Trending Time Series Models with Endogeneity

A thesis submitted for the degree of

Doctor of Philosophy

by

Li Chen

B.Sci., B.Econ. (Xiamen University, China)

Department of Econometrics and Business Statistics

Monash University

April 22, 2017

Contents

List of Tables

List of Figures

Declaration

I hereby declare that this thesis contains no material which has been accepted for the award of any other degree or diploma in any university or equivalent institution, and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

[redacted]

Li Chen

Acknowledgements

Four years ago, I was very thrilled and honored to commence my PhD candidature and become a member at the Department of Econometrics and Business Statistics, Monash University. Time flies, and now, standing at the end of this rewarding journey, I am deeply indebted to many people, who have kindly helped me to learn, think, and grow up in the past four years.

First of all, I would like to express my sincere gratitude to Jiti Gao, Farshid Vahid, and David Harris for their excellent supervision. They are my role models who have broad knowledge, innovative thinking, and professional experience in Economics and Econometrics. It is impossible for me to complete this PhD thesis without their guidance and help.

The staff members at the Department of Econometrics and Business Statistics are the most friendly and helpful people I have ever met. I especially would like to thank Mervyn Silvapulle, Heather Anderson, and Catherine Forbes for teaching me the first year courses, as well as Gael Martin for her efforts as an excellent PhD Director. Meanwhile, I would like to thank Don Poskitt, Param Silvapulle, and David Frazier for their comments and suggestions to my PhD thesis. The fellow PhD students at EBS are a group of lovely and hard-working people, and life is much happier with their company. I thank Tingting Cheng, Huanjun Zhu, Fei Liu, Yan Meng, Weilun Zhou, Yiru Wang, Ming Kong, Nithi Sopitpongstorn, Ban Kheng Tan, Chuan Wang and Kai Xia for their help and wish them all the best with their research and future career.

As one of the world-leading institutions for Econometrics research, the department invites many visiting scholars every year. The department also encourages and supports PhD students to attend various academic activities such as conference meetings. As a young novice, I cherish every opportunity to meet with and learn from these leading economists and econometricians. I am grateful to John Chao, Yongmiao Hong, Cheng Hsiao, Robert Kohn, Oliver Linton, Peter C.B. Phillips, Peter Robinson, Aman Ullah, and Qiwei Yao for their constructive and helpful suggestions to me. Meanwhile, I would like to thank the rising stars in Econometrics including Chaohua Dong, Bin Peng, Yundong Tu, and Yanrong Yang. I received many constructive suggestions from them that enabled me to improve this thesis. I also want to sincerely thank Ying Fang and Haiqiang Chen for guiding me to the world of Econometrics.

As a full-time PhD student, I am indebted to the Monash Business School for providing me with the scholarships that covered my tuition fees and living expenses during my candidature. As the only recipient of the Donald Cochrane Postgraduate Research Scholarship in 2013, I would like to thank the Cochrane family for their generosity in establishing this prestigious award in honor of Professor Donald Cochrane.

Lastly, I would very much like to thank my parents for their constant love and support. I have no better words to express my love to them, and I decide to make this thesis as a special gift to them for their 30^{th} wedding anniversary.

Abstract

This thesis explores the problem of endogeneity in the trending time series regression model. The nonstationary regressor in the model is assumed to follow a nonlinear trend-stationary process instead of a unit root process. We introduce a magnitude parameter that characterizes the strength of the trend in the nonstationary time series. The time trend is termed a *weak trend* when such parameter is exactly one. While it is called a *strong trend* when the parameter is greater than one.

The weak trend does not dominate the stationary error term in the regressor's data generating process so that the simple ordinary least squares estimator is biased and inconsistent due to the problem of endogeneity. To fix this issue, we employ a control function to decompose the endogenous correlation between the corresponding error terms. To avoid potential model misspecification, we let the control function be non-parametric. Replacing the regression error by such control function expression yields a semi-parametric partially linear model. We show that the conventional estimator remains valid (unbiased and $\sqrt{n}$-consistent), although we fail to satisfy the usual identifiability condition for the semi-parametric partially linear model.

On the contrary, the strong trend dominates the stationary error term that makes the OLS estimator consistent when the sample size tends to infinity. However, the statistical inference on the regression coefficient is substantially affected by endogeneity that the inference for the coefficients can be quite misleading unless we deal with the endogeneity issue appropriately. We propose a bias-corrected estimator that adjusts the endogeneity bias in the OLS estimator. The asymptotic results show that the new estimator is unbiased and consistent. Meanwhile, the simulation results indicate that the bias-correction method greatly improves the estimation accuracy as well as the inference performance for the coefficients.

The regression models and the estimation methods are applied to investigate the relationship between the logarithm of aggregate personal consumption and the logarithm of aggregate personal disposable income as an illustrative example. The result reveals significant endogenous correlations between the variables, and moreover, such correlation is found to be nonlinear rather than linear.

To summarize, the main objective of this thesis is to explore the effects of endogeneity on the trending regression models with nonlinear trend-stationary processes and propose effective methods to correct the endogeneity bias. In the process of achieving these goals, we reveal some interesting facts, such as

(1) it is difficult to discriminate between a unit root process and a nonlinear trend-stationary process;

(2) the strength of the trend matters a lot in the trending regression model;

(3) the behaviors and properties of certain conventional estimators are likely to be different for the models with nonstationary time series compared to those with stationary time series.

Chapter 1

Introduction

"No one understands trends, but everyone sees them in the data."

Laws and Limits of Econometrics

— Peter C. B. Phillips (2003)

1.1 Background and overview

Time series regression models are widely used in the economic and financial analysis. When estimating these models in practice, economists frequently encounter two major challenges. The first problem is the nonstationary trending feature of the time series data. As suggested in Andrews and McDermott (1995) and Krugman (1995), most of the empirical data, especially the macroeconomic aggregates, exhibit linear or nonlinear time trends. Such nonstationary characteristic violates the standard assumption that the time series should be stationary over time. Endogeneity is the second problem that we need to deal with in practice. The correlation between the explanatory variables and the regression error may lead to biased and inconsistent estimates of the coefficients. Various reasons may cause the problem of endogeneity, such as simultaneity, measurement errors, omitted variables, selection bias, etc. A popular example is the linear regression model of income and consumption

$$c_t = \alpha + \beta y_t + e_t, \tag{1.1}$$

for $t = 1, 2, ..., n$, where c_t and y_t represent the (logarithms of) aggregate personal consumption expenditure and disposable income. A simple plot of the two time series shows that both c_t and y_t exhibit upward trends, thus they are nonstationary across time. Meanwhile, due to the simultaneous determination of income and consumption, y_t and e_t are believed to be endogenously correlated. Therefore, it induces the problem of endogeneity. Since the assumptions for the classical linear regression (CLR) models are not satisfied[1], we cannot use the ordinary least squares (OLS) method to estimate the marginal propensity to consume.

This thesis deals with both problems of nonstationarity and endogeneity. We consider a general linear trending regression model formulated as

$$y_t = \alpha + x_t'\beta + e_t, \tag{1.2}$$

$$x_t = g(t) + v_t, \tag{1.3}$$

for $t = 1, 2, ..., n$, where x_t is a $k \times 1$ vector of trending time series, $g(\cdot)$ is a $k \times 1$ vector of functions representing the deterministic time trends in x_t. The problem of endogeneity occurs when the stationary error terms e_t and v_t are correlated.

The true form of the trend function in the generating process of the trending time series is usually unknown. We let $g(\cdot)$ be a nonparametric function rather than a pre-specified parametric form. Therefore, we avoid potential model misspecification, which may induce inconsistency in the estimation of the coefficients. The nonparametric form of $g(\cdot)$ provides sufficient flexibility to capture the nonstationary and nonlinear characteristics in the time trends. In this thesis, we use nonparametric kernel methods to estimate the trend term.[2] Since the estimation method is data-driven, the nonparametric estimate of $g(\cdot)$ is adaptive to the changes in the levels and slopes

[1] Here, both assumptions of exogeneity and stationarity are not satisfied.

[2] Other nonparametric methods are also applicable, for example, the Sieve method.

of the trends. In fact, equation (1.3) represents a class of nonlinear trend-stationary time series. By definition, stationary is achieved after removing the nonlinear trends. Moreover, as we will discuss later, it can be regarded as an alternative data generating mechanism to the unit root process for the nonstationary time series data.

Equation (1.2) forms a linear trending regression model. Since we do not impose constraints on the dependent variable y_t, the interpretation of the coefficient vector β depends on the nature of y_t.

One scenario is that y_t also contains a time trend as x_t, then the coefficient vector β represents the co-trending relationship between x_t and y_t. Hence, the equation (1.2) is analogous to the co-integration model first studied in Engle and Granger (1987). We also take into account the endogeneity problem, and the model investigated in this thesis is similar to the one discussed in Phillips and Hansen (1990), in which the authors studied the instrumental variable regressions for $I(1)$ processes.

Another scenario is that y_t is stationary, but the elements of x_t have trends so that co-trending occurs between the elements. A popular example is the predictive regression model commonly seen in the finance literature

$$y_t = \alpha + x_{t-1}'\beta + e_t, \tag{1.4}$$

where y_t represents asset returns, x_{t-1} is a vector of predictors such as the dividend-price ratio, the book-to-market ratio in Stambaugh (1999) and Welch and Goyal (2008). The predictors are usually quite persistent, with first order autocorrelations close to 1. In the literature, they are usually modeled as integrated processes with root equal or near to unity; see Campbell and Yogo (2006), Cai and Wang (2014). As we will address in detail later, it is also reasonable to model the predictors as (1.3) because it is hard to distinguish between a unit root process and a nonlinear trend-stationary process. When the predictors are modeled as (1.3), the persistence shown in the sample autocorrelation coefficients may come from the low-frequency information in the data, i.e., the nonlinear time trends. To balance both sides of equation (1.4), the coefficient β should represent both the co-trending relationship and the predictability of x_t. In other words, the linear combination $x'_{t-1}\beta$ itself forms a stationary process, which has predictive power on y_t.

As the endogeneity bias is caused only by the correlation between the stationary error terms, the magnitude of the time trend matters a lot to the consistency and the convergence rate of the simple OLS estimator. If the stationary term v_t were dominated by the time trend $g(\cdot)$, the endogeneity bias vanishes when the sample size goes to infinity. Otherwise, we cannot obtain consistent estimations of the coefficients when the time trends are weak.

