + 1} = \frac{\sum_{i=\lfloor n\delta\rfloor}^{\lfloor n(1-\delta)\rfloor}e_ie_{i+k}}{\lfloor n(1-\delta)\rfloor - \lfloor n\delta\rfloor + 1} + o_p(n^{-1/2}). \qquad (4.46)
$$

By the proof of Theorems 19–20, $\sum_{i=\lfloor n\delta\rfloor}^{\lfloor n(1-\delta)\rfloor}e_ie_{i+k} = \sum_{i=\lfloor n\delta\rfloor}^{\lfloor nt_0\rfloor}(e_ie_{i+k} - \gamma_{k,1}) + \sum_{i=\lfloor nt_0\rfloor+1}^{\lfloor n(1-\delta)\rfloor}(e_ie_{i+k} - \gamma_{k,2}) + \{\lfloor nt_0\rfloor - \lfloor n\delta\rfloor + 1\}\gamma_{k,1} + \{\lfloor n(1-\delta)\rfloor - \lfloor nt_0\rfloor\}\gamma_{k,2} = O_p(\sqrt{n}) + \{\lfloor nt_0\rfloor - \lfloor n\delta\rfloor + 1\}\gamma_{k,1} +$

$\{\lfloor n(1-\delta)\rfloor - \lfloor nt_0\rfloor\}\gamma_{k,2}$. Thus, by (4.46),

$$\hat{\gamma}_k(1-\delta) \to \frac{(t_0-\delta)\gamma_{k,1} + (1-\delta-t_0)\gamma_{k,2}}{1-2\delta}, \quad \text{in probability.} \tag{4.47}$$

Similarly, $\hat{\gamma}_k(t_0) \to \gamma_{k,1}$ in probability. Therefore, $\hat{\gamma}_k(t_0) - \hat{\gamma}_k(1-\delta) \to (\gamma_{k,1}-\gamma_{k,2})(1-\delta-t_0)/(1-2\delta)$ in probability. Recall $T_n^{(k)}(t)$ in (4.12). Then

$$\frac{T_n^{(k)}}{\sqrt{n}} \geq \frac{|T_n^{(k)}(t_0)|}{\sqrt{n}} = \frac{\lfloor nt_0\rfloor - \lfloor n\delta\rfloor + 1}{n}|\hat{\gamma}_k(t_0) - \hat{\gamma}_k(1-\delta)| \to c, \quad \text{in probability,}$$

where $c = |\gamma_{k,1} - \gamma_{k,2}|(1-\delta-t_0)(t_0-\delta)/(1-2\delta) > 0$.

It remains to show $\hat{\tau}(k) = o_p(\sqrt{n})$. By (4.44) and (4.47), $\hat{e}_i\hat{e}_{i+k} - \hat{\gamma}_k(1-\delta) = e_ie_{i+k} + O_p(\epsilon_n)(|e_i| + |e_{i+k}|) + O_p(1)$, uniformly over i. Then it can be easily seen that

$$\hat{\tau}^2(k) \leq |\widehat{\mathrm{cov}(\eta_0, \eta_0)}| + 2\sum_{j=1}^{I} |\widehat{\mathrm{cov}(\eta_0, \eta_j)}| = O_p(I). \tag{4.48}$$

Since $I = o(n)$, we conclude that $\hat{\tau}(k)/\sqrt{n} \to 0$ in probability, completing the proof. $\diamondsuit$

Chapter 5

On Time-varying Exogenous Autoregressive Models

5.1 Model and Methodology

Denote by $C^k[0,1]$ the set of functions on $[0,1]$ with k-th order continuous derivatives.

5.1.1 Time-varying exogenous autoregressive models

For the classical AR(p) model $X_i = \alpha + \sum_{j=1}^{p} \phi_j X_{i-j} + \varepsilon_i$, the current observation depends on the past p observations. As a natural extension, the AR(p) model with exogenous inputs introduces some exogenous variables that can affect the current observation. Specifically, the AR(p) model with exogenous time series inputs $\{Z_i\}$ of order q is given by

$$\text{ARX}(p,q): \quad X_i = \alpha + \sum_{j=1}^{p} \phi_j X_{i-j} + \sum_{r=1}^{q} \beta_r Z_{i+1-r} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{5.1}$$

where $\{X_i\}$ are the main time series of interest, $\{Z_i\}$ are external time series inputs, $\{\varepsilon_i\}$ are i.i.d. errors, and $(\alpha, \phi_1, \ldots, \phi_p, \beta_1, \ldots, \beta_q)^T$ are unknown parameters. Compared to the ordinary AR model, the exogenous AR model has the capability of incorporating external information outside the AR model. One serious restriction of model (5.1) is that the coefficients $(\alpha, \phi_1, \ldots, \phi_p, \beta_1, \ldots, \beta_q)^T$ are the same across time, i.e., time-invariant.

To model time-varying structure, we propose the generalization of (5.1):

$$X_i = \alpha(i/n) + \sum_{j=1}^{p} \phi_j(i/n) X_{i-j} + \sum_{r=1}^{q} \beta_r(i/n) Z_{i+1-r} + \sigma(i/n)\varepsilon_i, \quad i = 1, \ldots, n, \tag{5.2}$$

where $(\alpha(t), \phi_1(t), \ldots, \phi_p(t), \beta_1(t), \ldots, \beta_q(t), \sigma(t))^T$ are functions in $t \in [0,1]$. We call (5.2) the time-varying-coefficient AR model with exogenous inputs, denoted by TV-ARX(p, q).

The proposed model (5.2) has several nice features over (5.1). First, (5.2) allows the coefficients to vary over time and thus it can model time-varying nonstationary pattern. Second, we do not impose any structural (other than some smoothness) assumptions on the functions $(\alpha(t), \phi_1(t), \ldots, \phi_p(t), \beta_1(t), \ldots, \beta_q(t), \sigma(t))$. That is, we let the data speak for themselves. Third, if we assume that the involved functions are continuous, then they can be approximated by local constants within each small local time window, and consequently model (5.2) is a locally stationary process. This local stationarity feature allows us to use local data to address model estimation and inference. Furthermore, by incorporating the exogenous inputs $\{Z_i\}$, model (5.2) is more flexible than the locally stationary AR model in Subba Rao (1970) and the piecewise stationary AR model in Davis et al. (2006).

5.1.2 Local approximation and derivative process

As shown in the literature [see Dahlhaus (2012) for a survey], under appropriate conditions, at each time $t \in [0,1]$, locally stationary processes can be approximated by a stationary process. To illustrate the idea, we consider the nonlinear time-varying AR (TV-AR) model

$$Y_i = G(i/n, Y_{i-1}, \ldots, Y_{i-p}) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{5.3}$$

where $G(t, y_1, \ldots, y_p)$ is a measurable function that is continuous in $t \in [0,1]$. At each given time point $t \in [0,1]$, the local stationary approximation of (5.3) is

$$Y_i(t) = G(t, Y_{i-1}(t), \ldots, Y_{i-p}(t)) + \varepsilon_i, \quad i = 1, \ldots, n. \tag{5.4}$$

Since t is fixed, $\{Y_i(t)\}_{i\in\mathbb{Z}}$ is a stationary nonlinear AR process. To study locally stationary processes, an important step is to bound $|Y_i - Y_i(t)|$, i.e., the distance between the nonstationary process $\{Y_i\}_{i\in\mathbb{Z}}$ in (5.3) and the stationary process $\{Y_i(t)\}_{i\in\mathbb{Z}}$ in (5.4).

Example 7. Consider the first-order linear TV-AR model $Y_i = a(i/n)Y_{i-1}+\varepsilon_i$ for a function $a(\cdot) \in C^1[0,1]$ such that $\sup_{t\in[0,1]}|a(t)| < 1$. At any time $t \in [0,1]$, the approximation process is $Y_i(t) = a(t)Y_{i-1}(t) + \varepsilon_i$, which has the stationary solution (t is fixed): $Y_i(t) = \sum_{j=0}^{\infty} a(t)^j \varepsilon_{i-j}, i \in \mathbb{Z}$. Furthermore, the uniform approximations hold:

$$\begin{aligned} Y_i &= Y_i(t) + O_p(|i/n - t| + 1/n), \\ Y_i &= Y_i(t) + (i/n - t)Y_i'(t) + O_p(|i/n - t|^2 + 1/n). \end{aligned} \tag{5.5}$$

where $Y_i'(t) = \partial Y_i(s)/\partial s|_{s=t}$ is the derivative process. The two approximations can be viewed as the local constant and local linear Taylor-type expansions of Y_i around $Y_i(t)$. In particular, if $i/n \to t$, then $|Y_i - Y_i(t)| \xrightarrow{p} 0$. The same result can be extended to the p-th order TV-AR $Y_i = \sum_{j=1}^{p} a_j(i/n)Y_{i-1}+\varepsilon_i$ for appropriate functions $a_1(\cdot), \ldots, a_p(\cdot)$. We refer the reader to Dahlhaus (2012) and references therein for detailed discussion.

Now we extend the idea in Example 7 to the TV-ARX model (5.2). First, we need to impose some assumptions on the exogenous time series inputs $\{Z_i\}_{i\in\mathbb{Z}}$. Motivated by Example 7, we assume that $\{Z_i\}_{i\in\mathbb{Z}}$ is a locally stationary process as defined below.

Assumption 3. For $\{Z_i\}_{i\in\mathbb{Z}}$ in (5.2), there exists a stationary process $\{Z_i(t)\}_{i\in\mathbb{Z}}$ such that $\sup_{t\in[0,1]}\{\mathbb{E}[Z_i^2(t)] + \mathbb{E}[Z_i'^2(t)]\} < \infty$ and the uniform approximations hold:

$$\begin{aligned} Z_i &= Z_i(t) + O_p(|i/n - t| + 1/n), \\ Z_i &= Z_i(t) + (i/n - t)Z_i'(t) + O_p(|i/n - t|^2 + 1/n). \end{aligned} \tag{5.6}$$

Assumption 3 is satisfied for many locally stationary processes, including the TV-AR model in Example 7 and the time-varying ARCH model [Dahlhaus and Subba Rao (2006)].

Assumption 4. (i) $\{\varepsilon_i\}$ are i.i.d. with $\mathbb{E}(\varepsilon_i) = 0$, $\mathbb{E}(\varepsilon_i^2) = 1$, $\mathbb{E}(\varepsilon_i^4) < \infty$. (ii) $\alpha(\cdot)$, $\phi_1(\cdot), \ldots, \phi_p(\cdot)$, $\beta_1(\cdot), \ldots, \beta_q(\cdot) \in C^2[0,1]$, $\sigma(\cdot) \in C^1[0,1]$, $\sigma(\cdot) > 0$ (iii) There exists $\rho > 1$ such that, for each $t \in [0,1]$, all roots of $\sum_{j=1}^{p} \phi_j(t)z^j = 1$ have modulus greater than ρ.

Assumption 4(ii) ensures local continuity. Intuitively, if the data can jump without any continuity, it would be impossible to draw inference. Assumption 4(iii) is in line with the usual causal AR assumption that the AR polynomial has all roots outside the unit disc.

As in Example 7, under Assumptions 3–4, by the argument in the proof of Theorem 2.3 in Dahlhaus (1996), the nonstationary process $\{X_i\}_{i\in\mathbb{Z}}$ in (5.2) has the stationary approximation process:

$$X_i(t) = \alpha(t) + \sum_{j=1}^{p} \phi_j(t) X_{i-j}(t) + \sum_{r=1}^{q} \beta_r(t) Z_{i+1-r}(t) + \sigma(t)\varepsilon_i, \quad i = 1, \ldots, n. \tag{5.7}$$

For each fixed t, $\{X_i(t)\}_{i\in\mathbb{Z}}$ is a stationary ARX of the form (5.1). Furthermore, similar to (5.5)–(5.6), the local constant and local linear Taylor-type approximations hold:

$$\begin{aligned} X_i &= X_i(t) + O_p(|i/n - t| + 1/n), \\ X_i &= X_i(t) + (i/n - t)X_i'(t) + O_p(|i/n - t|^2 + 1/n). \end{aligned} \tag{5.8}$$

By the first equation, $X_i \approx X_i(t)$ for $i/n \approx t$, leading to local stationarity.