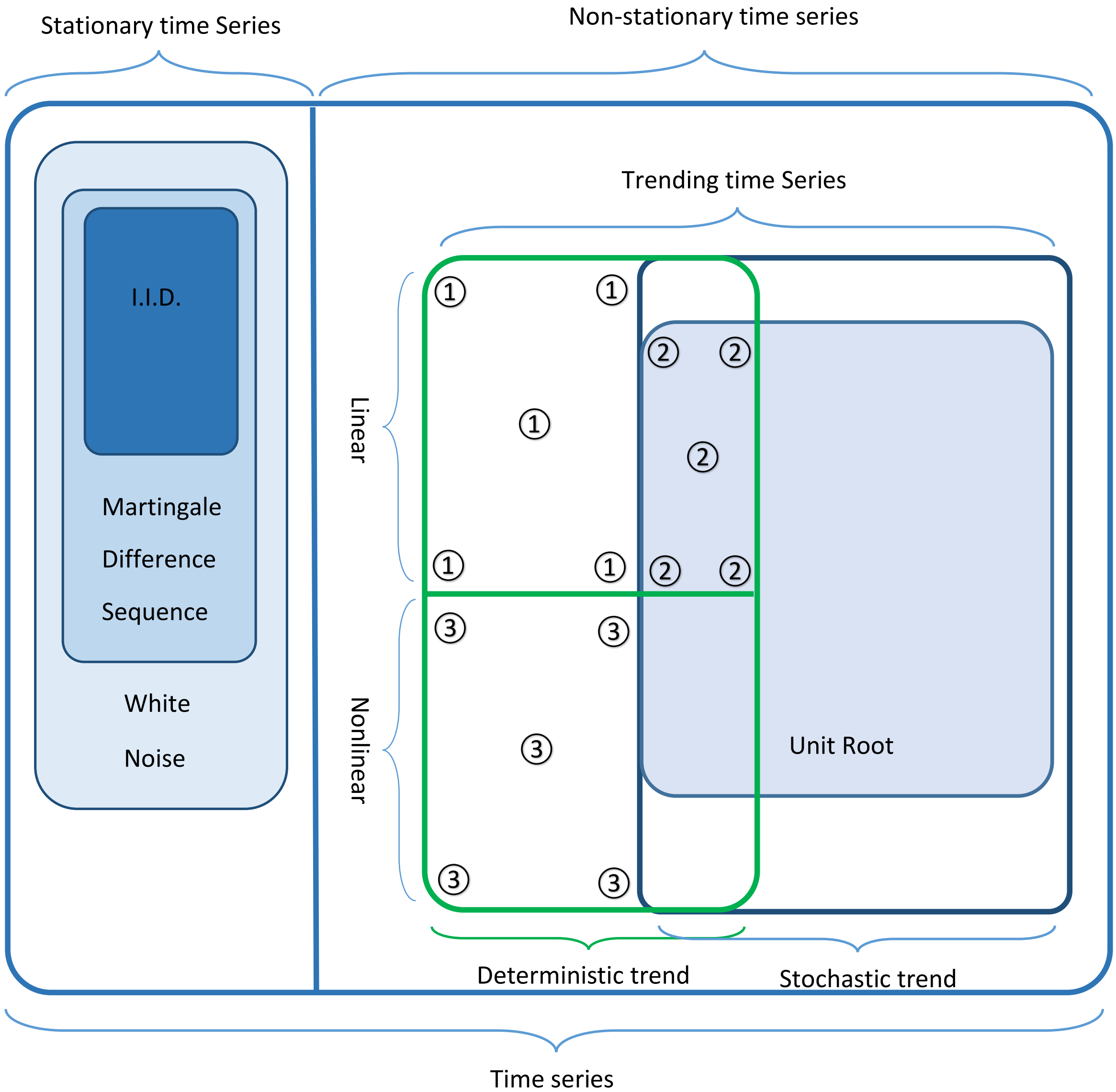
The primary objective of this thesis is to establish the estimation and inference methods for the coefficients in the trending regression model with endogeneity. We propose two methods (the nonparametric control function approach and the bias-correction method) to solve the problem of endogeneity in both cases with weak and strong trends respectively. We will show the properties of these estimators as well as some numerical and empirical examples in subsequent chapters.

1.2 The deterministic and stochastic trending time series

In the recent twenty years, nonstationary trending time series, as well as their regression models, have gained much attention. The challenges of the trending time series econometrics have been extensively discussed in Phillips (2001, 2003, 2005, 2010) and White and Granger (2011). As stated in these papers, trends are full of mysteries, and the sentence *'No one understands trends, but everyone sees them in the data'* as quoted at the beginning of this chapter has been one of the laws of modern econometrics. Trends contain a vast amount of information, and they have significant implications on various economic phenomena, such as the structural breaks, economic bubbles, and business cycles.

The generating process of the time series data is crucial for estimating the time series models. Figure 1.1 shows the basic categorization of time series. The data can either be stationary or nonstationary. Econometricians have established a systematic framework with hundreds of models for the stationary time series data. However, this thesis aims at constructing nonstationary time series models, in particular, the regression models for the trending time series represented by area ③.

Figure 1.1: The categorization of time series data.



In the trending time series data, trends can be either/both deterministic or/and stochastic. For the deterministic trend, it can be either linear or nonlinear. Area ① in

Figure 1.1 is the usual trend-stationary process, which is a stationary process around a linear deterministic trend. Specifically, it can be written as

$$x_t = a + bt + v_t, \tag{1.5}$$

for $t = 1, 2, .., n$, where v_t is a sequence of stationary innovations. Area ② is the unit root process. For instance, a random walk process with drift

$$x_t = \delta + x_{t-1} + v_t, \tag{1.6}$$

for $t = 1, 2, ..., n$, where $\delta \neq 0$ is the drift term and v_t is stationary as well. It is easy to find that the non-zero drift term δ would form a linear deterministic time trend because

$$x_t = \delta t + \sum_{s=1}^{t} v_s + x_0, \tag{1.7}$$

where x_0 is the initial value.

Area ③ is the type of trending time series we are going to explore in this thesis. It is a stationary process fluctuating around a nonlinear deterministic trend as equation (1.3).

In practice, the true generating process of the empirical time series is usually unknown and even unknowable. As suggested in Harvey (1997), the trend component can hardly be specified as a linear function of time unless the length of the time series is relatively short. In other words, nonlinear trends are often seen in the trending time series with a relatively longer horizon. Also, in terms of model specification, it is too restrictive to assume that the level and the slope parameters of the trend in the time series are constants over a long time period. Consequently, it is necessary to adopt the nonlinear time trend models to accommodate the nonlinear and nonstationary characteristics in the data.

By definition, a deterministic trend always map a given time point to a deterministic value. Otherwise, the trend is stochastic since its position is random at time t. Apparently, polynomial trends with constant coefficients are deterministic. For example, a linear trend $g(t) = 1 + 2t$, or a quadratic trend $g(t) = 1 + 2t + t^2$. On the contrary, stochastic trend appears when the level and slope parameters are not deterministic over time. For instance,

$$x_t = \alpha_t + \beta_t t + v_{1t}, \tag{1.8}$$

where $\alpha_t = \alpha_{t-1} + v_{2t}, \beta_t = \beta_{t-1} + v_{3t}$, for $t = 1, 2, ..., n$, $v_{it} \overset{i.i.d.}{\sim} N(0, \sigma_i^2)$, $\sigma_i^2 > 0$ for $i = 1, 2, 3$. The most intensively studied stochastic trend process is the unit root process, typically the random walk without drift

$$x_t = x_{t-1} + v_t, \tag{1.9}$$

for $t = 1, 2, ..., n$, where $v_t \overset{i.i.d.}{\sim} N(0, \sigma_v^2)$. The nonstationarity of the random walk process originates from the growing unconditional variance of x_t, i.e., $var[x_t] = t\sigma_v^2$, although its unconditional mean is always a constant over time. In this case, the level and the slope of the realized trend in x_t are stochastic instead of deterministic at any given time point t.

The structural form of a trend-stationary process is simply a stationary process about a deterministic time trend. One can obtain a stationary time series by removing such linear or nonlinear time trend. While for the unit root process, stationarity is achieved after taking the difference of the time series. Therefore, it is also named the *difference-stationary process*.

There are sharp differences between the economic interpretations of the two kinds of data generating processes. As the time series fluctuates about a deterministic time trend, the effects of the innovations in the trend-stationary process die out quickly, and they only cause transitory changes to the time series. While in the unit root process, the shocks cause permanent shifts to the time series as they are 100% accumulated without any loss. Economists are interested in revealing the data generating process of the empirical data, as it is critical to determining whether certain economic event would cause temporary or permanent effects on the economic variable.

Since the 1980s, economists and econometricians have established various unit root tests to statistically distinguish between the unit root process and the (trend-) stationary process. The *Augmented Dickey-Fuller unit root test* and the *Phillips-Perron unit* *root test* are the most widely used statistical tests; see Fuller (1976), Dickey and Fuller (1979) and Phillips and Perron (1988). The null and the alternative hypothesis in these tests are established as

$$\mathbb{H}_0: \quad x_t \text{ has a unit root}, \tag{1.10}$$

$$\mathbb{H}_1: \quad x_t \text{ is (trend) stationary}, \tag{1.11}$$

where x_t is the time series to be tested. Note that a deterministic time trend can be included in the test equation. Therefore the alternative hypothesis is allowed to be a trend-stationary process[3].

Testing for unit roots in the empirical time series data has been a necessary procedure before applying the time series models. Economists are aware that unlike in the stationary case, the estimation and inference of the coefficients are greatly different when the time series contain unit roots. In Nelson and Plosser (1982), fourteen U.S. macroeconomic time series are examined using the ADF test, and for most of the time series, the authors failed to reject the null hypothesis of unit root. This paper has been cited frequently, and similar results are also found in Said and Dickey (1984) and Perron (1988), where the error terms in the test equation are allowed to be serially correlated.

In conducting the unit root tests, the probability of making the *Type I error* is controlled by the significance level we choose, given that the tests do not suffer from size distortions. For instance, under 5% significance level, the probability of incorrectly rejecting $\mathbb{H}_0$, while in fact the null hypothesis is true, is less than 5%. Hence, it means that rejecting the null hypothesis is usually a statistically reliable conclusion for the hypothesis tests.

However, the probability of making the *Type II error* may be substantially large if the unit root test has power problems. The ADF tests are found to have low power against the alternative hypothesis of autoregressive process with roots near unity in

[3]In fact, the alternative hypothesis in the ADF test only allows for a linear deterministic time trend.

DeJong et al. (1992). In other words, the null hypothesis is rarely rejected even though the null hypothesis is false while the alternative hypothesis is true. Meanwhile, Rudebusch (1993) found that the existence of a unit root is quite uncertain in the U.S. real GNP. Diebold and Rudebusch (1991) found that the ADF test also has low power against the alternative of fractionally integrated time series. Therefore, in such unit root tests of (1.10) and (1.11), failing to reject the null hypothesis does not necessarily imply the existence of unit root in the time series due to the power problems.

A straightforward way to fix this problem is swapping the null and alternative hypothesis so that we are testing stationary against unit root. Kwiatkowski et al. (1992) propose the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) unit root test, in which unit root is taken as the alternative hypothesis. Thus, rejecting the null hypothesis strongly supports the existence of unit root in the time series. In DeJong and Whiteman (1991), the authors employ the Bayesian methods and find that only two of the Nelson-Plosser series contain unit roots. Nevertheless, if we fail to reject both null hypothesis in the ADF and KPSS tests, we are still ignorant about the existence of unit root in the data.

As discussed in the previous paragraphs, the unit root conclusion is questionable due to the power problems of the unit root tests. In practice, if the data exhibits an upward or downward trend, a linear function of time is usually included in the test equation of the ADF type tests. Specifically,

$$\Delta x_t = \underbrace{\alpha_0 + \delta t}_{\text{linear trend}} + \rho x_{t-1} + \alpha_1 \Delta x_{t-1} + \ldots + \alpha_p \Delta x_{t-p} + e_t, \tag{1.12}$$

where $\alpha_0 + \delta t$ is used to capture the linear deterministic trend in the data. However, as mentioned at the beginning of this section, Harvey (1997) suggests that for the empirical data, the deterministic trend is not necessarily linear. Perron (1989) establishes a unit root test in which one trend break[4] is allowed in both the null and the alternative hypotheses. Bierens (1997) uses Chebyshev polynomials to replace the linear deterministic trend terms in the test equation so that the test allows for nonlinear de-

[4] It is a special case for nonlinear trends.

terministic trends. Both papers find that for some time series in the Nelson-Plosser data set, we should reject the null hypothesis of unit root, which, however, is not rejected by the usual ADF test with linear trends.

To conclude, since 'unit root' is the null hypothesis to be tested and the power of the ADF test is low in some cases, the unit root conclusion is not statistically reliable. On the other hand, when the deterministic trend is allowed to be nonlinear functions of time, it is more likely to reach the conclusion that the time series is nonlinear trend-stationary rather than unit root. Therefore, it is reasonable to model the trending time series as a nonlinear trend-stationary process as (1.3).

1.3 An illustrative example

In the previous section, we addressed some problems and conflicting results in determining the existence of unit roots. In this section, we illustrate this phenomenon by an example with simulated time series as plotted in Figure 1.2.

The time series in the four subfigures are generated as the following steps.

(1) In the first subfigure, x_{1t} is a simulated random walk process without drift, i.e., $x_{1t} = x_{1t-1} + v_t$, where $v_t \overset{i.i.d.}{\sim} N(0,1)$ for $t = 1, 2, ..., 300$. As the trend is stochastic, for each realization, you may get different paths of the time series sequence. Without loss of generality, we focus on one realization of the random walk process as shown in the figure.

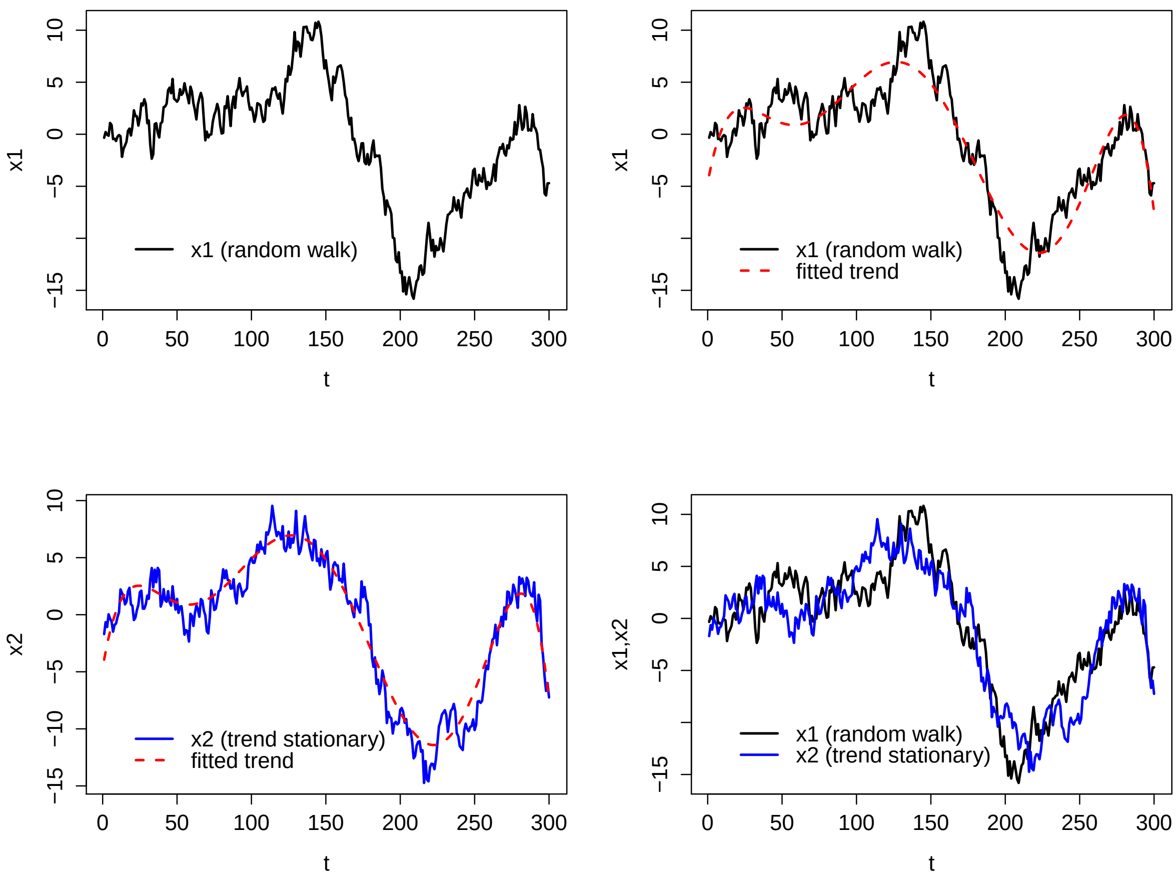
(2) We then suppose that the true data generating process of x_{1t} is unknown and approximate x_{1t} using a time polynomial.[5] Specifically, we run the following regression

$$x_{1t} = \alpha_0 + \alpha_1 t + ... + \alpha_6 t^6 + e_t, \tag{1.13}$$

and obtain the least squares estimates of $\alpha_0, ..., \alpha_6$. The spuriously estimated time

[5] Alternatively, one can also use nonparametric methods to fit a time trend.

Figure 1.2: Unit root process versus nonlinear trend-stationary process



trend is $\widehat{g}(t) = \widehat{\alpha}_0 + \widehat{\alpha}_1 t + ... + \widehat{\alpha}_6 t^6$, plotted as the red dashed line in the second subfigure.

(3) We create a nonlinear trend-stationary process x_{2t} in the third figure. The blue solid line in the lower-left subfigure is $x_{2t} = \widehat{g}(t) + u_t$, where $\widehat{g}(t)$ is the fitted trend in the previous step and u_t is a stationary AR(1) process. In particular, let $u_t = 0.6u_{t-1} + \eta_t$ for $u_0 = 0$ and $\eta_t \overset{i.i.d.}{\sim} N(0,1)$, for $t = 1, 2, ..., 300$.

(4) Lastly, the unit root process x_{1t} and the nonlinear trend-stationary process x_{2t} are plotted together in the fourth subfigure.

Visually, based merely on the fourth subfigure, one is hard to tell which line is a unit root process and which line represents a trend-stationary process as their paths are very close to each other. Statistical tests also fail to distinguish between the two processes. The ADF test, Phillips-Perron test and the KPSS test all suggest that both x_{1t} and x_{2t} exhibit unit roots. On the other hand, removing the time trend (the red dashed line) from both series gives two residual sequences, and all the unit root tests suggest that they are stationary time series under the 5% significance level. This result leads us to an opposite conclusion that the two sequences are nonlinear trend-stationary processes.

To conclude, given a sequence of time series (especially when the sample size is small), the unit root time series and the nonlinear trend-stationary time series are hardly distinguishable. Therefore, it is apparently necessary to study the nonlinear trend-stationary process as well as its regression models.

1.4 Detrending methods and their problems

As discussed in the first section, the stationary assumption for most of the time series models is violated when the time series exhibit trends. In practice, economists usually transform the non-stationary time series into stationary versions, and this process is called 'detrending', or 'the stationarization of the trending time series'.

Various detrending methods are investigated in Canova (1998). The author finds that different detrending methods lead to different patterns of estimated trends and cycles. In the section, I summarize some of the detrending methods and discuss their problems.

(1) *Fitting the polynomial functions of time*

Fitting a polynomial trend is the simplest and the oldest method for detrending. When the time series is trend-stationary, one only needs to approximate the sequence using a pre-specified polynomial function of time, and then subtract the estimated trend to obtain a stationary process. For example, suppose that

$$x_t = \sum_{k=0}^{p} \alpha_k t^k + e_t, \tag{1.14}$$

where the polynomial order p is known and e_t is a stationary process. We regress x_t over the polynomials of time and obtain the residuals as the detrended version of the time series

$$\widehat{e_t} = x_t - \sum_{k=0}^{p} \widehat{\alpha_k} t^k \sim I(0), \tag{1.15}$$

where $\widehat{\alpha_k}$ is the OLS estimate of α_k for $k = 0, 1, ..., p$. In most of the cases, we only include a linear time trend and let $p = 1$.

(2) *Taking difference*

Taking difference[6] is another commonly used method to eliminate the trends in the nonstationary time series. When the series is an integrated process, for example, an $I(d)$ process with d being a positive integer, then taking the difference of x_t for d times gives a stationary process. That is, if $x_t \sim I(d)$, then $\Delta^d x_t \sim I(0)$.

(3) *The Hodrick and Prescott's filter*

The HP-filter method was established in Hodrick and Prescott (1997), and it is widely applied by macroeconomists. The HP-filter separates the trend and cycle components by solving the optimization problem

$$\min_{g_1,...,g_t,...,g_n} \sum_{t=1}^{n} (x_t - g_t)^2 + \lambda \sum_{t=2}^{n} \Big((g_{t+1} - g_t) - (g_t - g_{t-1}) \Big)^2, \tag{1.16}$$

where $\{g_t\}_{t=1}^n$ is the trend sequence to be estimated. The smoothness of the estimated trend depends on the smoothing parameter, which is usually denoted as λ. In practice, λ takes different values for different frequencies of the time series. For example, $\lambda = 100, 1600, 14400$ for the yearly, quarterly and monthly data respectively.

[6]Here we only consider time series data. Sometimes, differencing for the cross-sectional data are quite useful in practice, see the first chapter of Yatchew (2003).

Some other detrending methods are also discussed in Canova (1998), such as *the Beveridge-Nelson's method, the frequency domain method, the unobserved component model, the one-dimensional index model, the model of common deterministic trends*, and *the model of common stochastic trends*.

Since we are usually ignorant about the true generating process of the time series data, these detrending methods should be used with caution. At the same time, we should be aware of the specific assumptions before using these methods. For example, the Hodrick-Prescott filter can only be applied to an $I(2)$ process. In the literature, these methods are criticized in many research papers.

First, trend-elimination throws away a vast amount of information in the data. Particularly, long-run information with low frequency is wiped out from the time series, and only short-run disturbances of high frequency information are left behind. Rao (2010) argues that the regressions using the differenced variables are useless to verify economic theories as they only reflect the relationship between short-run variables rather than the long term equilibrium relationships. As shown by an example in Cochrane (2012), the relationship between the original nonstationary data is quite significant, while the scatter plot shows little correlation between the differenced sequences. Moreover, the time series sometimes need to be differenced more than once before achieving stationarity. Eventually, little information is maintained in the stationary time series after differencing for several times.

Second, as the true data generating process is unknown for most of the time, misuse of the detrending methods usually lead to severe statistical problems. Nelson and Kang (1981) suggest that if we remove an estimated polynomial trend from an $I(1)$ process, we may introduce pseudo-periodic behavior in the detrended series. Therefore, the regression results make no sense as we have artificially introduced the autocorrelations. Moreover, for a nonlinear trend-stationary process, one may need to take the difference more than once to obtain a stationary process. For example, if $x_t = t^2 + v_t$, where v_t is an $I(0)$ process, the second order difference of x_t contains moving average unit roots as $\Delta^2 x_t = 2 + \Delta^2 v_t$ is a non-invertible process. For the time series with both global and local trends as $x_t = a + bt + g(t) + v_t$, where $g(t)$ is a local weak trend and v_t is stationary, differencing only eliminates the strong global trend $a + bt$, but the weak local trend $\Delta g(t)$ still remains in Δx_t.

To conclude, the trend-elimination methods delete much more useful information than expected. Meanwhile, they may cause other statistical issues that lead to complicated econometric problems. To overcome these problems, we propose trending regression models that directly deal with the nonstationary time series instead of their stationarized versions.

1.5 Nonstationary trending time series models

As we discussed in the previous section, detrending the nonstationary time series may cause unexpected problems. The verification of particular economic theories requires the application of trending time series models that regress the nonstationary time series directly.

The prevalence of the unit root conclusions leads to the popularity of the co-integration models, in which the combination of the integrated time series forms a stationary process. Such combinations are considered as equilibrium relationships in economics. Since the deviations from the equilibrium states are stationary over time, co-integration indicates stable relationships in the long-run.

In the two influential papers by Granger (1981) and Engle and Granger (1987), the authors investigate the representation, estimation, and testing of the co-integration models. Specifically, a co-integration relationship between x_t and y_t takes the form as

$$y_t = x_t'\beta + e_t, \tag{1.17}$$

where $\{x_t, y_t\}$ are integrated processes, e_t is an I(0) process, β is called the cointegrating coefficient (vector).

However, in practice, it is not easy to find the evidence of co-integration when the length of the sample period is relatively long. That is because the true co-integration coefficient (vector) may be time-varying, and a misspecified constant coefficient (vector) can hardly describe the changes in the cointegrating relationship. Hansen (1992) and Quintos and Phillips (1993) established the Lagrange multiplier tests to examine the parameter consistency in the co-integration models. To allow for the changes in the coefficients, Park and Hahn (1999) developed the co-integration regression model with time-varying coefficients of the form

$$y_t = x_t'\beta_t + u_t, \tag{1.18}$$

where x_t is a k-dimensional vector of I(1) processes, $\beta_t = \beta(t/n)$ is a smooth function defined on $[0,1]$ representing the slowly changing coefficients. [7] The paper also studied the U.S. automobile demand, and the authors showed that the time series data are cointegrated with smoothly varying coefficients. However, they are not cointegrated when the coefficients are restricted as constants for the whole sample period.

It is usually necessary to include nonlinear and nonparametric time trends in many time series and panel data regression models. For example, Gao and Hawthorne (2006) investigate the climate time series data using a semi-parametric model. In Cai (2007), the author proposes a varying coefficient trending time series model, where the error terms are allowed to be autocorrelated. However, both papers require the regressors to be stationary. Hence, they can not deal with the regression between trending time series.

In panel data models, nonlinear time trends are also taken into account to capture the trending feature. Robinson (2012) considers a nonparametric trending regression with cross-sectional dependence, where the error terms are allowed to be correlated and heteroscedastic over the cross-section. Chen et al. (2012) study a semi-parametric trending panel data model with cross-sectional dependence. By incorporating unknown nonlinear deterministic trends in both the regression equation and the data

[7] Recently, Phillips et al. (2017) also consider the same model by a kernel approach.

generating process, the model is capable of accommodating the nonstationarity in the data. Unfortunately, none of these models considers the endogeneity issue, which we will take into account in this thesis.