For each fixed t, let $\Phi(B) = 1 - \sum_{j=1}^{p} \phi_j(t) B^j$ be the AR polynomial with B being the backshift operator. Under Assumption 4(iii), $\Phi(B)^{-1} = \sum_{j=0}^{\infty} \psi_j(t) B^j$ for some $\{\psi_j(t)\}_{j\geq 0}$ satisfying $\psi_j(t) = O(\rho^{-j})$. By the well-known theory for causal AR models, $\{X_i(t)\}_{i\in\mathbb{N}}$ in (5.7) has the causal representation

$$\begin{aligned} X_i(t) &= \frac{\alpha(t)}{1 - \sum_{j=1}^{p} \phi_j(t)} + \sum_{r=1}^{q} \beta_r(t) \Phi(B)^{-1} Z_{i+1-r}(t) + \sigma(t)\Phi(B)^{-1}\varepsilon_i \\ &= \frac{\alpha(t)}{1 - \sum_{j=1}^{p} \phi_j(t)} + \sum_{r=1}^{q} \left[\beta_r(t) \sum_{j=0}^{\infty} \psi_j(t) Z_{i+1-r-j}(t)\right] + \sigma(t) \sum_{j=0}^{\infty} \psi_j(t)\varepsilon_{i-j}. \end{aligned} \tag{5.9}$$

Since $\{Z_i(t)\}_{i\in\mathbb{Z}}$ is stationary for each fixed t, we can easily show

$$\mathbb{E}[X_i(t)] = g(t), \quad \text{where} \quad g(t) = \frac{\alpha(t) + \mathbb{E}[Z_0(t)] \sum_{r=1}^{q} \beta_r(t)}{1 - \sum_{j=1}^{p} \phi_j(t)}. \tag{5.10}$$

5.1.3 Nonparametric estimation and asymptotic theory

In this section we assume that the orders (p, q) in (5.2) are known, and the case with unknown (p, q) will be discussed in Section 5.1.5 below. Define the vectors

$$U_i = (1, X_{i-1}, \ldots, X_{i-p}, Z_i, \ldots, Z_{i+1-q})^T, \tag{5.11}$$

$$\theta(\cdot) = (\alpha(\cdot), \phi_1(\cdot), \ldots, \phi_p(\cdot), \beta_1(\cdot), \ldots, \beta_q(\cdot))^T. \tag{5.12}$$

Then we can rewrite (5.2) as

$$X_i = U_i^T \theta(i/n) + \sigma(i/n)\varepsilon_i, \quad i = 1, \ldots, n. \tag{5.13}$$

We can view U_i and $\theta(\cdot)$ as the covariates and time-varying nonparametric functional coefficients, respectively. Therefore, model (5.13) can be viewed as a varying-coefficient model [Hastie and Tibshirani (1993)] with time series data, and it possesses both the flexibility of nonparametric modeling and the nice interpretability of linear models.

Suppose we wish to estimate $(\theta(\cdot), \sigma^2(\cdot))$ at $t \in (0, 1)$. Assuming independence between U_i and ε_i. By Taylor's linear approximation $\theta(i/n) \approx \theta(t) + (i/n - t)\theta'(t)$, we have

$$\mathbb{E}(X_i|U_i) = U_i^T \theta(i/n) \approx U_i^T \theta(t) + [(i/n - t)U_i^T]\theta'(t), \quad \text{for } i/n \approx t.$$

Therefore, in the neighborhood of $i/n \approx t$, model (5.13) can be approximated by a linear model with covariates U_i and $(i/n - t)U_i$ and corresponding coefficients $\theta(t)$ and $\theta'(t)$. We adopt the following local linear kernel smoothing estimation of $(\theta(t), \theta'(t))$:

$$\Big(\hat{\theta}(t), \hat{\theta}'(t)\Big) = \operatorname*{argmin}_{(\theta,\vartheta)} \sum_{i=1}^{n} \Big[X_i - U_i^T\theta - (i/n - t)U_i^T\vartheta\Big]^2 K_i(t), \quad K_i(t) = K\Big(\frac{i/n - t}{b_n}\Big), \tag{5.14}$$

for a kernel function $K(\cdot)$ and a bandwidth $b_n > 0$. In Section 5.1.4 we show that the local linear estimation can adapt to the unknown local stationarity structure of $U_i(t)$.

It can be shown that the $\hat{\theta}(t)$ component of the solution to (5.14) is given by

$$\hat{\theta}(t) = (M_0 - M_1 M_2^{-1} M_1)^{-1}(N_0 - M_1 M_2^{-1} N_1), \tag{5.15}$$

where

$$M_j = \sum_{i=1}^{n} (i/n - t)^j U_i U_i^T K_i(t), \quad j = 0, 1, 2; \tag{5.16}$$

$$N_j = \sum_{i=1}^{n} (i/n - t)^j U_i X_i K_i(t), \quad j = 0, 1. \tag{5.17}$$

To study the asymptotic property of $\hat{\theta}(t)$, we impose the following conditions:

Assumption 5. (i) For each i, Z_i in (5.2) is a function of some i.i.d. innovations $\{\eta_j\}_{j\leq i}$ with $\mathbb{E}(\eta_i) = 0$ and $\mathbb{E}(\eta_i^4) < \infty$, and $\{\eta_i\}_{i\in\mathbb{Z}}$ is independent of the errors $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ in (5.2). (ii) In Assumption 3, $\{Z_i(t)\}_{i\in\mathbb{Z}}$ is a functional-coefficient process

$$Z_i(t) = \mu(t) + \sum_{j=0}^{\infty} \alpha_j(t)\eta_{i-j}, i = 1, \ldots, n, \tag{5.18}$$

where $\mu(t), \alpha_j(t) \in C^1[0, 1]$ and $\sup_{t\in[0,1]} |\alpha_j(t)| = O(\varrho^j)$ for some $\varrho \in (0, 1)$.

The moving average linear process condition in Assumption 5(ii) is not restrictive. In fact, by Wold's decomposition, any zero-mean covariance-stationary process can be written as a moving average linear process (assuming no deterministic component). The exponentially decaying rate on $\sup_{t\in[0,1]} |\alpha_j(t)|$ can be weakened to some polynomial rate.

Assumption 6. *(i) $K(\cdot)$ has bounded support, is symmetric and continuously differentiable, and $\int_{\mathbb{R}} K(u)du = 1$. (ii) $b_n \to 0$ and $nb_n^3 \to \infty$.*

Recall the stationary approximation processes $\{X_i(t)\}_{i\in\mathbb{Z}}$ in (5.7) and $\{Z_i(t)\}_{i\in\mathbb{Z}}$ in Assumption 3. Define the stationary approximation of U_i in (5.11) as

$$U_i(t) = (1, X_{i-1}(t), \ldots, X_{i-p}(t), Z_i(t), \ldots, Z_{i+1-q}(t))^T. \tag{5.19}$$

Since $K(\cdot)$ has bounded support, in (5.14) it suffices to consider $i/n - t = O(b_n)$. Thus, under $nb_n^3 \to \infty$ in Assumption 6(ii), by (5.6) and (5.8), we have the uniform approximations

$$U_i = U_i(t) + O_p(b_n) \quad \text{and} \quad U_i = U_i(t) + (i/n - t)U_i'(t) + O_p(b_n^2). \tag{5.20}$$

Theorem 23. *Suppose Assumptions 3–6 hold. Then, for any* $t \in (0, 1)$,

$$H_n(t) := \sqrt{nb_n}\left[\hat{\theta}(t) - \theta(t) - b_n^2 \mu_K \frac{\theta''(t)}{2} + o_p(b_n^2)\right] \Rightarrow N\Big(0, s(t)\Big), \tag{5.21}$$

where $\mu_K = \int_{\mathbb{R}} u^2 K(u)du$, $U_i(t)$ *is defined as in (5.19), and*

$$s(t) \quad = \quad \sigma^2(t)\Big\{\mathbb{E}[U_0(t)U_0(t)^T]\Big\}^{-1} \int_{\mathbb{R}} K^2(u)du.$$

Furthermore, $H_n(t)$ *and* $H_n(t')$ *defined in (5.21) are asymptotically independent for* $t \neq t'$.

5.1.4 Local linear versus local constant estimation: Local linear estimation is adaptive to model nonstationarity

In (5.14), we approximate $\theta(i/n)$ by a local linear function. An alternative approach is the local constant regression (i.e., approximate $\theta(i/n)$ by the constant $\theta(t)$):

$$\tilde{\theta}(t) = \operatorname*{argmin}_{\theta} \sum_{i=1}^{n} \Big[X_i - U_i^T \theta\Big]^2 K_i(t) = \left[\sum_{i=1}^{n} U_i U_i^T K_i(t)\right]^{-1} \sum_{i=1}^{n} U_i X_i K_i(t). \tag{5.22}$$

It is well-known that the local linear estimation can reduce the boundary effect. In this section we show another interesting phenomenon that the local linear estimation $\hat{\theta}(t)$ in (5.14) can adapt to the unknown nonstationarity structure whereas the local constant estimation $\tilde{\theta}(t)$ in (5.22) can be strongly adversely affected by the local stationarity.

First we briefly review nonparametric estimation for model $Y_i = f(x_i) + \varepsilon_i$, where x_i are design points, ε_i are noises, and $f(\cdot)$ is an unknown function of interest. Suppose we use bandwidth b_n, and denote by $\hat{f}_{\mathrm{LC}}(x)$ and $\hat{f}_{\mathrm{LL}}(x)$, respectively, the local constant and local linear estimation of $f(x)$. It is well-known that [see Fan and Gijbels (1996)],

(a) For random-design case, i.e., when x_i are i.i.d. random points with density function $p_X(x)$, the local linear estimation $\hat{f}_{\mathrm{LL}}(x)$ has asymptotic bias $b_n^2 \mu_K f''(x)/2$ with $\mu_K = \int_{\mathbb{R}} u^2 K(u) du$, compared to the bias $b_n^2 \mu_K [f''(x)/2 + f'(x) p'_X(x)/p_X(x)]$ of the local constant estimation $\hat{f}_{\mathrm{LC}}(x)$.

(b) For fixed-design case $x_i = i/n$, $\hat{f}_{\mathrm{LL}}(x)$ and $\hat{f}_{\mathrm{LC}}(x)$ have the same bias $b_n^2 \mu_K f''(x)/2$. Intuitively, this can be explained by treating $x_i = i/n$ as "uniformly distributed" points in the random-design case (a) so that the "density" $p_X(x) = 1$ and $p'_X(x) = 0$.

From the above discussion (b), it is tempting to conjecture that, for the fixed-design model (5.13), the local constant estimation $\tilde{\theta}(t)$ in (5.22) and the local linear estimation $\hat{\theta}(t)$ in (5.14) have the same asymptotic bias. Surprisingly, Theorem 24 gives a negative answer.

Theorem 24. *Under the same conditions in Theorem 23, for $\tilde{\theta}(t)$ in (5.22), we have*

$$\sqrt{nb_n}\left[\tilde{\theta}(t) - \theta(t) - b_n^2 \mu_K \Big(\frac{\theta''(t)}{2} + B(t)\Big) + o_p(b_n^2)\right] \Rightarrow N\Big(0, s(t)\Big), \tag{5.23}$$

where $s(t)$ and μ_K are defined as in Theorem 23, and

$$B(t) = \Big\{\mathbb{E}[U_0(t) U_0(t)^T]\Big\}^{-1} \mathbb{E}\Big[U_0(t) U'_0(t)^T + U'_0(t) U_0(t)^T\Big] \theta'(t). \tag{5.24}$$

Compared to (5.21), the bias $b_n^2 \mu_K [\theta''(t)/2 + B(t)]$ in (5.23) has two sources. The first source $b_n^2 \mu_K \theta''(t)/2$ is the same as (5.21) and comes from the local approximation of $\theta(i/n)$. This first source of bias is an intrinsic feature of any nonparametric kernel smoothing local estimation. The second source $b_n^2 \mu_K B(t)$ comes from the local stationarity approximation in (5.20). If $U_i(t)$ does not depend on t (i.e., time-invariant), then $U'_0(t) = 0$ and $B(t) = 0$.