1.6 The endogeneity bias and the strength of the trend

The endogeneity problem affects the performance of the simple OLS estimator when it occurs in the regression model. Specifically, in the univariate linear regression model with stationary time series, the OLS estimator is defined as[8]

$$\widehat{\beta}_{OLS} = \left(\sum_{t=1}^{n} x_t^2\right)^{-1} \left(\sum_{t=1}^{n} x_t y_t\right) = \beta + \left(\frac{1}{n}\sum_{t=1}^{n} x_t^2\right)^{-1} \left(\frac{1}{n}\sum_{t=1}^{n} x_t e_t\right), \tag{1.19}$$

where endogeneity suggests that $\mathrm{E}[x_t e_t] \neq 0$, hence $\widehat{\beta}_{OLS}$ is biased and inconsistent for β as $n \to \infty$.

When x_t is nonstationary as equation (1.3), the strength of the trend matters. Recall that x_t is generated by

$$x_t = g(t) + v_t,$$

where v_t is stationary and $\mathrm{E}[v_t] = 0, \mathrm{E}[v_t^2] = \sigma_v^2 < \infty$. Suppose that $\sum_{t=1}^{n} g(t)^2 = O(n^d)$ for some $d \geq 1$. Then as $n \to \infty$, we have

$$\frac{1}{n^d}\sum_{t=1}^{n} x_t^2 \longrightarrow_P Q, \tag{1.20}$$

for some $0 < Q < \infty$. The value of d is determined by the strength of the trend in x_t, i.e., stronger trends lead to larger values of d. Then the endogeneity bias in the OLS estimator can be written as

$$\widehat{\beta}_{OLS} - \beta = \left(\sum_{t=1}^{n} x_t^2\right)^{-1} \left(\sum_{t=1}^{n} x_t e_t\right) = \left(\frac{1}{n^d}\sum_{t=1}^{n} x_t^2\right)^{-1} \left(\frac{1}{n^d}\sum_{t=1}^{n} x_t e_t\right), \tag{1.21}$$

where d is the magnitude parameter for $g(t)$. Therefore, the OLS estimator is consistent when $d > 1$. In other words, when the trend is strong, $n^{-d}\sum_{t=1}^{n} x_t e_t = O_P(n^{1-d}) = o_P(1)$,

[8] For simplicity, we only consider the univariate regression model.

and the endogeneity bias vanished when $n \to \infty$. Otherwise, it is inconsistent when $d = 1$ because the trend is too weak to dominate the stationary component and the endogenous correlation causes permanent bias in the estimator. Phillips and Hansen (1990) study the endogeneity in the co-integration regression that the regressor follows a pure random walk process. In their model, $d = 2$ and the OLS estimator is consistent. However, the limit distribution of $n(\widehat{\beta}_{OLS} - \beta)$ is not centered around zero because of endogeneity[9]. In this thesis, we propose two different estimation methods for the weak trend regression when $d = 1$ and the strong trend regression when $d > 1$.

1.7 The relationship between income and consumption

In this thesis, we re-examine the relationship between the aggregate personal consumption expenditure and the disposable income as an empirical example, which has been extensively studied in many research papers. This simple regression is related to the permanent income hypothesis (PIH), which has been a very popular topic since the 1970s in many research papers.

Hall (1978) solves the consumer's optimization problem under the condition of rational expectations. He concludes that the consumption should follow a random walk process that the changes are not predictable under the permanent income hypothesis, given that the real interest rate is a constant value over time. In other words, consumption tracks the permanent income, and it is not sensitive to the changes of current income.

However, opposite conclusions are found in some other papers. Flavin (1981) develops a structural econometric model of consumption to estimate the excess sensitivity of consumption to current income. Such excess sensitivity should be zero under the permanent income hypothesis. The empirical result shows a strong effect of excess

[9] When the time series are $I(1)$, the consistency of the OLS estimator is not affected by endogeneity because the stochastic trend is very strong that dominates the time series. However, since the endogenous correlation causes a bias in the limit distribution of $n(\widehat{\beta}_{OLS} - \beta)$, the inference is severely affected.

sensitivity in the consumption to current income. Therefore, the permanent income hypothesis is rejected. Similar results are also found in Flavin (1984) and Bernanke (1985). While in these papers, they assume that the time series of income is a stationary process around a deterministic trend.

Mankiw and Shapiro (1985) noticed the statistical evidence in Nelson and Plosser (1982) that income and consumption exhibit unit roots. They showed that excess sensitivity was favorable if we ignored the unit root and conducted inappropriate detrending, which could bring spurious cycles in the transformed data; also see Nelson and Kang (1981, 1984). Therefore, the Flavin's conclusions are likely to be biased and not reliable.

King et al. (1991) proposed a co-integration model with known cointegrating vector $(1, -1)$ for the logarithms of consumption and income, and the model is regarded as a special version of the permanent income hypothesis. Han and Ogaki (1997) considered the co-integration between both the stochastic trend and the deterministic trend, and they find that both trends are cointegrated, implying that the post-war U.S. saving rate is stable in the long-run.

Meanwhile, the estimation of such regression relationship of consumption over income also suffers from the problem of endogeneity as both of them belong to a system of simultaneous equations. Phillips and Hansen (1990) developed the bias-correction method for the co-integration models with endogeneity. In Hansen and Phillips (1990), they applied the bias-correction method to the co-integration model of the per capita personal consumption over the per capita personal income. In their paper, the permanent income hypothesis is not rejected as unit coefficient 1 is included in the 95% confidence interval of the estimated coefficient.

We also consider the relationship between income and consumption in this thesis. However, we treat the logarithms of income and consumption as nonlinear trend-stationary time series. Also, we deal with the endogeneity issue using the proposed methods when estimating the linear regression models.

1.8 The contributions and the organization of the thesis

The main objective of this thesis is to establish an estimation method for the linear trending time series models with endogeneity. The following items are the main contributions of this thesis.

(1) The thesis explains in detail that it is difficult to distinguish between a unit root process and a nonlinear trend-stationary process.

(2) The thesis studies the linear regression model with endogenous trending time series, and the trending time series are assumed to be stationary processes about nonlinear and nonparametric time trends.

(3) A trending magnitude parameter is defined to characterize the strength of the trend. Therefore, the time trends are categorized into weak and strong kinds.

(4) For the weak trend case, we use a nonparametric control function to deal with the endogenous correlation. I prove that the conventional estimator for the semi-parametric partially linear model is still unbiased and consistent, although the population version of the identifiability condition is not satisfied.

(5) For the strong trend case, we first discussed the asymptotic properties of the simple OLS estimators and then propose a bias-corrected estimator to adjust the endogeneity bias in the simple OLS estimator. The bias-correction procedure significantly improves the performance of the t-tests in the inferences on the coefficients.

(6) Both methods have the advantage that there is no need to find instrumental variables because we have used the information in the time trends. Another advantage is that the limit distributions in the asymptotic results are *normal distributions* instead of non-standard distributions such as Brownian Motions for the co-integration models. This brings much convenience in conducting the hypothesis tests for the coefficients.

The remaining parts of the thesis are organized as follows. Chapter 2 deals with the endogeneity issue in the weak trending regression model by using a nonparametric control function approach. Chapter 3 first explores the asymptotic properties of the simple OLS estimators, and then introduces the bias-corrected estimator that adjusts the endogeneity bias. The performances of the estimators in Chapter 2 and Chapter 3 are examined by Monte Carlo simulations in Chapter 4. The results are followed by an empirical example to illustrate the implementation steps of the methods proposed in this thesis. Chapter 5 concludes the thesis and several future research directions are discussed in Chapter 6. Appendix A presents a 'constructed instrumental variable approach' to solve the endogeneity problem. [10] The detailed mathematical proofs of the main Theorems, as well as the Lemmas, are provided in the appendices.

[10] This method was considered in the first year of my PhD candidature when I started to study this research topic. For the completeness of my PhD research, I include this method in the Appendix.

Chapter 2

Endogeneity in the weak trending regression

2.1 Weak trends and the control function approach

In this section, we consider the weak trending time series regression model with endogeneity. In the literature, the time trend term is commonly written as

$$x_t = g(\tau_t) + v_t, \tag{2.1}$$

for $t = 1, 2, ..., n$, where $\tau_t = t/n$ and v_t is a stationary $I(0)$ process. Suppose that $g(\cdot)$ is a continuous function defined on $[0, 1]$ and square integrable, then equation (2.1) forms a nonstationary time series with weak trend since

$$\frac{1}{n}\sum_{t=1}^{n} g(\tau_t)^2 \longrightarrow \int_0^1 g(\tau)^2 d\tau < \infty, \tag{2.2}$$

as $n \to \infty$. Therefore, the trending parameter $d = 1$ and

$$\frac{1}{n}\sum_{t=1}^{n} x_t^2 \longrightarrow_P \int_0^1 g(\tau)^2 d\tau + \sigma_v^2, \tag{2.3}$$

where σ_v^2 is the variance of v_t. The advantage of writing the weak trend as $g(\tau_t)$ is that it allows for the accumulation of information on the compact interval $[0, 1]$. In other words, for any $\tau \in [0,1]$, the information about the trend function $g(\cdot)$ in the small neighborhood of τ grows with the sample size n. Therefore, the nonparametric estimation of the trend function $g(\cdot)$ is getting more accurate. Therefore, the nonparametric estimate of $\widetilde{g}(\tau_t)$ is consistent when n goes to infinity.

Remark 2.1.1. The function $g(\tau_t)$ is a standardized version of the form $\widetilde{g}(t)$ in the previous chapter[1]. In fact, we have re-defined the trend function $\widetilde{g}(t)$ as $g(\tau_t)$. It is straightforward that there is a one-to-one correspondence between $g(\tau_t)$ and $\widetilde{g}(t)$, hence, for estimation purposes, there is no difference to estimate $g(\tau_t)$ instead of $\widetilde{g}(t)$.

With equation (2.1) as the data generating process of x_t, we then set up the weak trending regression model with endogeneity as

$$y_t = \alpha + x_t'\beta + e_t, \tag{2.4}$$

$$x_t = g(\tau_t) + v_t, \tag{2.5}$$

where e_t and v_t are endogenously correlated. The ordinary least square estimator is defined as

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \begin{pmatrix} 1 & n^{-1}\sum_{t=1}^n x_t \\ n^{-1}\sum_{t=1}^n x_t & n^{-1}\sum_{t=1}^n x_t x_t' \end{pmatrix}^{-1} \begin{pmatrix} n^{-1}\sum_{t=1}^n y_t \\ n^{-1}\sum_{t=1}^n x_t y_t \end{pmatrix}. \tag{2.6}$$

Replacing y_t by equation (2.4), we have

$$\begin{pmatrix} \widehat{\alpha} - \alpha \\ \widehat{\beta} - \beta \end{pmatrix} = \begin{pmatrix} 1 & n^{-1}\sum_{t=1}^n x_t \\ n^{-1}\sum_{t=1}^n x_t & n^{-1}\sum_{t=1}^n x_t x_t' \end{pmatrix}^{-1} \begin{pmatrix} n^{-1}\sum_{t=1}^n e_t \\ n^{-1}\sum_{t=1}^n x_t e_t \end{pmatrix}. \tag{2.7}$$

The problem of endogeneity implies that $\mathrm{E}[v_t e_t] \neq 0$. Since $g(\tau_t)$ is deterministic in (2.5), the simple OLS estimators of α and β are biased and inconsistent in that by the Law of Large Numbers, $n^{-1}\sum_{t=1}^n x_t e_t$ does not converge to zero as the sample size n goes to infinity.