From Theorem 23, we conclude that the local linear estimation $\hat{\theta}(t)$ can adapt to the unknown local stationarity of $U_i(t)$ in the sense that its bias $b_n^2 \mu_K \theta''(t)/2$ does not depend on the nonstationarity structure of $U_i(t)$). By contrast, from Theorem 24, the bias of the local constant estimation $\tilde{\theta}(t)$ heavily depends on the process $U_i(t)$ as well as its derivative process $U'_i(t)$. The latter is highly undesirable since in practice we have no control on the underlying processes $U_i(t)$ and $U_i'(t)$. Our simulation study in Section 5.3.1 also confirms the superior performance of $\hat{\theta}(t)$. Therefore, in practice we recommend $\hat{\theta}(t)$.

5.1.5 Bandwidth selection and order selection

To implement the proposed methods, two practical issues are the bandwidth and order selection. The dependence and nonstationarity make the latter problem difficult, especially when both the bandwidth and order are unknown.

(i) Bandwidth selection

Bandwidth selection is a challenging issue in nonparametric regression. For local linear estimation of conditional mean function, Ruppert et al. (1995)'s automatic plug-in bandwidth selector works for independent data and tends to undersmooth dependent data. For nonparametric estimation of time trend with dependent data, the optimal bandwidth b_n^* usually satisfies $b_n^* = \nu b_n^o$, where b_n^o is the optimal bandwidth when ignoring the dependence and ν is some variance correction factor due to dependence. However, it is generally difficult to estimate ν; see the discussions in Altman (1990) and Wu and Zhao (2007).

Due to dependence, nonstationarity, and unknown order (p, q), bandwidth selection becomes even more challenging for our TV-ARX model. We propose a partial solution here. Recall the mean trend function $g(t) = \mathbb{E}[X_i(t)]$ in (5.10). Similar to (5.14), $g(t)$ can be estimated by the local linear regression (for a bandwidth h_n):

$$\Big(\hat{g}(t), \hat{g}'(t)\Big) = \underset{(g,g')}{\operatorname{argmin}} \sum_{i=1}^{n} \Big[X_i - g - (i/n - t)g'\Big]^2 K\Big(\frac{i/n - t}{h_n}\Big). \tag{5.25}$$

From (5.10), $g(t)$ is a function of $\theta(t) = (\alpha(t), \phi_1(t), \ldots, \phi_p(t), \beta_1(t), \ldots, \beta_q(t))^T$. Therefore, a good bandwidth h_n for estimating $g(t)$ is also a reasonable bandwidth choice for estimating $\theta(t)$. Motivated by this, we propose the following bandwidth choice for b_n in (5.14):

(a) In (5.25), use Ruppert et al. (1995)'s automatic plug-in bandwidth selector (implemented by the `R` command `dpill`) to obtain the optimal bandwidth h_n^* for estimating the mean trend function $g(t)$. This optimal bandwidth h_n^* ignores the dependence.

(b) Let $b_n = \nu h_n^*$, where ν is the variance correction factor due to dependence. Our simulation studies show that $\nu \in [2, 3]$ usually performs quite well. See Section 5.3.1.

In addition to the easy implementation, the above bandwidth selection has another appealing feature that it works regardless of the unknown order (p, q): the nonparametric mean function $g(t)$ in (5.10) automatically takes into account the unknown order (p, q). In Section 5.3.1 we compare the selected bandwidth to the true optimal bandwidth.

(ii) Order selection

For a model with k parameters, the order selection is usually achieved by minimizing some information criterion (IC) of the form $n \log \hat{\sigma}^2 + \ell(k)$, where $\hat{\sigma}^2$ is the noise variance estimator and $\ell(k)$ is some non-decreasing function in k penalizing the model complexity. For the TV-ARX model (5.2), we replace $n \log \hat{\sigma}^2$ in (5.26) with the aggregated version over $\hat{\sigma}^2(1/n), \ldots, \hat{\sigma}^2(n/n)$ and choose (p, q) by minimizing $\mathrm{IC}(p, q)$:

$$\mathrm{IC}(p, q) = \sum_{i=1}^{n} \log \hat{\sigma}^2(i/n) + \ell(p + q + 1). \tag{5.26}$$

For example, the Akaike information criterion (AIC) uses $\ell(k) = 2k$ and the Bayesian information criterion (BIC) uses $\ell(k) = k \log n$. However, there seems to be no universal rule on determining whether one should use AIC or BIC [Yang (2005)]. Our simulation study in Section 5.3.1 suggests that the BIC performs quite well, whereas the AIC always selects too large models. Therefore, we recommend BIC for the TV-ARX model.

The above order-independent bandwidth selection plays an important role in (5.26). First, without the order-independent bandwidth it would be infeasible to evaluate (5.26). Second, when the bandwidth also plays a role in $\mathrm{IC}(p, q)$, we cannot distinguish the effect of the bandwidth from that of the model order.

5.1.6 Some inference problems

In this section we discuss inference for $\theta(\cdot)$ in (5.12).

First, we can estimate $\sigma^2(t)$ based on the squared residuals:

$$\hat{\sigma}^2(t) = \frac{\sum_{i=1}^n [X_i - U_i^T \hat{\theta}(i/n)]^2 K_i(t)}{\sum_{i=1}^n K_i(t)}. \tag{5.27}$$

For $s(t)$ in (5.21), motivated by (5.44) in the proof section, we have the following estimator:

$$\hat{s}(t) = \hat{\sigma}^2(t) \left[\frac{1}{nb_n} \sum_{i=1}^n U_i U_i^T K_i(t) \right]^{-1} \int_{\mathbb{R}} K^2(u) du. \tag{5.28}$$

(i) Bias issue

In Theorem 23, the bias $b_n^2 \mu_K \theta''(t)/2$ contains the unknown derivative $\theta''(t)$. It is generally more difficult to estimate the second derivative $\theta''(t)$ than to estimate the function $\theta(t)$ itself. Furthermore, it requires an additional bandwidth to estimate the unknown derivative. In this section, we follow the convention in the nonparametric estimation literature and ignore the bias. In Section 5.2, we propose a sieve-wild bootstrap procedure, which can mimic the dependence and nonstationarity structure of the original data and therefore automatically incorporate the bias into the inference procedure.

(ii) Simultaneous confidence band

Let $\theta_k(\cdot)$ be the k-th component of $\theta(\cdot)$. We can use Theorem 23 to construct an asymptotic $(1-\alpha)$ point-wise confidence interval for $\theta_k(t)$ at each t. This point-wise confidence interval covers $\theta(t)$ with asymptotically correct probability for each given t.

Compared to a point-wise confidence interval, a simultaneous confidence band (SCB) can capture the overall variability over time. A pair of curves $(\ell_1(\cdot), \ell_2(\cdot))$ is an asymptotic $(1-\alpha)$ SCB for $\theta_k(\cdot)$ on a discrete set $\mathcal{T} = \{0 < t_1 < t_2 < \cdots < t_L < 1\}$ if

$$\lim_{n\to\infty} \mathbb{P}\{\ell_1(t) \le \theta_k(t) \le \ell_2(t), \quad \text{for all } t \in \mathcal{T}\} = 1 - \alpha. \tag{5.29}$$

Let $\hat{\theta}_k(t)$ be the k-th component of $\hat{\theta}(t)$ in (5.14) and $\hat{s}_k(t)$ be the k-th component of $\hat{s}(t)$ in (5.28). By the asymptotic independence between $\hat{\theta}(t)$ and $\hat{\theta}(t')$ (Theorem 23) for given

$t \neq t'$,

$$\mathbb{P}\{W_n \leq x\} \to \mathbb{P}\{|N(0,1)| \leq x\}^L, \quad \text{where} \quad W_n = \max_{t \in \mathcal{T}} \frac{\sqrt{nb_n}|\hat{\theta}_k(t) - \theta_k(t)|}{\sqrt{\hat{s}_k(t)}}. \tag{5.30}$$

Thus, an asymptotic $(1-\alpha)$ SCB for $\theta_k(\cdot)$ can be constructed as (ignoring the bias)

$$\left(\hat{\theta}_k(t) - z_{[1+(1-\alpha)^{1/L}]/2}\sqrt{\frac{\hat{s}_k(t)}{nb_n}}, \quad \hat{\theta}_k(t) + z_{[1+(1-\alpha)^{1/L}]/2}\sqrt{\frac{\hat{s}_k(t)}{nb_n}}\right), \quad t \in \mathcal{T}, \tag{5.31}$$

where $z_{[1+(1-\alpha)^{1/L}]/2}$ is the $[1+(1-\alpha)^{1/L}]/2$ standard normal quantile. This SCB covers the entire curve $\theta_k(t)$ on $t \in \mathcal{T}$ with asymptotic probability $(1-\alpha)$.

(iii) Testing the constancy of some coefficients

As in (ii) above, let $\theta_k(\cdot)$ be the k-th component of $\theta(\cdot)$. In practice, it is of interest to examine whether θ_k is indeed time-varying or constant. Specifically, we consider testing

$$H_0 : \theta_k(\cdot) \equiv \theta_k \text{ for some constant } \theta_k \text{ (i.e., a constant function).}$$

Since the SCB in (5.31) covers $\theta_k(\cdot)$ with asymptotically correct probability regardless of the form of $\theta_k(\cdot)$, we can use it as a reference quantity to test H_0 via the following procedure:

(a) Under H_0, estimate θ_k by $\hat{\theta}_k = L^{-1}\sum_{r=1}^{L}\hat{\theta}_k(t_r)$, where $t_1, \ldots, t_L$ are the grid points in $\mathcal{T}$ defined in (ii) above.

(b) If the constant line $\hat{\theta}_k$ is not entirely contained within the SCB in (5.31), we reject H_0.

If we wish to test $\theta_k(\cdot) \equiv 0$, we can check whether the SCB covers the horizontal zero line.

(iv) Testing the significance of exogenous variables

In (5.2), it is of interest to test whether the exogenous inputs $\{Z_i\}$ have a significant contribution to the main time series $\{X_i\}$. Specifically, we consider testing the null hypothesis

$$H_0 : \beta_1(\cdot) = \cdots = \beta_q(\cdot) \equiv 0. \tag{5.32}$$

Denote by $\hat{\beta}(\cdot)$ the $(\beta_1(\cdot), \ldots, \beta_q(\cdot))$ components of $\hat{\theta}(\cdot)$ in (5.14), and by $\hat{s}_\beta(\cdot)$ the corresponding $q \times q$ sub-matrix (corresponding to the $\hat{\beta}$ components) of $\hat{s}(t)$ in (5.28). By Theorem 23, under H_0, for each t, $T_n(t) := nb_n\hat{\beta}(t)^T[\hat{s}_\beta(t)]^{-1}\hat{\beta}(t) \Rightarrow \chi^2_q$, the chi-squared distribution with q degrees of freedom. A large value of $T_n(t)$ indicates a significant departure of $(\beta_1(t), \ldots, \beta_q(t))$ from zero at the point t. Therefore, a global measure of the deviation of $(\beta_1(\cdot), \ldots, \beta_q(\cdot))$ from zero on the set $\mathcal{T}$ (defined as in (ii) above) is

$$T_n = \max_{t \in \mathcal{T}} T_n(t), \quad \text{where} \quad T_n(t) = nb_n\hat{\beta}(t)^T[\hat{s}_\beta(t)]^{-1}\hat{\beta}(t). \tag{5.33}$$

By the asymptotic independence between $\hat{\beta}(t)$ and $\hat{\beta}(t')$ for $t \neq t'$ (Theorem 23), under H_0,

$$T_n \Rightarrow \max\{\chi^2_q(1), \cdots, \chi^2_q(L)\}, \tag{5.34}$$

where $\chi^2_q(1), \cdots, \chi^2_q(L)$ are i.i.d. chi-squared random variables with q degrees of freedom.

5.2 Sieve-wild Bootstrap Inference

In finite sample applications, the bias and some asymptotically negligible terms may play a non-negligible role, an issue that becomes even more serious if the sample size is small. Various bootstrap methods have been introduced in the hope of yielding better empirical performance. In this section we introduce a sieve-wild bootstrap for our TV-ARX model.

To develop a bootstrap method for (5.2), we need to address the dependence and nonstationarity. If we know the underlying time series model subject to some parameters, the sieve bootstrap [Bühlmann (1997)] creates bootstrap samples by recursively using the same time series model with estimated parameters and resampled residuals. Thus, the sieve bootstrap can preserve the dependence. For independent observations, the wild bootstrap in Liu (1988) can adapt to the heterogeneity in errors. Therefore, to preserve both the dependence and nonstationarity, we propose combining the two bootstraps as follows:

(a) (**Calculate residuals**) Use $\hat{\theta}(\cdot)$ in (5.14) to compute

$$e_i = X_i - \left[\hat{\alpha}(i/n) + \sum_{j=1}^{p} \hat{\phi}_j(i/n) X_{i-j} + \sum_{r=1}^{q} \hat{\beta}_r(i/n) Z_{i+1-r}\right].$$

By the consistency of $\hat{\theta}(\cdot)$, these residuals $\{e_i\}_{i=1}^n$ are estimates of $\{\sigma(i/n)\varepsilon_i\}_{i=1}^n$.

(b) (**Wild bootstrap**) Generate i.i.d. $\{r_i\}_{i=1}^n$ satisfying $\mathbb{P}\{r_i = 1\} = \mathbb{P}\{r_i = -1\} = 1/2$ and obtain $e_i^b = e_i r_i$.

(b) (**Wild bootstrap**) Generate i.i.d. $\{r_i\}_{i=1}^n$ satisfying $\mathbb{P}\{r_i = 1\} = \mathbb{P}\{r_i = -1\} = 1/2$ and obtain $e_i^b = e_i r_i$.

(c) (**Sieve bootstrap**) Compute the bootstrap data $\{(X_i^b, Z_i)\}_{i=1}^n$ recursively through

$$X_i^b = \hat{\alpha}(i/n) + \sum_{j=1}^{p} \hat{\phi}_j(i/n) X_{i-j}^b + \sum_{r=1}^{q} \hat{\beta}_r(i/n) Z_{i+1-r} + e_i^b, \quad i = 1, \ldots, n. \tag{5.35}$$

In this section, the superscript "b" stands for the bootstrap version.

The bootstrap model (5.35) is a TV-ARX model of the same form (5.2) with estimated coefficients and new noises. Thus, this sieve bootstrap captures the dependence structure of the original data. Furthermore, the wild bootstrap error-generating-mechanism in step (b) can mimic the nonstationarity. To see this, note that

$$e_i^b = e_i r_i \approx \sigma(i/n)\varepsilon_i r_i = \sigma(i/n)\varepsilon_i^*, \quad \text{with} \quad \varepsilon_i^* = \varepsilon_i r_i.$$

Therefore, the nonstationarity term $\sigma(i/n)$ is automatically built into the bootstrap error e_i^b with the original innovation ε_i being replaced by the new innovation ε_i^*. The new innovations ε_i^* share several characteristics with ε_i, including same mean $\mathbb{E}(\varepsilon_i^*) = \mathbb{E}(\varepsilon_i) = 0$, same variance $\mathrm{var}(\varepsilon_i^*) = \mathrm{var}(\varepsilon_i)$, and same magnitude. Furthermore, if ε_i has a symmetric distribution, then ε_i^* has the same distribution as ε_i. See Davidson and Flachaire (2008) for more discussions.

For illustration, consider the bootstrap distribution of W_n in (5.30). With the bootstrap data $\{(X_i^b, Z_i)\}_{i=1}^n$ in (5.35), we use (5.14) and (5.27) with the same bandwidth b_n in (5.14) to obtain the bootstrap estimates $\hat{\theta}^b(t)$ and $\hat{s}^b(t)$. The bootstrap version of W_n is

$$W_n^b = \max_{t \in \mathcal{T}} \frac{\sqrt{nb_n}|\hat{\theta}_k^b(t) - \hat{\theta}_k(t)|}{\sqrt{\hat{s}_k^b(t)}}. \tag{5.36}$$

We can repeat the bootstrap procedure (b)–(c) above to obtain many realizations of W_n^b in (5.36), and use the empirical distribution of these realizations as the bootstrap distribution of W_n (opposed to the theoretical distribution in (5.30)). The $(1-\alpha)$ bootstrap SCB is

$$\left(\hat{\theta}_k(t) - z_{1-\alpha}^b \sqrt{\frac{\hat{s}_k(t)}{nb_n}}, \quad \hat{\theta}_k(t) + z_{1-\alpha}^b \sqrt{\frac{\hat{s}_k(t)}{nb_n}}\right), \quad t \in \mathcal{T}, \tag{5.37}$$

where $z_{1-\alpha}^b$ is the $(1-\alpha)$ sample quantile of the realizations W_n^b.

Similarly, to obtain the bootstrap distribution of T_n in (5.33), we consider the following bootstrap version of T_n:

$$T_n^b = \max_{t \in \mathcal{T}} T_n^b(t), \quad \text{where} \quad T_n^b(t) = nb_n[\hat{\beta}^b(t) - \hat{\beta}(t)]^T [\hat{s}_\beta^b(t)]^{-1} [\hat{\beta}^b(t) - \hat{\beta}(t)]. \tag{5.38}$$

Here $\hat{\beta}^b(t)$ and $\hat{s}_\beta^b(t)$ have the same meaning as $\hat{\beta}(t)$ and $\hat{s}_\beta(t)$ in (5.33) but using bootstrap data. The bootstrap distribution of T_n is the empirical distribution of realizations of T_n^b.

5.3 Numerical Results

5.3.1 Monte Carlo Study

In this section we evaluate the finite sample performance of the proposed methods. In our numerical analysis we use the standard normal kernel. Consider the TV-ARX(2,2) model:

$$\begin{aligned} X_i &= \alpha(i/n) + \phi_1(i/n)X_{i-1} + \phi_2(i/n)X_{i-2} + \beta_1(i/n)Z_i + \beta_2(i/n)Z_{i-1} + \sigma(i/n)\varepsilon_i, \\ Z_i &= i/n + 0.2\textstyle\sum_{r=0}^{3}(i/n)^r \eta_{i-r}. \end{aligned} \tag{5.39}$$

The exogenous inputs $\{Z_i\}$ follow a time-varying moving average of order 3. We use sample size $n = 600$ and coefficients $\alpha(t) = 0.5t$, $\phi_1(t) = -0.4t$, $\phi_2(t) = 0.4\cos(2\pi t)$, $\beta_1(t) =$

$0.5\exp(t)$, $\beta_2(t) = -0.6t$, $\sigma(t) = 0.2\sqrt{1+t^2}$. Throughout, we evaluate the estimation and inference methods on the set of 19 evenly spaced points $\mathcal{T} = \{0.05, 0.10, \ldots, 0.95\}$, and we approximate integral $\int_{0.05}^{0.95}$ in (5.40) below on the discrete set $\mathcal{T}$.

Bandwidth selection

For the estimator $\hat{\theta}(t)$ in (5.14), its performance can be evaluated by the mean integrated squared error (MISE) on the interval $[0.05, 0.95]$:

$$\mathrm{MISE}(\hat{\theta}) = \mathbb{E}[\mathrm{ISE}(\hat{\theta})], \quad \text{where} \quad \mathrm{ISE}(\hat{\theta}) = \int_{0.05}^{0.95} [\hat{\theta}(t) - \theta(t)]^T [\hat{\theta}(t) - \theta(t)] dt. \tag{5.40}$$

The true optimal bandwidth b_n^* is the minimizer of $\mathrm{MISE}(\hat{\theta})$. By averaging 1000 realizations of $\mathrm{ISE}(\hat{\theta})$ in (5.40), we find that the true optimal bandwidth b_n^* is about 0.15.

Now we consider the bandwidth selection described in Section 5.1.5(i). For h_n in the time trend estimation (5.25), the median among 1000 realizations of h_n selected by Ruppert et al. (1995)'s plug-in bandwidth selector is about 0.07. This bandwidth tends to undersmooth the data due to ignorance of the dependence. As discussed in the step (b) of Section 5.1.5(i), we consider $b_n = \nu \times 0.07$ for some $\nu > 1$. Specifically, we evaluate our estimation and inference methods using different bandwidths $b_n = 0.10, 0.15, 0.20, 0.25$. For these 4 bandwidths, the corresponding $\mathrm{MISE}(\hat{\theta})$ is $0.035, 0.031, 0.038, 0.053$, respectively. We have also tried some smaller bandwidths $b_n = 0.05, 0.07$, but they led to clear undersmoothing.

Order selection

As discussed above, the optimal bandwidth is about 0.15. To examine the sensitivity of the bandwidth choice to the proposed order selection procedure, we consider four choices of bandwidth $b_n = 0.10, 0.15, 0.20, 0.25$ and minimize the target function (5.26) over 25 combinations of (p, q) with $p = 0, 1, \ldots, 4$ and $q = 0, 1, \ldots, 4$. Using BIC $\ell(p+q+1) = (p+q+1)\log n$ in (5.26), Table 5.1 presents the counts of selected orders among 100 realizations. With $b_n = 0.10, 0.15, 0.20, 0.25$, the percentage of correct order selection $(2, 2)$ is 16%, 43%, 56%, 53%, respectively. Since there are 25 combinations of (p, q), the percentage of correct order selection is reasonably good, especially for larger bandwidths $b_n = 0.15, 0.20, 0.25$.

To get more insights into the *incorrect order selection* cases in Table 5.1, Table 5.2

Table 5.1: BIC order selection for (5.39). The "other" category includes all orders (p, q) with $p \leq 1$ or $q \leq 1$.

	Counts of selected orders (p, q) among 100 realizations									
b_n	(2,2)	(2,3)	(2,4)	(3,2)	(3,3)	(3,4)	(4,2)	(4,3)	(4,4)	other
0.10	16	8	9	5	7	7	18	13	17	0
0.15	43	10	11	12	1	4	10	6	3	0
0.20	56	9	8	12	3	3	8	1	0	0
0.25	53	4	5	13	1	0	21	0	3	0

tabulates one such typical realization of the BIC values for bandwidth $b_n = 0.25$ and different choices of (p, q). In this case, although BIC selects $(4, 2)$, BIC$(2, 2)$ is only 0.1% larger than BIC$(4, 2)$. As a practical approach, if we choose the relatively simpler model among several models with comparable BIC values, then $(2, 2)$ is a more reasonable choice than $(4, 2)$. By contrast, BIC(p, q) with $p \leq 1$ or $q \leq 1$ is substantially larger than BIC$(2, 2)$. In fact, if we choose the smallest order among all orders whose BIC values are within the 0.5% range of the smallest BIC, then the procedure almost always selects the correct order.

Table 5.2: One realization of BIC values in (5.26) (bandwidth $b_n = 0.25$). In each cell, the bracketed numbers are the order (p, q) and the negative number is the corresponding BIC value.

(0,0)	−1347	(0,1)	−1493	(0,2)	−1566	(0,3)	−1568	(0,4)	−1565
(1,0)	−1359	(1,1)	−1598	(1,2)	−1649	(1,3)	−1651	(1,4)	−1646
(2,0)	−1411	(2,1)	−1664	(2,2)	−1701	(2,3)	−1700	(2,4)	−1696
(3,0)	−1407	(3,1)	−1664	(3,2)	−1700	(3,3)	−1696	(3,4)	−1693
(4,0)	−1410	(4,1)	−1666	(4,2)	−1703	(4,3)	−1701	(4,4)	−1699

In Table 5.1, it is interesting to note that the "other" category (last column) has zero counts, which means that the BIC tends to select larger models ($p \geq 2$ and $q \geq 2$) than smaller models ($p \leq 1$ or $q \leq 1$). Since larger models include smaller models as a special case (by letting the corresponding coefficients be zero), it is better to select relatively larger models than smaller ones that miss some important variables.

When using the AIC $\ell(p + q + 1) = 2(p + q + 1)$ in (5.26), we find that the AIC always selects unnecessarily large models. For example, with $b_n = 0.25$, the number of selected orders is 1, 0, 4, 2, 1, 6, 11, 17, 58, 0, corresponding to the 10 selected orders in Table 5.1. Therefore, in practice, we recommend BIC.

Comparison of local linear and local constant estimation

To empirically examine the result in Section 5.1.4, in (5.40) we compute the MISE by averaging 1000 realizations of ISE. We define the relative efficiency gain of the local linear estimation $\hat{\theta}(\cdot)$ in (5.14) compared to the local constant estimation $\tilde{\theta}(\cdot)$ in (5.22) as

$$\mathrm{REG} = \frac{\mathrm{MISE}(\tilde{\theta}) - \mathrm{MISE}(\hat{\theta})}{\mathrm{MISE}(\tilde{\theta})} \times 100\%.$$

For bandwidths $b_n = 0.10, 0.15, 0.20, 0.25,$ the corresponding REG is 67%, 82%, 82%, 79%, respectively. Thus, there is a substantial efficiency gain by using the local linear estimation. This is in good agreement with the result in Section 5.1.4.