To deal with the problem of endogeneity, we employ a nonparametric control function approach. The control function describes the endogenous correlation between the

[1]To discriminate the trend functions, $\widetilde{g}(t)$ here is the same as $g(t)$ in the previous chapter.

error terms of e_t and v_t in equations (1.2) and (1.3). Specifically, we assume

$$e_t = \lambda(v_t) + u_t, \tag{2.8}$$

where $\lambda(v_t) = \mathrm{E}[e_t|v_t]$ and $u_t = e_t - \mathrm{E}[e_t|v_t]$. This approach is also followed by Amihud and Hurvich (2004) and Cai and Wang (2014) for solving the endogeneity problem in the predictive regression models. But they assume a linear functional form for $\lambda(v_t)$. We aim to avoid any potential misspecification arising from this linearity assumption and let the control function be an unspecified nonparametric form. Replacing e_t in (1.2) with (2.8) yields a semi-parametric partially linear model

$$y_t = \alpha + x_t'\beta + \lambda(v_t) + u_t. \tag{2.9}$$

Since u_t is assumed to be uncorrelated with v_t and x_t, the problem of endogeneity disappears in the augmented model. We also assume $\mathrm{E}[\lambda(v_t)] = 0$, so that we don't have the problem in identifying the intercept term α and the nonlinear term $\lambda(v_t)$ in model (2.9). However, the cost of using the nonparametric control function is that we need to assume some regularity conditions on v_t since it is used as the smoothing parameter in the nonparametric kernel estimation.

The semi-parametric partially linear model (2.9) has been extensively studied and widely applied; see Robinson (1988), Härdle et al. (2000), Gao (2007), Li and Racine (2007), as well as many other related papers. Our setting differs from those in the literature in two ways. First is that the disturbance v_t is not observed and must be estimated from x_t. Second is that the regressors in (2.9), x_t and v_t, differ only by the deterministic trend $g(\tau_t)$, implying that $\mathrm{E}(x_t|v_t) = x_t$ and a potential identification problem for β — in fact, the usual identification condition for β in partial linear models (see Robinson (1988)) is not satisfied here. However, we show in the subsequent sections that the conventional estimators remain valid for the trending time series regression since the sample identifiability condition can be satisfied.

2.2 Estimation method

2.2.1 Nonparametric kernel estimation method

The nonparametric kernel estimation method is commonly applied to estimate the density functions and the regression functions. For example, in the nonparametric regression model,

$$y_t = m(x_t) + e_t, \tag{2.10}$$

$m(x_t)$ is the conditional mean $\mathrm{E}[y_t|x_t]$, and e_t is the usual stationary and ergodic error term. One can estimate the conditional mean function nonparametrically using the local constant or local polynomial kernel estimation methods. The local constant estimator (also called the Nadaraya-Watson estimator) is defined as

$$\widehat{m}(x_t) = \widehat{\mathrm{E}}_h[y_t|x_t] = \sum_{s=1}^{n} w_{ns}(t) y_s, \tag{2.11}$$

where

$$w_{ns}(t) = \frac{K_1\left(\frac{x_s - x_t}{h}\right)}{\sum_{q=1}^{n} K_1\left(\frac{x_q - x_t}{h}\right)}, \tag{2.12}$$

for $s = 1, 2, ..., n$, $K_1(\cdot)$ is the kernel function and h is the bandwidth.

The nonparametric kernel method can also be applied to estimate the time trend in the data generating process of (2.5).

Replacing x_t by τ_t, we obtain the local constant estimator for the weak trend term in (2.5)

$$\widehat{g}(\tau_t) = \sum_{s=1}^{n} w_{ns}^{*}(t) x_s, \tag{2.13}$$

where

$$w_{ns}^{*}(t) = \frac{K_2\left(\frac{\tau_s - \tau_t}{b}\right)}{\sum_{q=1}^{n} K_2\left(\frac{\tau_q - \tau_t}{b}\right)}, \tag{2.14}$$

for $s = 1, 2, ..., n$, $K_2(\cdot)$ is another kernel function and b is another bandwidth.

2.2.2 Model identification and estimation

In this section, we outline the estimation method for β in (2.9). Equation (2.9) has the form of a partially linear regression model, with the additional complication that v_t is unobserved. Suppose first, however, that v_t were observed. Then, following the usual approach (Robinson (1988)) to partial linear estimation, taking the expectations of (2.9) conditional on v_t and subtracting this from (2.9) gives

$$y_t - \mathrm{E}[y_t|v_t] = (x_t - \mathrm{E}[x_t|v_t])'\beta + (u_t - \mathrm{E}[u_t|v_t]), \tag{2.15}$$

leaving β as the coefficient in a linear regression of $y_t - \mathrm{E}(y_t|v_t)$ on $x_t - \mathrm{E}(x_t|v_t)$. In this case there appears to be an identification problem because the deterministic nature of the trends $g(\tau_t)$ implies that $\mathrm{E}[x_t|v_t] = x_t$, leading to β disappearing from the model (2.15). However this apparent identification problem can be shown not to apply to the sample version of (2.15). We do the subtraction with their sample versions of the conditional means as (2.11) and obtain

$$y_t - \widehat{\mathrm{E}}_h[y_t|v_t] = (x_t - \widehat{\mathrm{E}}_h[x_t|v_t])'\beta + (\lambda(v_t) - \widehat{\mathrm{E}}_h[\lambda(v_t)|v_t]) + (u_t - \widehat{\mathrm{E}}_h[u_t|v_t]), \tag{2.16}$$

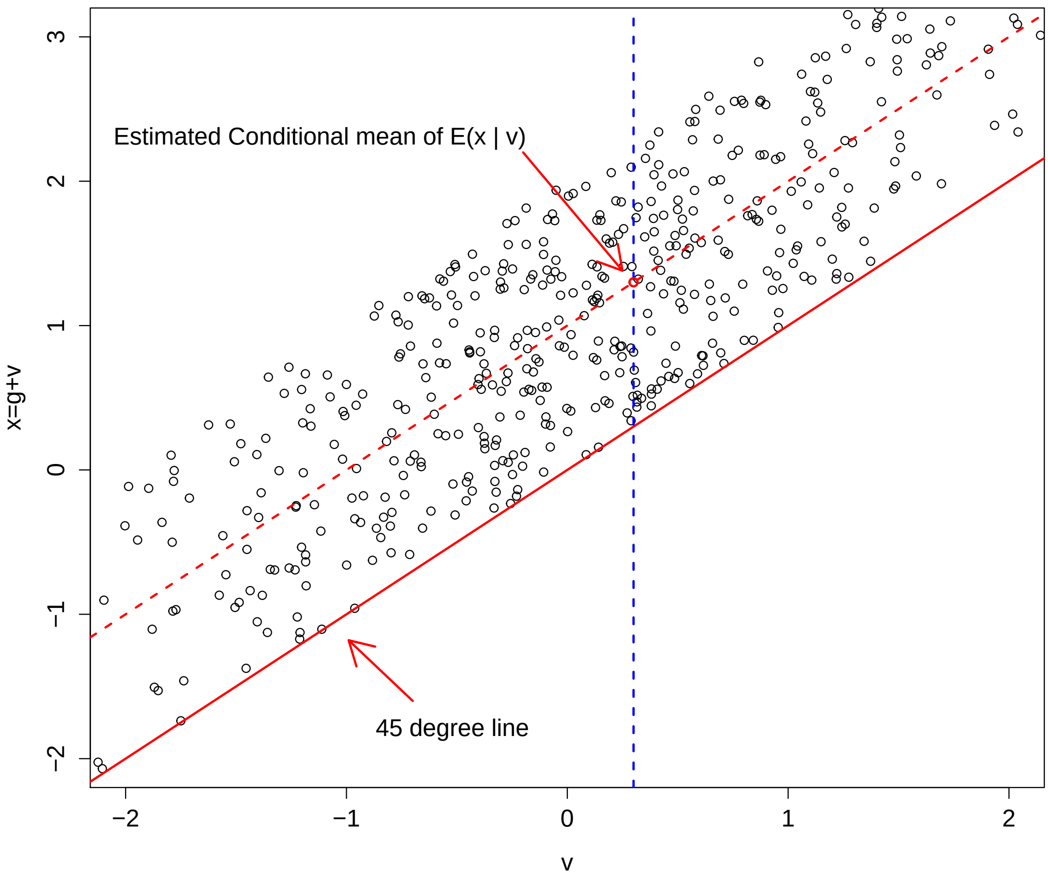
which includes the additional estimation error $\lambda(v_t) - \widehat{\mathrm{E}}_h[\lambda(v_t)|v_t]$ (shown to be negligible for sufficiently large n). In nonparametric analysis of stationary time series, $\widehat{\mathrm{E}}_h[x_t|v_t]$ is consistent for $\mathrm{E}[x_t|v_t]$ as the sample size n goes to infinity, but this becomes invalid when deterministic time trends are present in x_t. To explain this inconsistency, suppose that x_t follows equation(2.1), and $\widehat{\mathrm{E}}[x_t|v_t]$ is the local constant kernel estimator for the conditional mean $\mathrm{E}[x_t|v_t]$ defined as

$$\widehat{\mathrm{E}}_h[x_t|v_t] = \sum_{s=1}^{n} w_{ns}(t)x_s = \sum_{s=1}^{n} w_{ns}(t)g(\tau_s) + \sum_{s=1}^{n} w_{ns}(t)v_s = \widehat{\mathrm{E}}_h[g(\tau_t)|v_t] + \widehat{\mathrm{E}}_h[v_t|v_t]. \tag{2.17}$$

Since v_s for $s = 1,\dots,n$ is the smoothing variable in the weight function $w_{ns}(t)$, whose value is not linked with the time index, we then have

$$\widehat{\mathrm{E}}_h[g(\tau_t)|v_t] = \sum_{s=1}^{n} w_{ns}(t)g(\tau_s) \longrightarrow_P \overline{g}, \tag{2.18}$$

Figure 2.1: The inconsistency of the conditional mean estimator.



where $\overline{g} = \int_0^1 g(\tau)d\tau$. Graphically, we consider the scatter plot of x_t versus v_t in Figure 2.1. The red 45° solid line represents the function of $m(v) = v$. Therefore, each point in the scatter plot of x_t versus v_t ends up with an upward shift from the 45° line (suppose that the trend function $g(\cdot)$ is always positive). Again, since the value of v_t is not linked with the value of the time index t, the amount of the shift is random for a given value of v_t. Therefore, the sample version of the expected value of x_t given $v_t = v$ is approximately $v + \overline{g}$, where $\overline{g} = \int_0^1 g(\tau)d\tau$, which is the limit of the average of $g(\tau_t)$.

This implies that $x_t - \widehat{\mathrm{E}}_h[x_t|v_t]$ is not a constant value of zero in (2.16), even though the population expression shows that $x_t - \mathrm{E}[x_t|v_t] = 0$ for all t. Therefore, β can be identifiable in (2.16). The condition $g(\tau_t)$ is not a constant for all t is analogous to the time trend $g(\tau_t)$ being a relevant instrumental variable for x_t.

Formally, the identifiability condition for β in the partially linear models (Robinson (1988), Gao (2007), Li and Racine (2007)) is that the matrix

$$\Sigma = \mathrm{E}\Big[(x_t - \mathrm{E}[x_t|v_t])(x_t - \mathrm{E}[x_t|v_t])'\Big] \tag{2.19}$$

should be positive definite, but this does not hold in our context because $\mathrm{E}[x_t|v_t] = x_t$. However, with further assumptions, the sample identifiability condition

$$\widetilde{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n}\Big(x_t - \widehat{\mathrm{E}}_h[x_t|v_t]\Big)\Big(x_t - \widehat{\mathrm{E}}_h[x_t|v_t]\Big)' \longrightarrow_P Q, \tag{2.20}$$

can be satisfied with Q being positive definite, see Theorem 2.3.1 below. We wish to emphasize this as it is the key condition that makes the conventional estimator work. The smoothing operation over x_t with respect to the sequence v_t, for $t = 1, 2, ..., n$, is in fact an averaging process for $g(\tau_t)$, yielding $\overline{g}$ as $n \to \infty$. From the perspective of model estimation, the trending feature in x_t provides additional information to help with the estimation of the coefficients. In this way, the coefficients β can be identified and estimated. Of course, further assumptions are necessary to guarantee that the estimator is well-defined.