Simultaneous confidence band: sieve-wild bootstrap versus asymptotic theory

First, we compare the theoretical quantile $z_{[1+(1-\alpha)^{1/L}]/2}$ in (5.31) to the bootstrap quantile $z^b_{1-\alpha}$ in (5.37). The bootstrap quantile $z^b_{1-\alpha}$ depends on the specific realization of data $\{(X_i, Z_i)\}_{i=1}^n$. Since it is computationally expensive to compute $z^b_{1-\alpha}$ for each of a large number of realizations, to facilitate computation we adopt the following approach:

(a) Simulate 60 realizations of the data $\{(X_i, Z_i)\}_{i=1}^n$.

(b) For each realization above, we simulate 50 bootstrap realizations to compute 50 realizations of the bootstrap statistic W^b_n in (5.36).

(c) Pool together all $60 \times 50 = 3000$ realizations of the bootstrap statistic W^b_n, and use their sample $(1-\alpha)$ quantile as $z^b_{1-\alpha}$.

Table 5.3 tabulates the obtained bootstrap quantiles $z^b_{1-\alpha}$ for each of the five coefficients $\alpha(\cdot), \phi_1(\cdot), \phi_2(\cdot), \beta_1(\cdot), \beta_2(\cdot)$, using bandwidths $b_n = 0.10, 0.15, 0.20, 0.25$ and levels $1-\alpha =$ 95% and 90%. The bootstrap quantiles are consistently larger than the theoretical quantiles $z_{[1+(1-\alpha)^{1/L}]/2}$ in (5.31) (3.00 for 95% level or 2.77 for 90% level). A possible explanation is that, the asymptotic convergence in (5.30) ignores many negligible terms, which, however, may play a significant role in the finite sample setting. As a result, the theoretical distribution does not take into account the variability from such negligible terms and thus produces smaller quantiles. By contrast, the bootstrap method allows such small terms to speak for themselves in the bootstrap procedure, leading to better finite sample approximation. This is also supported by the empirical coverage probabilities below.

Table 5.3: Bootstrap quantiles $z^b_{1-\alpha}$ obtained using the procedure above. The theoretical quantile $z_{[1+(1-\alpha)^{1/L}]/2}$ in (5.31) is 3.00 for $1-\alpha = 95\%$ or 2.77 for $1-\alpha = 90\%$.

	$1-\alpha = 95\%$					$1-\alpha = 90\%$				
b_n	$\alpha(\cdot)$	$\phi_1(\cdot)$	$\phi_2(\cdot)$	$\phi_2(\cdot)$	$\phi_2(\cdot)$	$\alpha(\cdot)$	$\phi_1(\cdot)$	$\phi_2(\cdot)$	$\phi_2(\cdot)$	$\phi_2(\cdot)$
0.10	4.65	4.53	4.77	4.06	4.32	4.04	3.91	4.14	3.56	3.76
0.15	5.08	4.72	5.23	4.36	4.61	4.32	4.10	4.67	3.89	4.02
0.20	5.55	5.24	6.15	4.95	4.97	4.81	4.64	5.59	4.29	4.40
0.25	6.38	5.80	7.93	5.40	5.30	5.46	5.09	7.20	4.61	4.59

Next, we compare the performance of the asymptotic theory SCB in (5.31) and the sieve-wild bootstrap SCB in (5.37). Using the quantiles in Table 5.3, Table 5.4 presents the coverage probabilities, based on 1000 realizations, of the two SCBs for each of the five coefficients. For coefficients $\alpha(\cdot), \phi_1(\cdot), \beta_1(\cdot), \beta_2(\cdot)$, the bootstrap SCB has coverage probabilities close to the nominal level, and the performance is relatively robust against the bandwidth choice; by contrast, the asymptotic theory SCB suffers from serious under coverage. For coefficient $\phi_2(\cdot)$, the bootstrap SCB performs reasonably well for smaller bandwidths $b_n = 0.10$ and 0.15 but poorly for larger bandwidths $b_n = 0.20$ and 0.25. Since $\phi_2(\cdot)$ has the most curvature, a large b_n tends to oversmooth such curvature and leads to poor coverage.

Table 5.4: Empirical coverage probabilities of the bootstrap SCB (5.37) and asymptotic theory SCB (5.31).

		$1-\alpha = 95\%$					$1-\alpha = 90\%$				
b_n		$\alpha(\cdot)$	$\phi_1(\cdot)$	$\phi_2(\cdot)$	$\beta_1(\cdot)$	$\beta_2(\cdot)$	$\alpha(\cdot)$	$\phi_1(\cdot)$	$\phi_2(\cdot)$	$\beta_1(\cdot)$	$\beta_2(\cdot)$
0.10	Bootstrap	93.6	95.7	93.0	93.5	93.2	87.7	89.5	85.4	88.2	86.9
	Theory	67.6	73.0	55.3	77.2	71.8	60.4	64.6	48.4	69.1	65.6
0.15	Bootstrap	94.1	93.9	85.1	93.9	94.8	87.2	87.8	71.5	89.1	89.2
	Theory	62.4	65.2	15.2	72.6	69.9	55.8	59.2	10.4	65.6	64.6
0.20	Bootstrap	95.0	95.3	37.1	95.7	94.0	91.4	92.0	20.8	90.6	89.1
	Theory	59.5	62.4	0.1	68.3	61.9	52.6	53.3	0.0	61.4	54.5
0.25	Bootstrap	95.5	96.3	14.3	93.4	94.3	88.5	91.3	4.7	86.9	88.8
	Theory	42.4	52.3	0.0	56.1	55.8	36.9	44.6	0.0	48.7	47.8

Test for significance of exogenous inputs

Using the same notation in (5.39), we consider data from the following model

$$X_i = \alpha(i/n) + \phi_1(i/n)X_{i-1} + \phi_2(i/n)X_{i-2} + \delta\Big[\beta_1(i/n)Z_i + \beta_2(i/n)Z_{i-1}\Big] + \sigma(i/n)\varepsilon_i \quad (5.41)$$

Here the parameter δ controls the contribution of the exogenous inputs $\{Z_i\}$. In particular, if $\delta = 0$, then (5.41) becomes the TV-AR model (without exogenous inputs):

$$H_0 : X_i = \alpha(i/n) + \phi_1(i/n)X_{i-1} + \phi_2(i/n)X_{i-2} + \sigma(i/n)\varepsilon_i. \tag{5.42}$$

To test the significance of exogenous inputs (test H_0), we use the sieve-wild bootstrap method in Section 5.2 to obtain the bootstrap distribution of the test statistic T_n in (5.33). As in the SCB construction above, to facilitate computation we use 20×50 realizations (20 realizations of original data with 50 bootstrap realizations each) to obtain the bootstrap critical value of T_n. To examine the sensitivity of the bandwidth choice, we use four bandwidths $b_n = 0.10, 0.15, 0.20, 0.25$. Based on 200 realizations, Table 5.5 below tabulates the empirical size ($\delta = 0$) and power ($\delta = 0.1, 0.2, 0.3, 0.4$) at two significance levels 5% and 10%.

The empirical size is close to the significance level. As the model deviates from H_0, i.e., as δ moves away from zero, the power of the test quickly increases. Also, the test is quite robust against the bandwidth choice.

Table 5.5: Empirical size ($\delta = 0$) and power ($\delta = 0.1, 0.2, 0.3, 0.4$) at two significance levels 5% and 10%.

	5% significance, value of $\delta =$					10% significance, value of $\delta =$				
b_n	0.0	0.1	0.2	0.3	0.4	0.0	0.1	0.2	0.3	0.4
0.10	0.08	0.15	0.53	0.91	1.00	0.13	0.23	0.64	0.97	1.00
0.15	0.06	0.09	0.44	0.90	1.00	0.11	0.21	0.62	0.98	1.00
0.20	0.05	0.11	0.48	0.85	1.00	0.09	0.19	0.65	0.95	1.00
0.25	0.05	0.13	0.53	0.86	1.00	0.12	0.21	0.64	0.95	1.00

5.3.2 Application: The Stock Return-Inflation Puzzle

One main goal of the financial market research is to study how stock returns are affected by different factors. Among many other important factors such as unemployment and government's monetary policy, inflation plays an essential role in consumer spending, which in turn can affect stock market. In a seminal work, Fama (1981) argued that, as inflation increases, real variables (business activities) slow down, thus leading to decline in stock returns. Prior to Fama's theoretical framework, post-war empirical data [Bodie (1976); Nelson (1976)] also suggested that there is a significant negative correlation between stock returns and inflation rate. These empirical finding and Fama's theoretical work contradict the Fisher hypothesis that nominal asset returns (including interest rate) move one-to-one with inflation rate. This phenomenon is often called the stock return-inflation puzzle. Since then, there has been an extensive literature on how stock markets are affected by inflation.

We refer the reader to Nelson (1976), Kaul (1987), Balduzzi (1995), Gallagher and Taylor (2002), and Lee (2010) for further discussions.

When studying the predictability of stock returns from interest rate, Campbell (1987) (footnote 5 on page 376) pointed out that if the interest rate is nonstationary, then the asymptotic theory behind his statistical inference may be problematic. Similarly, in the context of stock return-inflation puzzle, the aforementioned empirical and theoretical work may be invalid if the inflation rate has unknown time-varying nonstationarity. As a modeling issue, it is hard to justify that we have more information about the dynamics of inflation rate than that about the return-inflation relationship. In other words, any assertion about the return-inflation relationship could be a result of a mis-specified assumption (e.g., stationarity or a particular model specification) on the inflation rate. Therefore, it is desirable to develop statistical methods that are robust against unknown nonstationarity and dependence, and our nonparametric TV-ARX model provides a reasonable framework.

A commonly used measure of inflation rate is the percentage change of the Consumer Price Index (CPI), which is a statistical measure, over time, of the prices of goods and services in major expenditure groups. The CPI data employed in this section is from Bureau of Labor Statistics *http://data.bls.gov/cgi-bin/surveymost*, from January 1982 to December 2012, with a total of 372 monthly observations. The right plot in Figure 5.1 is a time series plot of the inflation rate series. It clearly shows a complicated time-varying pattern (e.g., the time-varying variation) that cannot be captured by a simple parametric curve. In fact, Kwiatkowski et al. (1992)'s stationarity test for the absolute value of inflation rate has a p-value 0.09 and thus the stationarity may be questioned.

We apply our TV-ARX model to study how stock returns are affected by the past returns and inflation rate. Our model differs from the existing ones in that we allow such relationship to vary over time in a fully nonparametric way. For stock returns, we consider

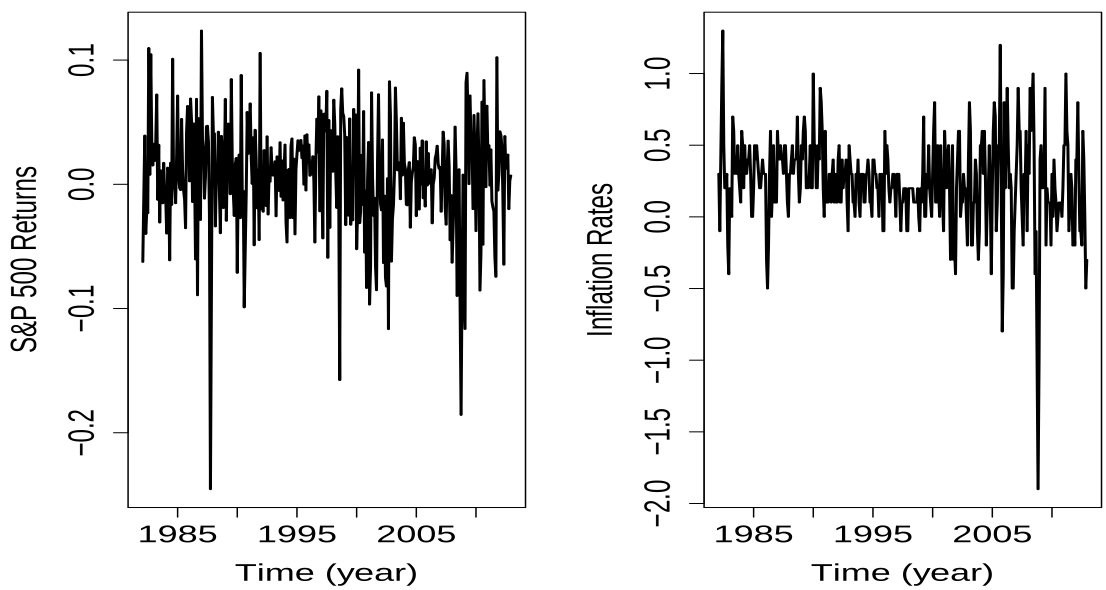

Figure 5.1: Time series plots of S&P 500 index return (left plot) and CPI-based inflation rate (right plot) during January, 1982–December, 2012.

S&P 500 index monthly returns $\log(S_i) - \log(S_{i-1})$, where S_i is the index at month i. The left plot of Figure 5.1 is a plot of the return series over the same time period 1982–2012. In (5.2) we let $X_i = \log(S_i) - \log(S_{i-1})$ be the return and $\{Z_i\}$ be the inflation rate.

For bandwidth h_n in the nonparametric mean regression (5.25), Ruppert et al. (1995)'s plug-in bandwidth is 0.08. Thus, by the discussion in Section 5.1.5(i), we use $b_n = 0.16$. Minimizing BIC(p, q) [cf. (5.26) with $\ell(k) = k\log(n)$] over $0 \le p, q \le 6$, we obtain the order $(p, q) = (1, 4)$. To examine the robustness of the selected order, we also consider two other bandwidths $b_n = 0.12, 0.20$. For $b_n = 0.12$, it selects the same order $(1, 4)$; for $b_n = 0.20$, it selects order $(0, 4)$ but BIC(0,4)$= -2308.9$ is almost identical to BIC(1,4)$= -2308.7$. Given the consistency of the selected order, it is reasonable to use $(p, q) = (1, 4)$.

The solid curves in Figure 5.2 are the estimated intercept function $\hat{\alpha}(\cdot)$ and coefficient functions $\hat{\phi}(\cdot), \hat{\beta}_1(\cdot), \ldots, \hat{\beta}_4(\cdot)$. The plots of $\hat{\beta}_1(\cdot), \ldots, \hat{\beta}_4(\cdot)$ show some interesting feature. Note that $\hat{\beta}_1(\cdot)$ is negative over the entire time period, which provides another empirical support for Fama's theory that stock returns are negatively correlated with inflation. However, the positiveness of $\hat{\beta}_2(\cdot)$ suggests that return X_i is positively correlated with the lagged inflation Z_{i-1}. Interestingly, $\hat{\beta}_3(\cdot)$ is negative for about 3/4 of the time period, whereas $\hat{\beta}_4(\cdot)$ is half-negative and half-positive. The plots also suggest clear time-varying effects: while $\hat{\beta}_1(\cdot)$ and $\hat{\beta}_3(\cdot)$ increase over time, $\hat{\beta}_4(\cdot)$ has a downward trend and $\hat{\beta}_2(\cdot)$ shows a downward and then flat trend.

In Figure 5.2 we also include the sieve-wild bootstrap SCB (see Section 5.2) for each of the 6 curves at 95% (dot dashed) and 90% (dashed) levels. Since the SCB for $\beta_1(\cdot), \ldots, \beta_4(\cdot)$ does not *entirely* cover the zero line (horizontal dotted), we have strong evidence to reject the four separate null hypotheses $\beta_1(\cdot) \equiv 0, \ldots, \beta_4(\cdot) \equiv 0$, and we conclude that the four coefficients are significantly different from zero. Although the significant regions where the zero line falls outside the SCB mostly happen near the boundary, they are not just some boundary issue but real. First, the functions are estimated on $[0.05, 0.95]$ to avoid the boundary issue. In fact, the significant regions would have been much larger if we had estimated the curves on a larger subinterval of $[0, 1]$. Second, we use local linear regression, which is well-known to reduce the boundary effect.

To test the joint significance of $\beta_1(\cdot), \ldots, \beta_4(\cdot)$ [cf. the null hypothesis in (5.32)], we use the test statistic T_n in (5.33). The sieve-wild bootstrap p-value is 0.047, which presents strong evidence to reject the joint null hypothesis. Therefore, the inflation rates $(Z_i, Z_{i-1}, Z_{i-2}, Z_{i-3})$ are jointly significant in predicting the return X_i, although in an overall negative correlation way. This agrees with the conclusion from the above SCB analysis.

Finally, since the SCB for $\alpha(\cdot)$ and $\phi(\cdot)$ contain the zero line, we fail to reject the null hypotheses $\alpha(\cdot) \equiv 0$ and $\phi(\cdot) \equiv 0$. Thus, unlike the inflation rate, historical returns carry no statistically significant information about future returns.

In summary, our TV-ARX modeling supports previous studies that stock returns are significantly negatively correlated with inflation, but our study also shows that the return-inflation relationship varies with time in a nonparametric way.

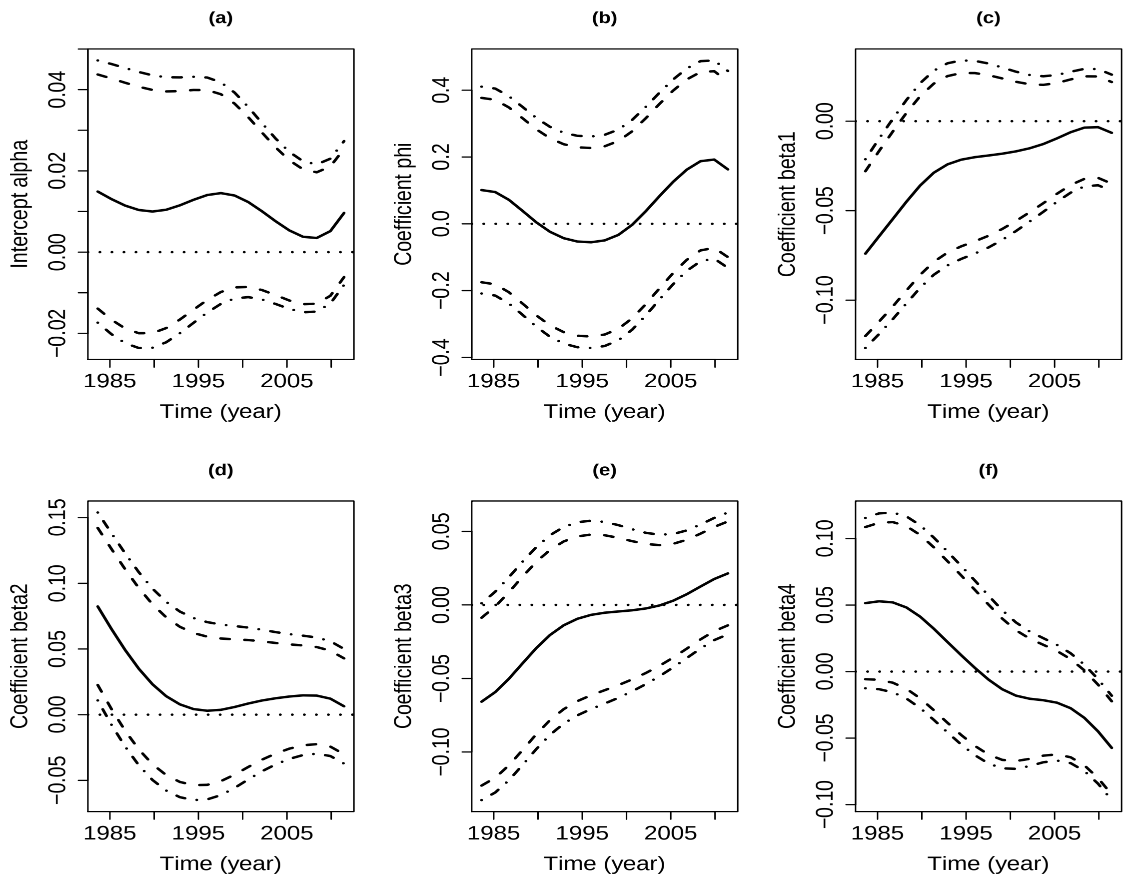

Figure 5.2: Plots of the estimated curves along with their 95% SCB (dot dashed) and 90% SCB (dashed). (a): intercept $\hat{\alpha}(\cdot)$; (b): coefficient $\hat{\phi}(\cdot)$; (c)–(f): coefficients $\hat{\beta}_1(\cdot), \ldots, \hat{\beta}_4(\cdot)$. The estimation and SCB construction are on $[0.05, 0.95]$ to avoid boundary issue. For better interpretation, we linearly transform the domain $[0, 1]$ of the functional curves to the calendar period January, 1982–December, 2012, and the subinterval $[0.05, 0.95]$ is then transformed accordingly.

5.4 Proofs

For a random variable Z, denote by $\|Z\|_q = [\mathbb{E}(|Z|^q)]^{1/q}$ the L_q norm. Recall $K_i(t) = K\{(i/n - t)/b_n\}$ in (5.14). Lemmas 5–6 follow from simple calculus and we omit the proofs.

Lemma 5. *Suppose Assumption 6 holds. Then for any $t \in [\epsilon, 1-\epsilon]$ with $\epsilon > 0$,*

$$\sum_{i=1}^{n} \left(\frac{i/n - t}{b_n} \right)^r K_i(t) = nb_n \int_{\mathbb{R}} u^r K(u) du + O(1), \quad r = 0, 1, \ldots.$$

Lemma 6. *Let $\{D_i\}_{i\in\mathbb{Z}}$ be a stationary process. Define $I_n = \sum_{i=1}^n D_i K_i(t)$.*

(i) If $\mathbb{E}(|D_0|) < \infty$, then $I_n = O_p(nb_n)$.

(ii) If $\mathrm{cov}(D_0, D_k) = O(\lambda^k)$ for some $\lambda \in (0,1)$, then $\mathrm{var}(I_n) = O(nb_n)$.

Lemma 7. *Suppose $\{a_j\}_{j\geq 0}$ and $\{b_j\}_{j\geq 0}$ satisfy $|a_j| + |b_j| = O(\tau^j)$ for some $\tau \in (0,1)$.*

(i) Let $\zeta_i = \sum_{j=0}^\infty a_j \varepsilon_{i-j}$ for some iid $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ with $\mathbb{E}(\varepsilon_i^4) < \infty$. Then $\mathrm{cov}(\zeta_0, \zeta_k) = O(\tau^k)$, and for any given $r \in \mathbb{Z}$, $\mathrm{cov}(\zeta_0\zeta_r, \zeta_k\zeta_{k+r}) = O(\tau^k)$.

(ii) Let $\xi_i = \sum_{j=0}^\infty b_j \eta_{i-j}$ for some iid $\{\eta_i\}_{i\in\mathbb{Z}}$, independent of $\{\varepsilon_i\}_{\in\mathbb{Z}}$, with $\mathbb{E}(\eta_i^4) < \infty$. Then, for any given $r \in \mathbb{Z}$, $\mathrm{cov}(\zeta_0\xi_r, \zeta_k\xi_{k+r}) = O(\tau^k)$.