Based on the preceding discussion of identification, the estimation of β when v_t is observed could be based on the usual partial linear regression estimator

$$\widetilde{\beta} = \left(\sum_{t=1}^{n}(x_t - \widehat{\mathrm{E}}_h[x_t|v_t])(x_t - \widehat{\mathrm{E}}_h[x_t|v_t])'\right)^{-1}\sum_{t=1}^{n}(x_t - \widehat{\mathrm{E}}_h[x_t|v_t])(y_t - \widehat{\mathrm{E}}_h[y_t|v_t]), \tag{2.21}$$

which is the same as the one in Robinson (1988). The sample identifiability condition (2.20), along with some regularity conditions given below, are sufficient for the consistency of $\widetilde{\beta}$. However, $\widetilde{\beta}$ is infeasible since v_t is not observable. Therefore we estimate the trends $g(\tau_t)$ using nonparametric regression of x_t on τ_t, and define the estimator of v_t as the residuals

$$\widehat{v_t} = x_t - \widehat{\mathrm{E}}_b[x_t|\tau_t], \tag{2.22}$$

where $\tau_t = t/n$ and a different bandwidth b is used for the regression on τ_t. The feasible estimator of β is then obtained by replacing v_t in (2.21) using $\widehat{v_t}$

$$\widehat{\beta} = \left(\sum_{t=1}^{n} (x_t - \widehat{\mathrm{E}}_h[x_t|\widehat{v_t}])(x_t - \widehat{\mathrm{E}}_h[x_t|\widehat{v_t}])' \right)^{-1} \sum_{t=1}^{n} (x_t - \widehat{\mathrm{E}}_h[x_t|\widehat{v_t}])(y_t - \widehat{\mathrm{E}}_h[y_t|\widehat{v_t}]). \tag{2.23}$$

It is not necessary that the choice of kernels in the nonparametric regressions $\widehat{\mathrm{E}}_b[x_t|\tau_t]$, $\widehat{\mathrm{E}}_h[x_t|\widehat{v_t}]$ and $\widehat{\mathrm{E}}_h[y_t|\widehat{v_t}]$ should be the same. Finally, the intercept term can be estimated by

$$\widehat{\alpha} = \frac{1}{n} \sum_{t=1}^{n} (y_t - x_t' \widehat{\beta}). \tag{2.24}$$

2.3 The main results

2.3.1 Assumptions

We first make the following assumptions for establishing the asymptotic results.

Assumption 2.3.1. Assume that $g(\tau) = (g_1(\tau), g_2(\tau), ..., g_k(\tau))'$ is a $k \times 1$ vector of functions and each $g_i(\cdot)$ is a continuous and bounded function defined on $[0,1]$ with continuous derivatives of up to the second order for $i = 1, 2, ..., k$. We also assume that

$$Q = \int_0^1 \left(g(\tau) - \overline{g}\right)\left(g(\tau) - \overline{g}\right)' d\tau, \tag{2.25}$$

is a $k \times k$ positive definite matrix when $k > 1$ and a positive scalar when $k = 1$, where $\overline{g} = \int_0^1 g(\tau) d\tau$.

Assumption 2.3.1 regulates the weak trend components in the regressors. The positive definiteness of the Q matrix rules out the case of collinearity when $k > 1$. When $k = 1$, Q is always positive as long as the trend $g(\tau)$ is not a horizontal line. This condition ensures that the coefficients can be identified properly and the estimator is well-defined.

Assumption 2.3.2. $\lambda(\cdot)$ is a continuous and differentiable function defined on $\mathbb{R}^k \to \mathbb{R}^1$. Denote the first order derivative of $\lambda(\cdot)$ as $\zeta(z) = \lambda'(z)$.

Assumption 2.3.3. (i) The error term u_t is a stationary α-mixing time series with mixing-coefficient $\alpha(\cdot)$ satisfying $\sum_{d=1}^{\infty} \alpha^{\frac{\delta}{2+\delta}}(d) < \infty$, for some $\delta > 0$ such that $\mathrm{E}\left[|u_t|^{2+\delta}\right] < \infty$.
(ii) The error term v_t is strictly stationary with zero mean and finite variance. The two sequences u_t and v_s are independent for $t, s = 1, 2, ...n$.

Assumption 2.3.3(i) contains the standard requirements for the stationary α-mixing time series. The independence of u_t and v_t guarantees that the endogenous correlation is separable and thereby can be represented by the control function (2.8). It is a strong condition that u_t and v_s are independent for all $t, s = 1, 2, ..., n$, however, we need this assumption to prove the asymptotic results. Meanwhile, we address the following restrictions on v_t to regulate the weak dependence instead of using the mixing-conditions.

Assumption 2.3.4. Let $f(z)$ be the marginal distribution of the strictly stationary process v_t for $t = 1, 2, ..., n$. Let $f_{t_1,...,t_p}(z_1, ..., z_p)$ be the joint probability density function of $v_{t_1}, v_{t_2}, ..., v_{t_p}$ for $p > 1$. Assume that for $p = 2, 3, ..., 6$, the joint and marginal densities satisfy

$$\sum_{\substack{t_1,t_2,...,t_p=1 \\ t_1 \neq t_2 \neq ... \neq t_p}}^{n} \int \cdots \int \left| f_{t_1,t_2,...,t_p}(z_1, z_2, ..., z_p) - \prod_{i=1}^{p} f(z_i) \right| dz_1 ... dz_p = O(n^{p-1}). \tag{2.26}$$

In addition, we introduce the following technical assumptions that are useful for the proofs.

1. $\max\limits_{t_1,...,t_p} \int \frac{f_{t_1,...,t_p}(z,...,z)}{f(z)^2} dz < \infty$, for $p = 2, 3$.
2. $\max\limits_{t_1,...,t_p} \int \frac{\zeta(z)^2 f_{t_1,...,t_p}(z,...,z)}{f(z)^4} dz < \infty$, for $p = 3, 5$.
3. $\max\limits_{t_1,...,t_6} \left| \int \frac{\zeta(z_1)\zeta(z_2) f^{(4)(3,4,5,6)}(z_1,z_2,z_1,z_2,z_1,z_2)}{f(z_1)^2 f(z_2)^2} dz_1 dz_2 \right| < \infty$.
4. $\sum_{t_1 \neq ... \neq t_6} \iint \frac{\zeta(z_1)\zeta(z_2)}{f(z_1)^2 f(z_2)^2} |f_{t_1,...,t_6}(z_1, z_1, z_1, z_2, z_2, z_2) - f(z_1)^3 f(z_2)^3| dz_1 dz_2 = O(n^5)$.

5. $|\int \zeta(z) f'(z) dz| < \infty$.

6. When $p = 6$, we have

$$\max_{t_5, t_6} \sum_{\substack{t_1, t_2, \ldots, t_4 = 1 \\ t_1 \neq t_2 \neq \ldots \neq t_4}}^{n} \iint \left| f_{t_1, t_2, \ldots, t_6}(z_1, z_1, z_1, z_2, z_2, z_2) - f(z_1)^3 f(z_2)^3 \right| dz_1 dz_2 = o(n^4).$$

7. The partial derivatives satisfies

$$\max_{t_1, t_2, \ldots, t_p} \iint \left| f_{t_1, t_2, \ldots, t_p}^{(p)(q_1, q_2, \ldots q_p)}(z_1, \ldots, z_1, z_2, \ldots, z_2) \right| dz_1 dz_2 < \infty,$$

for $p \leq 6$ and $1 \leq q_1 < q_2 < \ldots < q_p \leq p$.

It is a special situation that v_t is unobserved in the nonparametric component $\lambda(v_t)$ and it works as the smoothing variable in the nonparametric estimation when computing the conditional expectations. Therefore, instead of assuming that v_t is a stationary α-mixing process as u_t, we place restrictions on its marginal and joint probability density functions for the convenience of proving the Theorems.

The restriction in Assumption 2.3.4 is reasonable for the weakly dependent time series v_t as the joint probability density converges to the product of marginal densities when the distances between the time indexes become sufficiently large, i.e., they are asymptotically independent.

The estimation process has two kernel functions and bandwidths involved. When we estimate the conditional expectations using the smoothing variable v_t, for example, $\widehat{\mathrm{E}}_h[x_t|v_t]$ and $\widehat{\mathrm{E}}_h[y_t|v_t]$, the kernel function is denoted as $K_1(\cdot)$ with bandwidth h. When we estimate the nonparametric time trend $\widehat{g}(\tau_t)$, the kernel function is represented by $K_2(\cdot)$ with bandwidth b. For the kernel functions and bandwidths, we have the following assumptions.

Assumption 2.3.5. $K_1(\cdot)$ and $K_2(\cdot)$ are symmetric and continuous kernel functions. We require that $\int K_i(u) du = 1, \int u K_i(u) du = 0, \int K_i^2(u) du < \infty$, $\int u u' K_i(u) du < \infty$, for $i = 1, 2$.

Assumption 2.3.6. As $n \to \infty$, the bandwidths h and b satisfy $h \to 0, b \to 0, nh^{2k} \to \infty, nb^2 \to \infty, nh^{5k} < \infty, nb^5 < \infty$, and $b/h^k \to 0$.

The conditions in Assumption 2.3.5 can be easily satisfied. For example, the Epanechnikov kernel function $K(u) = 0.75(1-u^2)\mathbf{1}_{(|u|\leq 1)}$, or the Gaussian Kernel $K(u) = e^{-\frac{u^2}{2}}/\sqrt{2\pi}$. The conditions in Assumption 2.3.6 are reasonable when the bandwidth is selected as the usual optimal bandwidth, which has the same order as $O_p(n^{-1/5})$. Meanwhile, bandwidth b should converge to 0 faster than h^k to ensure the consistency of the estimators.

2.3.2 Asymptotic results

A special feature in our estimation of the semi-parametric partially linear model is that the smoothing variable v_t in the nonlinear component is unobservable. It is therefore replaced by its estimated value $\widehat{v_t}$ in the estimator. Hence, in all the analysis below, we follow a two-step procedure by first considering the properties of the infeasible estimator $\widetilde{\beta}$ and then address that the distance between the feasible estimator $\widehat{\beta}$ and the infeasible estimator $\widetilde{\beta}$ is a small quantity that converges to zero in probability as $n \to \infty$. Thus $\widehat{\beta}$ follows exactly the same asymptotic properties of $\widetilde{\beta}$ when the distance between them converges to 0 sufficiently fast. Recall that the infeasible estimator is defined as

$$\widetilde{\beta} = \left(\frac{1}{n}\sum_{t=1}^{n}\widetilde{x_t}\widetilde{x_t}'\right)^{-1}\left(\frac{1}{n}\sum_{t=1}^{n}\widetilde{x_t}\widetilde{y_t}\right),$$

where $\widetilde{x_t} = x_t - \widehat{\mathrm{E}}_h[x_t|v_t]$ and $\widetilde{y_t} = y_t - \widehat{\mathrm{E}}_h[y_t|v_t]$ are the modified versions of x_t and y_t. The first Theorem ensures that the estimator is well defined.

Theorem 2.3.1. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$, we have

$$\widetilde{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n}\widetilde{x_t}\widetilde{x_t}' \longrightarrow_P Q, \tag{2.27}$$

where $\widetilde{\Sigma}_n$ is defined as (2.20) and Q is defined in Assumption 2.3.1.

Remark 2.3.1. In fact, $\widehat{\Sigma}_n$ can be regarded as a generalized version of the 'variance-covariance matrix' that measures the variation in the time trends. Particularly, in the univariate case, a relatively flat time trend leads to a small value of Q, which indicates relatively insufficient information in the trending component. Thus it causes relatively large standard errors in the estimator.

Replacing $\widetilde{y}_t$ in $\widetilde{\beta}$ using (2.16), we have

$$\widetilde{\beta} - \beta = B_n + \left(\sum_{t=1}^{n} \widetilde{x}_t \widetilde{x}_t'\right)^{-1} \left(\sum_{t=1}^{n} \widetilde{x}_t \widetilde{u}_t\right), \tag{2.28}$$

where $B_n = \left(\sum_{t=1}^{n} \widetilde{x}_t \widetilde{x}_t'\right)^{-1} \left(\sum_{t=1}^{n} \widetilde{x}_t \widetilde{\lambda}(v_t)\right)$, $\widetilde{\lambda}(v_t) = \lambda(v_t) - \widehat{\mathrm{E}}_h[\lambda(v_t)|v_t]$ and $\widetilde{u}_t = u_t - \widehat{\mathrm{E}}_h[u_t|v_t]$. Rearranging the equation and multiply both sides by $\sqrt{n}$, we obtain

$$\sqrt{n}\left(\widetilde{\beta} - \beta - B_n\right) = \left(\frac{1}{n}\sum_{t=1}^{n} \widetilde{x}_t \widetilde{x}_t'\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \widetilde{x}_t \widetilde{u}_t\right). \tag{2.29}$$

Lemma 2.3.1. Let Assumptions 2.3.3 to 2.3.6 hold. As $n \to \infty$, we have

$$\left\|\sqrt{n} B_n\right\| = o_p(1). \tag{2.30}$$

This Lemma indicates that the potential bias term is negligible in the asymptotic results. It then leads to the following Theorem.

Theorem 2.3.2. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$, we have

$$\sqrt{n}\left(\widetilde{\beta} - \beta\right) \longrightarrow_D N(0, \Omega), \tag{2.31}$$

where $\Omega = Q^{-1}\Lambda_u$, and Λ_u is the long-run variance of u_t.

According to Theorem 2.3.1 and in conjunction with Lemma 2.3.1, it is straightforward to show Theorem 2.3.2. In the next step, we pass the asymptotic properties of the infeasible estimator to the feasible estimator by evaluating the distance between the two versions of estimators. We first address in the following Lemma that the distance between the information matrices is negligible.

Lemma 2.3.2. Let $\widehat{\Sigma}_n = n^{-1}\sum_{t=1}^{n}\widehat{x}_t\widehat{x}_t'$, where $\widehat{x}_t = x_t - \widehat{\mathrm{E}}_h[x_t|\widehat{v}_t]$. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$, we have

$$\left\|\widehat{\Sigma}_n - \widetilde{\Sigma}_n\right\| = o_p(1). \tag{2.32}$$

Equation (2.32) is the key intermediate result we need to prove. As the distance between $\widehat{\Sigma}_n$ and $\widetilde{\Sigma}_n$ is converging to zero in probability, we have

$$\|\widehat{\Sigma}_n - Q\| = \|\widehat{\Sigma}_n - \widetilde{\Sigma}_n + \widetilde{\Sigma}_n - Q\| \leq \|\widehat{\Sigma}_n - \widetilde{\Sigma}_n\| + \|\widetilde{\Sigma}_n - Q\| \longrightarrow_P 0, \tag{2.33}$$

which leads to the following Theorem.

Theorem 2.3.3. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$, we have

$$\widehat{\Sigma}_n = \frac{1}{n}\sum_{t=1}^{n}\widehat{x}_t\widehat{x}_t' \longrightarrow_P Q. \tag{2.34}$$

This Theorem shows that the feasible information matrix also converges to Q. We then move on to the asymptotic property of the feasible estimator. As in Theorem 2.3.2, we have shown that the infeasible estimator is asymptotically normal with convergence rate $\sqrt{n}$. The following lemma bounds the distance between the feasible and infeasible estimators.

Lemma 2.3.3. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$, we have

$$\left\|\sqrt{n}(\widehat{\beta} - \widetilde{\beta})\right\| = o_p(1). \tag{2.35}$$

Based on this Lemma, the asymptotic normality is easily obtainable as follows.

Theorem 2.3.4. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$, we have

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \longrightarrow_D N(0, \Omega), \tag{2.36}$$

where $\Omega = Q^{-1}\Lambda_u$ and Λ_u is the long-run variance of u_t.

Corollary 2.3.1. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$, we have

$$\sqrt{n}\widehat{\Omega}^{-1/2}\left(\widehat{\beta} - \beta\right) \longrightarrow_D N(0, I_k), \tag{2.37}$$

where I_k is the $k \times K$ identity matrix, $\widehat{\Omega} = \widehat{\Sigma}_n^{-1}\widehat{\Lambda}_u$, in which $\widehat{\Sigma}_n = n^{-1}\sum_{t=1}^{n}\widehat{x}_t\widehat{x}_t'$ and

$$\widehat{\Lambda}_u = \sum_{l=-p}^{p}\omega_l\widehat{\Gamma}_L(l), \tag{2.38}$$

where $\widehat{\Gamma}_L(l)$ is the l^{th} sample autocovariance of $\widehat{u}_t$ and $\widehat{u}_t = y_t - \widehat{\alpha} - x_t'\widehat{\beta} - \widehat{\lambda}(\widehat{v}_t)$. For $l = 1, 2, ..., p$, ω_l is a weight function that guarantees $\widehat{\Gamma}_L(l)$ nonnegative, p is a truncation parameter. For example, in Phillips and Perron (1988), they used $w_l = 1 - l/(p+1)$, which was first proposed in Newey and West (1987). Since both $\widehat{\Sigma}_n^{-1}$ and $\widehat{\Lambda}_u$ are consistent estimators for Q and Λ_u, $\widehat{\Omega}$ is a consistent estimator for Ω.

Remark 2.3.2. Theorem 3.4.1 shows that the estimator $\widehat{\beta}$ is $\sqrt{n}$-consistent. The convergence rate is slower than the rate in the co-integration regression with I(1) process due to the weak trend assumption. Since the weak trend $g(\tau_t)$ does not dominate the time series, the simple OLS estimator is therefore inconsistent due to endogeneity. This phenomenon is examined in the subsequent chapter of Monte Carlo simulations. Unlike the limit distribution of the coefficients in the co-integration regressions, the limit distribution of $\sqrt{n}(\widehat{\beta} - \beta)$ is Gaussian with zero mean. Therefore, it is more convenient to conduct hypothesis tests for the parameters in (1.2) than for those in the co-integration models.

2.4 Implementation steps and computational issues

Suppose that we are given the data set $\{y_t, x_{1t}, x_{2t}, ..., x_{kt}\}$ for $t = 1, 2, ..., n$ and the assumptions in the previous sections are satisfied. To obtain consistent estimates of the coefficients in (2.9), we follow the estimation steps below.

Step 1: Use nonparametric kernel methods[2] to estimate the weak trends in $x_{it} = g_i(\tau_t) + v_{it}$ for $\tau_t = t/n$ and $i = 1, 2, ..., k$, $t = 1, 2, ..., n$.

$$\widehat{g}_i(\tau_t) = \widehat{\mathrm{E}}_b[x_{it}|\tau_t]. \tag{2.39}$$

[2]In R, we apply nonparametric regression using the 'np' package developed by Jeffery Racine and Tristen Hayfield. https://cran.r-project.org/web/packages/np/np.pdf

Then, compute k sequences of residuals $\widehat{v_{it}} = x_{it} - \widehat{g_i}(\tau_t)$ for $i = 1, 2, ..., k$ respectively.

Step 2: Conditional on $\widehat{v_{1t}}, ..., \widehat{v_{kt}}$, obtain the expectations of $y_t, x_{1t}, ..., x_{kt}$ respectively.

$$\widehat{yv}_t = \widehat{\mathrm{E}}_h[y_t | \widehat{v_{1t}}, ..., \widehat{v_{kt}}], \tag{2.40}$$

$$\widehat{xv}_{it} = \widehat{\mathrm{E}}_h[x_{it} | \widehat{v_{1t}}, ..., \widehat{v_{kt}}]. \tag{2.41}$$

Note that for multivariate nonparametric regression, the kernel function is defined as the product of the kernel functions for each element, i.e.,

$$K(v_{1s}, ..., v_{ks}) = \Pi_{i=1}^{k} K\left(\frac{v_{is} - v_{it}}{h}\right). \tag{2.42}$$

Once we have the smoothed versions $\widehat{yv}_t$ and $\widehat{xv}_{it}$, compute the modified versions of the time series by

$$\widehat{y_t} = y_t - \widehat{yv}_t, \tag{2.43}$$

$$\widehat{x_{it}} = x_{it} - \widehat{xv}_{it}, \tag{2.44}$$

for $i = 1, 2, ..., k$ and $t = 1, 2, ..., n$.

Step 3: Apply the simple OLS method to estimate the coefficients of $\beta_1, ..., \beta_k$ in the linear regression model

$$\widehat{y_t} = \widehat{x_{1t}}\beta_1 + \widehat{x_{2t}}\beta_2 + ... + \widehat{x_{kt}}\beta_k + u_t. \tag{2.45}$$

Therefore,

$$\widehat{\beta} = (\widehat{\beta_1}, ..., \widehat{\beta_k})' = \left(\sum_{t=1}^{n} \widehat{x_t}\widehat{x_t'}\right)^{-1} \left(\sum_{t=1}^{n} \widehat{x_t}\widehat{y_t}\right). \tag{2.46}$$

Note that the intercept term is removed in the above regression. Once the estimated values of $\beta_1, ..., \beta_k$ have been obtained, the intercept term can be estimated by

$$\widehat{\alpha} = \frac{1}{n}\sum_{t=1}^{n}(y_t - x_{1t}\widehat{\beta_1} - ... - x_{kt}\widehat{\beta_k}). \tag{2.47}$$

Step 4: Compute the residual sequence $\widehat{e_t} = y_t - \widehat{\alpha} - x_{1t}\widehat{\beta_1} - ... - x_{kt}\widehat{\beta_k}$. Since the endogenous correlation is defined as $e_t = \lambda(v_{1t}, ..., v_{kt}) + u_t$, we can then uncover the control function using nonparametric kernel methods.

$$\widehat{\lambda}(v_t) = \widehat{\mathrm{E}}_h[\widehat{e_t} | \widehat{v_{1t}}, ..., \widehat{v_{kt}}]. \tag{2.48}$$

Remark 2.4.1. The estimation process involves nonparametric kernel estimation that we need to carefully select bandwidths of h and b. The selection of the bandwidths is the trade-off between the bias and the variance of the nonparametric estimates. In the literature, various bandwidth selection methods have been developed, such as the *rule-of-thumb and plug-in* method, the *cross-validation (CV)* method, and the *AIC type* methods, see Härdle and Vieu (1992), Fan and Gijbels (1996), Fan and Yao (2003), Li and Racine (2007), Cai (2007). In this thesis, we apply the cross-validation method based on a grid search procedure. The optimal bandwidth of h for $\widehat{\mathrm{E}}_h[x_t|v_t]$ is the one that minimizes the objective function

$$h_{opt} = \arg\min_{h} \sum_{t=1}^{n} \Big(x_t - \widehat{x}_{-1}(v_t, h)\Big)^2, \tag{2.49}$$

where $\widehat{x}_{-1}(v_t, h)$ is the leave-one-out kernel estimator of $\widehat{\mathrm{E}}_h[x_t|v_t]$. While for the optimal bandwidth b for the time trend estimation in $\widehat{\mathrm{E}}_b[x_t|\tau_t]$, as the error terms in (1.3) are allowed to be weakly dependent, we should apply a modified version of the cross-validation method by removing $2r+1$ data points around x_t, i.e., we remove the data points from x_{t-r} to x_{t+r} to ensure that the autocorrelation in v_t does not affect the selection of b. Therefore, the optimal bandwidth b_{opt} is selected by

$$b_{opt} = \arg\min_{b} \sum_{t=1}^{n} \Big(x_t - \widehat{g}_{-r}(\tau_t, b)\Big)^2, \tag{2.50}$$

and $\widehat{g}_{-r}(\tau_t, b)$ is the leave-(2r+1)-out estimator of $\widehat{\mathrm{E}}_b[x_t|\tau_t]$. Note that when the error terms are i.i.d., $r = 0$ is sufficient to eliminate the information at time t for cross-validation, i.e., the usual leave-one-out cross-validation method.

Chapter 3

Endogeneity in the strong trending regression

3.1 Strong trends and the OLS estimation

We noticed that all the trends have their orders of magnitudes. Recall the data generating process of x_t,

$$x_t = g(t) + v_t,$$

where $g(t)$ is the time trend and v_t is the stationary error term. We use d to denote the order of magnitude of the sum of squared trend. Specifically,

$$\frac{1}{n^d}\sum_{t=1}^{n} g(t)^2 \longrightarrow_P C, \tag{3.1}$$

for some $0 < C < \infty$ and $d \geq 1$.

In Chapter 2, we have already discussed the linear regression model where all the trends are weak ($d = 1$). The weak trend does not dominate the time series. Therefore, the endogenous correlation induces bias and inconsistency in the simple OLS estimator.

In this Chapter, we consider the linear regression model where all the nonstationary time series contain strong trends, i.e., $d > 1$ for all the trending time series. Thus the trend dominates the time series, and we show later that the simple OLS estimator is consistent. However, the limit distribution may not be centered around zero. Therefore, the OLS method provides consistent estimates, but the inference conclusions are not reliable due to the endogeneity problem.

3.2 Identification and assumptions

As we only consider strong trending time series, the time trends play dominating roles over the stationary disturbances. Formally, we regulate the nonlinear and nonparametric time trends by the assumption as follows.