Proof. (i) From $a_j = O(\tau^j)$, we have $\mathrm{cov}(\zeta_0, \zeta_k) = \mathbb{E}(\varepsilon_0^2)\sum_{j=0}^\infty a_j a_{j+k} = O(\tau^k)$. Let $\{\varepsilon_i'\}_{i\in\mathbb{Z}}$ be an iid copy of $\{\varepsilon_i\}_{i\in\mathbb{Z}}$. Without loss of generality, assume $r \geq 0$. For $k \geq r+1$, define

$$\zeta_k^* = \sum_{j=0}^{k-r-1} a_j \varepsilon_{k-j} + \sum_{j=k-r}^{\infty} a_j \varepsilon_{k-j}' \quad \text{and} \quad \zeta_{k+r}^* = \sum_{j=0}^{k-1} a_j \varepsilon_{k+r-j} + \sum_{j=k}^{\infty} a_j \varepsilon_{k+r-j}'.$$

By replacing $\{\varepsilon_i\}_{i\leq r}$ with the iid copy $\{\varepsilon_i'\}_{i\leq r}$, $(\zeta_k^*, \zeta_{k+r}^*)$ is independent of $\{\varepsilon_i\}_{i\leq r}$ and thus it is also independent of (ζ_0, ζ_r). Therefore, $\mathrm{cov}(\zeta_0\zeta_r, \zeta_k^*\zeta_{k+r}^*) = 0$, and

$$\begin{aligned} |\mathrm{cov}(\zeta_0\zeta_r, \zeta_k\zeta_{k+r})| &= |\mathrm{cov}(\zeta_0\zeta_r, \zeta_k\zeta_{k+r} - \zeta_k^*\zeta_{k+r}^*)| \\ &\leq \|\zeta_0\zeta_r\|_2 \times \|\zeta_k\zeta_{k+r} - \zeta_k^*\zeta_{k+r}^*)\|_2 \\ &\leq \|\zeta_0\zeta_r\|_2[\|(\zeta_k - \zeta_k^*)\zeta_{k+r}\|_2 + \|\zeta_k^*(\zeta_{k+r} - \zeta_{k+r}^*)\|_2] \\ &\leq \|\zeta_0\|_4 \cdot \|\zeta_r\|_4[\|\zeta_k - \zeta_k^*\|_4 \cdot \|\zeta_{k+r}\|_4 + \|\zeta_k^*\|_4 \cdot \|\zeta_{k+r} - \zeta_{k+r}^*\|_4] \end{aligned} \tag{5.43}$$

Clearly, ζ_k^* has the same distribution as ζ_k. Since $\mathbb{E}(\varepsilon_i^4) < \infty$, $\|\zeta_0\|_4 < \infty$. By construction,

$$\|\zeta_k - \zeta_k^*\|_4 = \left\| \sum_{j=k-r}^{\infty} a_j(\varepsilon_j - \varepsilon_j') \right\|_4 \leq \sum_{j=k-r}^{\infty} |a_j| \|\varepsilon_j - \varepsilon_j'\|_4 \leq 2\|\varepsilon_0\|_4 \sum_{j=k-r}^{\infty} |a_j| = O(\tau^k),$$

for each fixed r. Thus, by (5.43), we have $\mathrm{cov}(\zeta_0\zeta_r, \zeta_k\zeta_{k+r}) = O(\tau^k)$ for large enough k.

(ii) By the same argument in (i), we can prove the result through introducing (ζ_k^*, ξ_{k+r}^*) by replacing $\{\varepsilon_i\}_{i\leq 0}$ and $\{\eta_i\}_{i\leq r}$ in (ζ_k, ξ_{k+r}) with iid copies $\{\varepsilon_i'\}_{i\leq 0}$ and $\{\eta_i'\}_{i\leq r}$. ♢

Lemma 8. *Recall M_0, M_1 and M_2 in (5.16) and $U_i(t)$ in (5.19). Under Assumptions 3–6,*

$$\frac{M_0}{nb_n} \xrightarrow{p} \mathbb{E}[U_0(t)U_0^T(t)], \tag{5.44}$$

$$\frac{M_1}{nb_n^3} \xrightarrow{p} \mathbb{E}[U_0(t)U_0'(t)^T + U_0'(t)U_0(t)^T] \int_{\mathbb{R}} u^2 K(u)du, \tag{5.45}$$

$$\frac{M_2}{nb_n^3} \xrightarrow{p} \mathbb{E}[U_0(t)U_0(t)^T] \int_{\mathbb{R}} u^2 K(u)du. \tag{5.46}$$

Proof. [**Proof of (5.44)**]: Using $U_i = U_i(t) + o_p(1)$ in (5.20) and Lemma 6(i), we have

$$\frac{M_0}{nb_n} = \tilde{M}_0 + o_p(1), \quad \text{where} \quad \tilde{M}_0 = \frac{1}{nb_n}\sum_{i=1}^n U_i(t)U_i(t)^T K_i(t).$$

It suffices to show $\tilde{M}_0 \xrightarrow{p} \mathbb{E}[U_0(t)U_0^T(t)]$. By the stationarity of $\{U_i(t)\}_{i\in\mathbb{Z}}$ and Lemma 5 (recall $\int_{\mathbb{R}} K(u)du = 1$),

$$\mathbb{E}(\tilde{M}_0) = \mathbb{E}[U_0(t)U_0(t)^T]\frac{1}{nb_n}\sum_{i=1}^n K_i(t) \to \mathbb{E}[U_0(t)U_0(t)^T].$$

It remains to show that the variance of each element of the matrix $\tilde{M}_0$ goes to zero. Note that each element of $\tilde{M}_0$ can be written as $(nb_n)^{-1}\sum_{i=1}^n D_i(t)K_i(t)$, where $D_i(t)$ is one of the five forms: $X_{i-j}(t), X_{i-j}(t)X_{i-j'}(t), Z_{i-r}(t), Z_{i-r}(t)Z_{i-r'}(t), X_{i-j}(t)Z_{i-r}(t)$ for some $1 \le j, j' \le p$ and $0 \le r, r' \le q-1$. Substituting (5.18) into (5.9) and using the fact that ψ_j in (5.9) and $\sup_t |\alpha_j(t)|$ in (5.18) decay exponentially, we can show that $X_i(t)$ can be expressed as a linear combination of $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ and $\{\eta_i\}_{i\in\mathbb{Z}}$ with exponentially decaying coefficients. Thus, by Lemma 7, $\mathrm{cov}(D_0(t), D_k(t))$ decays exponentially fast in k. An application of Lemma 6(ii) then gives the bound $O[(nb_n)^{-1}] = o(1)$ on the variance of $\tilde{M}_0$.

[**Proof of (5.45)**]: By $U_i = U_i(t) + (i/n - t)U_i'(t) + O_p(b_n^2)$ in (5.20), we can write $(nb_n^3)^{-1}M_1 = J_1 + J_2 + J_3 + J_4$, where

$$J_1 = \frac{1}{nb_n^3}\sum_{i=1}^n U_i(t)U_i(t)^T K_i(t)(i/n - t),$$

$$J_2 = \frac{1}{nb_n^3}\sum_{i=1}^n U_i'(t)U_i'(t)^T K_i(t)(i/n - t)^3,$$

$$J_3 = \frac{1}{nb_n^3}\sum_{i=1}^{n}[U_i(t)U_i'(t)^T + U_i'(t)U_i(t)^T]K_i(t)(i/n-t)^2,$$
$$J_4 = \frac{O_p(1)}{nb_n}\sum_{i=1}^{n}[|U_i(t)| + |U_i'(t)|]K_i(t)|i/n-t|.$$

By the symmetry of $K(\cdot)$, $\int_{\mathbb{R}} uK(u)du = 0$. Thus, by Lemma 5,

$$\mathbb{E}(J_1) = \mathbb{E}[U_0(t)U_0(t)^T]\frac{1}{nb_n^2}\sum_{i=1}^{n}\left(\frac{i/n-t}{b_n}\right)K_i(t) = O\Big(\frac{1}{nb_n^2}\Big) \to 0.$$

By the same argument in the proof of (5.44), we can show that the variance of J_1 is bounded by $O[(nb_n^3)^{-1}] = o(1)$. Thus, $J_1 = o_p(1)$. Similarly, $J_2 = o_p(1)$. For J_3, by Lemma 5,

$$\begin{aligned}\mathbb{E}(J_3) &= \mathbb{E}[U_0(t)U_0'(t)^T + U_0'(t)U_0(t)^T]\frac{1}{nb_n}\sum_{i=1}^{n}\left(\frac{i/n-t}{b_n}\right)^2 K_i(t)\\ &\to \mathbb{E}[U_0(t)U_0'(t)^T + U_0'(t)U_0(t)^T]\int_{\mathbb{R}} u^2K(u)du.\end{aligned}$$

By the same argument in the proof of (5.44), the variance of J_3 is bounded by $O[(nb_n)^{-1}] = o(1)$. Finally, by Lemma 6(i), $J_4 = O(n^{-1})\sum_{i=1}^{n}[|U_i(t)| + |U_i'(t)|]K_i(t) = O_p(b_n) = o_p(1)$.

[**Proof of (5.46)**]: It follows from the same argument of (5.45). $\diamondsuit$

Lemma 9. *Suppose Assumptions 3–6 holds. Then, for any given $t \in (0,1)$, the CLT holds:*

$$C_n(t) := \frac{1}{\sqrt{nb_n}}\sum_{i=1}^{n}U_iK_i(t)\sigma(i/n)\varepsilon_i \Rightarrow N\bigg(0,\ \sigma^2(t)\mathbb{E}[U_0(t)U_0^T(t)]\int_{\mathbb{R}} K^2(u)du\bigg). \quad (5.47)$$

Furthermore, $C_n(t)$ and $C_n(t')$ are asymptotically independent for $t \neq t'$.

Proof. Write $d_i = U_iK_i(t)\sigma(i/n)\varepsilon_i$. For each i, let $\mathcal{F}_i$ be the σ-algebra generated by $\{(\varepsilon_j, \eta_{j+1})\}_{j\le i}$. By (5.2) and Assumption 5(i), U_i is a function of $\mathcal{F}_{i-1}$ and ε_i is independent of $\mathcal{F}_{i-1}$. Thus, $\mathbb{E}(d_i|\mathcal{F}_{i-1}) = U_iK_i(t)\sigma(i/n)\mathbb{E}(\varepsilon_i) = 0$ and $\{d_i\}_{i\in\mathbb{Z}}$ are martingale differences with respect to the filtration $\{\mathcal{F}_i\}_{i\in\mathbb{Z}}$. We shall apply the martingale CLT. For the conditional variance, by $\mathbb{E}(\varepsilon_i^2) = 1$ and the independence between ε_i and $\mathcal{F}_{i-1}$,

$$\frac{1}{nb_n}\sum_{i=1}^n \mathbb{E}(d_i d_i^T|\mathcal{F}_{i-1}) = \frac{1}{nb_n}\sum_{i=1}^n U_i U_i^T \sigma^2(i/n)K_i^2(t) \xrightarrow{p} \sigma^2(t)\mathbb{E}[U_0(t)U_0^T(t)]\int_{\mathbb{R}} K^2(u)du,$$

where the convergence follows from the same argument in the proof of (5.44) in Lemma 8.

To verify the Lindeberg condition, consider the $X_{i-j}, j = 1, \ldots, p$, component of U_i (the component $Z_{i+1-r}, r = 1, \ldots, q$, follows similarly). For any constant $c > 0$, define

$$L_n = \frac{1}{nb_n}\sum_{i=1}^n \mathbb{E}\Big\{[X_{i-j}K_i(t)\sigma(i/n)\varepsilon_i]^2 \mathbf{1}_{|X_{i-j}K_i(t)\sigma(i/n)\varepsilon_i| \geq c\sqrt{nb_n}}\Big|\mathcal{F}_{i-1}\Big\}.$$

By the boundedness of $K(\cdot)$ and $\sigma(\cdot)$, $|X_{i-j}K_i(t)\sigma(i/n)\varepsilon_i| \geq c\sqrt{nb_n}$ implies that $|\varepsilon_i| \geq c_1\sqrt{nb_n}/|X_{i-j}|$ for some constant $c_1 > 0$. Thus, by the independence between ε_i and $\mathcal{F}_{i-1}$,

$$\begin{aligned} L_n &= \frac{O(1)}{nb_n}\sum_{i=1}^n K_i^2(t)X_{i-j}^2 \mathbb{E}[\varepsilon_i^2 \mathbf{1}_{|\varepsilon_i| \geq c_1\sqrt{nb_n}/|X_{i-j}|}|\mathcal{F}_{i-1}] \\ &= \frac{O(1)}{nb_n}\sum_{i=1}^n K_i^2(t)X_{i-j}^2 G(\omega_i), \quad \text{where } G(z) = \mathbb{E}(\varepsilon_i^2\mathbf{1}_{|\varepsilon_i|>z}),\ \omega_i = \frac{c_1\sqrt{nb_n}}{|X_{i-j}|}. \end{aligned} \tag{5.48}$$

Since $\mathbb{E}(\varepsilon_i^2) < \infty$, we have $G(z) \to 0$ as $z \to \infty$. By (5.20), $X_{i-j} = X_{i-j}(t) + o_p(1)$. From $\mathbb{E}(\varepsilon_i^4) < \infty$ and $\mathbb{E}(\eta_i^4) < \infty$, we have $\max_{1\leq i\leq n}|X_{i-j}(t)| = O_p(n^{1/4})$. Thus, under condition $nb_n^3 \to \infty$, $\min_{1\leq i\leq n}\omega_i \to \infty$. Therefore, $G(\omega_i) = o_p(1)$ uniformly in i. Then, by (5.48),

$$L_n = \frac{o_p(1)}{nb_n}\sum_{i=1}^n K_i^2(t)X_{i-j}^2 = \frac{o_p(1)}{nb_n}\sum_{i=1}^n K_i^2(t)[X_{i-j}(t) + o_p(1)]^2 = o_p(1).$$

This proves the Lindeberg condition. By the martingale CLT, (5.47) holds.

By the bounded support of $K(\cdot)$, as $b_n \to 0$, $K_i(t)K_i(t') = 0$ for $t \neq t'$. The asymptotic independence between $C_n(t)$ and $C_n(t')$ can be shown using the Cramér-Wold device. $\diamondsuit$

Proof of Theorem 23. By (5.15), we have

$$\hat{\theta}(t) - \theta(t) = \left(\frac{M_0 - M_1M_2^{-1}M_1}{nb_n}\right)^{-1}\left\{\frac{N_0 - M_0\theta(t)}{nb_n} - \frac{M_1M_2^{-1}[N_1 - M_1\theta(t)]}{nb_n}\right\}. \tag{5.49}$$

By (5.44), $(nb_n)^{-1}M_0 \xrightarrow{p} \mathbb{E}[U_0(t)U_0^T(t)]$. Therefore, by Lemma 8,

$$\frac{M_0 - M_1 M_2^{-1} M_1}{nb_n} = \frac{M_0}{nb_n} - b_n^2 \frac{M_1}{nb_n^3}\left(\frac{M_2}{nb_n^3}\right)^{-1}\frac{M_1}{nb_n^3} \xrightarrow{p} \mathbb{E}[U_0(t)U_0^T(t)]. \tag{5.50}$$

Using $X_i - U_i^T\theta(t) = U_i^T[\theta(i/n) - \theta(t)] + \sigma(i/n)\varepsilon_i$, we have the bias-stochastic decomposition:

$$N_0 - M_0\theta(t) = \sum_{i=1}^n U_i U_i^T K_i(t)[\theta(i/n) - \theta(t)] + \sum_{i=1}^n U_i K_i(t)\sigma(i/n)\varepsilon_i. \tag{5.51}$$

By Taylor's expansion $\theta(i/n) - \theta(t) = (i/n - t)\theta'(t) + (i/n - t)^2\theta''(t)/2 + o(b_n^2)$,

$$\sum_{i=1}^n U_i U_i^T K_i(t)[\theta(i/n) - \theta(t)] = M_1\theta'(t) + M_2\theta''(t)/2 + o_p(nb_n^3). \tag{5.52}$$

Here the error term $o_p(nb_n^3)$ follows from Lemma 6(i). Similar to (5.51), we can rewrite

$$N_1 - M_1\theta(t) = \sum_{i=1}^n (i/n - t)U_i U_i^T K_i(t)[\theta(i/n) - \theta(t)] + \sum_{i=1}^n (i/n - t)U_i K_i(t)\sigma(i/n)\varepsilon_i.$$

For the two terms on the right hand side, by the same argument in (5.52), the first term equals $M_2\theta'(t) + O_p(nb_n^4)$; by the same argument in Lemma 9, the second term is of the order $O_p(\sqrt{nb_n^3})$. Note that both $O_p(nb_n^4)$ and $O_p(\sqrt{nb_n^3})$ are of the order $o_p(nb_n^3)$. Therefore,

$$M_1 M_2^{-1}[N_1 - M_1\theta(t)] = M_1 M_2^{-1}[M_2\theta'(t) + o_p(nb_n^3)] = M_1\theta'(t) + o_p(nb_n^3). \tag{5.53}$$

Here the last equality follows from $M_1 M_2^{-1} = O_p(1)$ [(5.45)–(5.46) in Lemma 8]. By (5.51)–(5.53),

$$\frac{N_0 - M_0\theta(t)}{nb_n} - \frac{M_1 M_2^{-1}[N_1 - M_1\theta(t)]}{nb_n} = \frac{M_2\theta''(t)}{2nb_n} + \frac{1}{nb_n}\sum_{i=1}^n U_i K_i(t)\sigma(i/n)\varepsilon_i + o_p(b_n^2),$$

which, combined with (5.49)–(5.50), (5.46), and Lemma 9, then completes the proof. ♢

Bibliography

[1] Zhao, Z. and Li, X., Inference for modulated stationary processes, *Bernoulli*, 19, 2013 pp. 205–227.

[2] Li, X. and Zhao, Z., Testing for changes in autocovariances of non-parametric time series models, *Journal of Statistical Planning and Inference*, 143, 2013 pp. 237–250.

[3] Efron, B., Bootstrap methods: another look at the jackknife,*The annals of Statistics*, Vol. 7, No. 1, 1979 pp. 1–26.

[4] Wu, C.F.J., Jackknife, bootstrap and other resampling methods in regression analysis,*The Annals of Statistics*, Vol. 14, No. 4,1986 pp. 1261–1295.

[5] Liu, R.Y., Bootstrap procedures under some non-iid models,*The Annals of Statistics*, Vol. 16, No. 4, 1988 pp. 1696–1708.

[6] Bai, J. and Perron, P., Estimating and testing linear models with multiple structural changes,*Econometrica*, Vol. 66, No. 1, 1998 pp. 47–78.

[7] Altman, NS, Kernel smoothing of data with correlated errors,*Journal of the American Statistical Association*, Vol. 85, No. 411, 1990 pp. 749–759.

[8] Horvath, L., The maximum likelihood method for testing changes in the parameters of normal observations,*The Annals of Statistics*, Vol. 21, No. 2, 1993 pp. 671–680.

[9] Dahlhaus, R., Fitting time series models to nonstationary processes,*The annals of Statistics* , Vol. 25, No. 1, 1997 pp. 1–37.

[10] , Fan, J. and Yao, Q., Nonlinear time series,*The annals of Statistics* , Springer-Verlag, New York, 2003.

[11] Aue, A. and Horváth, L. and Kokoszka, P. and Steinebach, J., Monitoring shifts in mean: Asymptotic normality of stopping times, Springer, 2008 pp. 515–530.

[12] Dahlhaus, R. and Polonik, W., Empirical spectral processes for locally stationary time series, *Bernoulli* , Vol. 15, No. 1, 2009 pp. 1–39.

[13] Abramowitz, M. and Stegun, I.A., *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, 1964.

[14] Altman, N.S., Kernel smoothing of data with correlated errors, *J. Amer. Statist. Assoc.*, 1990 pp. 749–759.

[15] Andrews, D.W.K., Heteroskedasticity and autocorrelation consistent covariance matrix estimation, *Econometrica* 1991 pp. 817–858.

[16] Aue, A., Hörmann, S., Horváth, L. and Reimherr, M., Break detection in the covariance structure of multivariate time series models, *Ann. Statist.*, 2009 pp. 4046–4087.

[17] Aue, A., Horváth, L., Hušková, M. and Kokoszka, P., Testing for changes in polynomial regression., *Bernoulli*, 2008a, pp. 637–660.

[18] Aue, A., Horváth, L., Kokoszka, P. and Steinebach, J., Monitoring shifts in mean: asymptotic normality of stopping times, *Test*, 2008b pp. 515–530.

[19] Bai, J. and Perron, P., Estimating and testing linear models with multiple structural changes, *Econometrica*, 1998 pp. 47–78.

[20] Berkes, I., Gombay, E. and Horváth, L., Testing for changes in the covariance structure of linear processes., *J. Statist. Plann. Inference*, 2009 pp. 2044–2063.

[21] Billingsley, P., *Convergence of Probability Measures.* New York: Wiley 1968.

[22] Brockwell, P.J. and Davis, R., *Time Series: Theory and Methods.* (2nd ed.). Springer-Verlag, New York, 1991.

[23] Davis, R.A., Huang, D.W. and Yao, Y.C., Testing for a change in the parameter values and order of an autoregressive model, *Ann. Statist.*, 1995 pp. 282-304.

[24] Galeano, P. and Peña, D., Covariance changes detection in multivariate time series, *J. Statist. Plann. Inference* 2007 pp. 194–211.

[25] Gombay, E., Change detection in autoregressive time series, *J. Multivariate Anal.* 2008 pp. 451–464.

[26] Horváth, L., The maximum likelihood method for testing changes in the parameters of normal observations. *Ann. Statist.*, 1993 pp. 671–680.

[27] Inclán, C. and Tiao, G.C., Use of cumulative sums of squares for retrospective detection of changes of variance, *J. Amer. Statist. Assoc.*, 1994 pp. 913–923.

[28] Kokoszka, P. and Leipus, R., Change-point estimation in ARCH models, *Bernoulli*, 2000 pp. 513–539.

[29] Kufner, A., John, O. and Fučik, S., *Function Spaces.* Noordhoff, Leyden, 1997.

[30] Lai, K.S., The puzzling unit root in the real interest rate and its inconsistency with intertemporal consumption behavior, *J. Int. Money Financ.*, 2008 pp. 140–155.

[31] Lavielle, M., Detection of multiple changes in a sequence of dependent variables, *Stoch. Proc. Appl.*, 1999 pp. 79–102.

[32] Lee, S., Ha, J. and Na, O., The CUSUM test for parameter change in time series models, *Scand. J. Statist.*, 2003 pp. 781–796.

[33] Ling, S., Testing for change points in time series models and limiting theorems for NED sequences, *Ann. Statist.*, 2007 pp. 1213–1237.

[34] Mei, Y., Sequential change-point detection when unknown parameters are present in the pre-change distribution, *Ann. Statist.*, 2006 pp. 92–122.

[35] Picard, D., Testing and estimating change-points in time series, *Adv. Appl. Probab.*, 1985 pp. 841–867.

[36] Politis, D.N., Adaptive bandwidth choice, *J. Nonparametric Statist.*, 2003 pp. 517–533.

[37] Priestley, M.B. and Chao, M.T., Non-parametric function fitting, *J. R. Statist. Soc. B*, 1972 pp. 385–392.

[38] Shao, X. and Wu, W.B., Asymptotic spectral theory for nonlinear time series, *Ann. Statist.*,2007 pp. 1773–1801.

[39] Shao, X. and Zhang, X., Testing for change points in time series, *J. Amer. Statist. Assoc.*, 2010 pp. 1228–1240.

[40] Wied, D., Arnold, M., Bissantz, N. and Ziggel, D., A new fluctuation test for constant variances with applications to finance, *Metrika*, 2012 Forthcoming.

[41] Wied, D., Krämer, W. and Dehling, H., Testing for a change in correlation at an unknown point in time using an extended functional delta method, *Econometric Theory*, 2012 pp. 570–589.

[42] Wu, W.B., Strong invariance principles for dependent random variables, *Ann. Probab.*, 2007 pp. 2294–2320.

[43] Wu, W.B. and Zhao, Z., Inference of trends in time series, *J. R. Statist. Soc. B*, 2007 pp. 391–410.

[44] Zhao, Z. and Li, X., Inference for modulated stationary processes, To appear, *Bernoulli*, 2012.

[45] Zhao, Z., Wei, Y. and Lin, D., Asymptotics of nonparametric L-1 regression models with dependent data. Manuscript, 2012.

Vita

Xiaoye Li

EDUCATION

Pennsylvania State University, University Park, PA

Ph.D. candidate, Statistics, 2008-2013

Dalian University of Technology, Dalian, China

M.S., Mathematics, 2006-2008

Dalian University of Technology, Dalian, China

B.S., Mathematics, 2002-2006

RESEARCH

- Statistical inference for non-stationary time series
- Change point test of autocovariance for non-stationary time series models
- Time-varying Exogenous Autoregressive Models

PUBLICATIONS

- Zhao, Z. and Li, X., Inference for modulated stationary processes, *Bernoulli*, 19, 2013.
- Li, X. and Zhao, Z., Testing for changes in autocovariances of non-parametric time series models, *Journal of Statistical Planning and Inference*, 143, 2013.

PROFESSIONAL ACTIVITIES

- Presenter, Joint Statistical Meeting, Miami, FL, Aug. 2011
- Presenter, Industrial Mathematical and Statistical Modeling Workshop, Raleigh, NC, Jul. 2010

TEACHING EXPERIENCE

- Led the Statistics Lab regarding the Applied Regression Analysis, Elementary Statistics, and Statistical Concepts and Reasoning courses
- Graded assignments and exams for Applied Time Series Analysis, Probability and Intermediate Applied Statistics courses