Assumption 3.2.1. Let $G = (g_1, g_2, ..., g_k)$ and $g_i = (g_i(1), g_i(2), ..., g_i(n))'$ for $i = 1, 2, .., k$. Assume that there exists a diagonal matrix $D = diag(n^{d_1/2}, ..., n^{d_k/2})$, such that as $n \to \infty$,

$$D^{-1} G' G D^{-1} \longrightarrow Q, \tag{3.2}$$

where $d_i > 1$ for $i = 1, 2, ..., k$ and Q is a positive definite matrix.

This assumption rules out weak trends since the parameter d_i is assumed to be greater than 1 for $i = 1, 2, ..., k$. Matrix Q is a $k \times k$ positive definite matrix with full rank, therefore, collinearity has been ruled out that none of the time trends can be represented by the rest of the time trends. It is shown later that this condition is necessary for identifying the coefficients.

Remark 3.2.1. Equation (3.2) simply implies that

$$\frac{1}{n^{d_{ij}}} \sum_{t=1}^{n} g_i(t) g_j(t) \longrightarrow Q_{ij}, \tag{3.3}$$

where $d_{ij} = (d_i + d_j)/2$, for $i, j = 1, 2, ..., k$.

For the purpose of nonparametric estimation, given the sample size n, it is equivalent to define a rescaled time trend $\widetilde{g}_i(\tau_t)$ that

$$\frac{1}{n^{d_{ij}}} \sum_{t=1}^{n} g_i(t) g_j(t) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{g_i(t)}{n^{\frac{d_i - 1}{2}}} \right) \left(\frac{g_j(t)}{n^{\frac{d_j - 1}{2}}} \right) = \frac{1}{n} \sum_{t=1}^{n} \widetilde{g}_i(\tau_t) \widetilde{g}_j(\tau_t) \longrightarrow Q_{ij}, \tag{3.4}$$

where $\widetilde{g}_i(\tau_t) = n^{-\frac{d_i-1}{2}} g_i(t)$ for $i, j = 1, 2, ..., k$.

Assumption 3.2.2. Let $(\epsilon_t, \eta_t')' = (\epsilon_t, \eta_{1t}, ..., \eta_{kt})'$ be a $k+1$ vector of i.i.d. innovations with mean 0 and $\sigma_1^2 = \mathrm{E}[\epsilon_1^2]$, $\mathbf{\Theta} = (\theta_1, ..., \theta_k)'$, $\theta_i = Cov(\epsilon_t, \eta_{it})$ and Σ_{22} is the $k \times k$ variance-covariance matrix of $\boldsymbol{\eta}_t = (\eta_{1t}, ..., \eta_{kt})'$ with σ_{ij} being the element at the i^{th} row and j^{th} column, i.e., $\sigma_{ij} = \mathrm{E}[\eta_{it}\eta_{jt}]$. Meanwhile, $\mathrm{E}[\epsilon_t^4] < \infty$, $\mathrm{E}[\eta_{it}^4] < \infty$ and $\mathrm{E}[\epsilon_t^4 \eta_{it}^4] < \infty$, for $i, j = 1, 2, ..., k$.

Assumption 3.2.3. The error terms are defined as linear processes with respect to the sequences of innovations defined above. Specifically,

$$e_t = \sum_{s=0}^{\infty} \phi_s \epsilon_{t-s} \triangleq \Phi(L)\epsilon_t, \tag{3.5}$$

$$v_t = \sum_{s=0}^{\infty} \boldsymbol{\psi}_s \boldsymbol{\eta}_{t-s} \triangleq \mathbf{\Psi}(L)\boldsymbol{\eta}_t, \tag{3.6}$$

where $\boldsymbol{\psi}_s = diag(\psi_{s,1}, ..., \psi_{s,k})$ is a $k \times k$ diagonal matrix. The coefficients satisfy $\sum_{s=0}^{\infty} \phi_s^2 < \infty$ and $\sum_{s=0}^{\infty} \psi_{s,i}^2 < \infty$, for $i = 1, 2, ..., k$.

Since $\boldsymbol{\psi}_s$ is a diagonal matrix, for each element of $\boldsymbol{v}_t$, we can write

$$v_{it} = \Psi_i(L)\eta_{i,t} = \sum_{s=0}^{\infty} \psi_{s,i}\eta_{i,t-s}. \tag{3.7}$$

Meanwhile, we define

$$f_{i,q}(L) = \sum_{s=0}^{\infty} \phi_s \psi_{s+q,i} L^s, \tag{3.8}$$

$$m_{i,q}(L) = \sum_{s=0}^{\infty} \phi_{s+q} \psi_{s,i} L^s, \tag{3.9}$$

where L is the *lag-operator*, for example, $L^i x_t = x_{t-i}$. Therefore, by the Beveridge-Nelson Decomposition, we have

$$\Phi(L) = \Phi(1) - (1-L)\widetilde{\Phi}(L), \tag{3.10}$$

$$\Psi_i(L) = \Psi(1) - (1-L)\widetilde{\Psi}_i(L), \tag{3.11}$$

$$f_{i,q}(L) = f_{i,q}(1) - (1-L)\widetilde{f}_{i,q}(L), \tag{3.12}$$

$$m_{i,q}(L) = m_{i,q}(1) - (1-L)\widetilde{m}_{i,q}(L), \tag{3.13}$$

where $\widetilde{\Phi}(L) = \sum_{s=0}^{\infty} \widetilde{\phi}_s L^s$, $\widetilde{\Psi}_i(L) = \sum_{s=0}^{\infty} \widetilde{\psi}_{s,i} L^s$, $\widetilde{f}_{i,q}(L) = \sum_{s=0}^{\infty} \widetilde{f}_{i,qs} L^s$, $\widetilde{m}_{i,q}(L) = \sum_{s=0}^{\infty} \widetilde{m}_{i,qs} L^s$, in which $\widetilde{\phi}_s = \sum_{p=s+1}^{\infty} \phi_p$, $\widetilde{\psi}_{s,i} = \sum_{p=s+1}^{\infty} \psi_{p,i}$, $\widetilde{f}_{i,qs} = \sum_{p=s+1}^{\infty} \phi_p \psi_{p+q,i}$, $\widetilde{m}_{i,qs} = \sum_{p=s+1}^{\infty} \phi_{p+q} \psi_{p,i}$.

Remark 3.2.2. In fact, Assumption 3.2.3 is a special case of the following linear process. Assume that $u_t = (e_t, v_t')$ and

$$u_t = \sum_{s=0}^{\infty} \gamma_s \mu_{t-s} \triangleq \Gamma(L)\mu_t, \tag{3.14}$$

where μ_t is an i.i.d. process with mean $\mathbf{0}$ and variance-covariance matrix Σ_μ.

Assumption 3.2.4. In addition, we propose the following conditions that are necessary for establishing the asymptotic results. Let H represent f or m in the BN decomposition and we assume

1. $\sum_{s=0}^{\infty} \widetilde{f}_{i,0s}^2 < \infty$.
2. $\sum_{q=1}^{\infty} \sum_{s=0}^{\infty} \widetilde{H}_{i,qs}^2 < \infty$.
3. $|\sum_{q_1=1}^{\infty} H_{i,q_1}(1) H_{j,q_1}(1)| < \infty$.
4. $|\sum_{p=1}^{n-1} \sum_{q_1=1}^{\infty} H_{i,q_1}(1) H_{j,q_1}(1) H_{i,p+q_1}(1) H_{j,p+q_1}(1)| < \infty$.
5. $\sum_{q_1=1}^{\infty} H_{i,q_1}(1)^2 H_{j,q_1}(1)^2 < \infty$.
6. $|\sum_{p=1}^{n-1} \sum_{q_1=1}^{\infty} \sum_{\substack{q_2=1 \\ q_2 \neq q_1}}^{\infty} H_{i,q_1}(1) H_{j,q_2}(1) H_{i,p+q_1}(1) H_{j,p+q_2}(1)| < \infty$.
7. $\sum_{q=1}^{\infty} H_{i,q}(1)^2 < \infty$.
8. $\lim_{n\to\infty} n^{1-d_j} \int_0^1 \int_{\tau_1}^1 \widetilde{g}_i(\tau_1)\widetilde{g}_i(\tau_2)\gamma(n(\tau_2-\tau_1), j) d\tau_1 d\tau_2 = 0$, where $\gamma(d,j) = \sum_{q_1=1}^{\infty} H_{j,q_1}(1) H_{j,d+q_1}(1)$.
9. $|\sum_{p=1}^{n-1} \sum_{q_1=1}^{\infty} H_{j,q_1}(1) H_{j,p+q_1}(1)| < \infty$.
10. $\sum_{q=1}^{\infty} H_{i,q}(1)^4 < \infty$.

11. $\sum_{q_1=1}^{\infty} \sum_{q_2=q_1+1}^{\infty} H_{i,q_1}(1)^2 H_{i,q_2}(1)^2 < \infty$.

These assumptions can be easily satisfied, particularly when the error term follow a stationary AR(p) or MA(q) process. For example, in condition 8, $\gamma(d,j)$ is a finite value when the time series follows an AR(1) process, hence the limit converges to 0 for $d_j > 1$. When $\mathbf{\Theta} \neq 0$, the innovations ϵ_t and η_t are correlated and therefore, the error terms v_t and e_t are correlated. Then it causes the problem of endogeneity in the regression.

We define the sample information matrix as $\widehat{Q} = D^{-1}\boldsymbol{X}'\boldsymbol{X}D^{-1}$. The strong trend condition indicates that the stationary disturbances in the regressors are dominated. Therefore, they can be ignored in the information matrix.

Theorem 3.2.1. Under Assumptions 3.2.1 to 3.2.4, as $n \to \infty$, we have

$$\widehat{Q} = D^{-1}\boldsymbol{X}'\boldsymbol{X}D^{-1} \longrightarrow_P Q, \tag{3.15}$$

where Q is a positive definite matrix defined in Assumption 3.2.1.

This Theorem shows that the variation in the time trends plays a central role in the identification and estimation of β. In the next section, we discuss the performance of the simple OLS estimator given that all the trends are strong, i.e., $d_i > 1$ for all i.

3.3 The simple OLS estimator

In this section, we investigate the asymptotic properties of the simple OLS estimator, which is defined as

$$\widehat{\beta}_{ols} = \left(\sum_{t=1}^{n} x_t x_t' \right)^{-1} \left(\sum_{t=1}^{n} x_t y_t \right). \tag{3.16}$$

In matrix form, it can be written as

$$\widehat{\beta}_{ols} = (\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'y). \tag{3.17}$$

In Theorem 3.2.1, we have shown that $D^{-1}\boldsymbol{X}'\boldsymbol{X}D^{-1} \longrightarrow_P Q$ as $n \to \infty$, in which Q is a positive definite matrix. Therefore, $D^{-1}\boldsymbol{X}'\boldsymbol{X}D^{-1}$ is invertible and

$$D\left(\widehat{\beta}_{ols} - \beta\right) = \left(D^{-1}\boldsymbol{X}'\boldsymbol{X}D^{-1}\right)^{-1} \left(D^{-1}\boldsymbol{X}'e\right). \tag{3.18}$$

Obviously, the convergence rates of the coefficients depend on the strength of the trends respectively. We show that

Theorem 3.3.1. Under Assumptions 3.2.1 to 3.2.4, as $n \to \infty$,

$$D\left(\widehat{\beta}_{ols} - \beta - D^{-1}\widehat{Q}^{-1}D^{-1}nb\right) \longrightarrow_D \mathcal{N}(0, Q^{-1}\Omega Q^{-1}), \tag{3.19}$$

where $b = \mathrm{E}[e_t v_t] = \left(\sum_{j=0}^{\infty}\phi_j \boldsymbol{\Psi}_j\right)\boldsymbol{\Theta}$, Ω is the asymptotic variance-covariance matrix of $D^{-1}\boldsymbol{X}'e$. Specifically, if $d_i = d_j = 1$, and let $\delta_{2j} = \mathrm{E}[\epsilon_t^2\eta_{jt}]$ and $\delta_{2ij} = \mathrm{E}[\epsilon_t^2\eta_{it}\eta_{jt}]$. We have

$$\begin{aligned}\Omega_{ij} =& \sigma_1^2\Phi(1)^2\mathbf{Q}_{ij} + f_{i,0}(1)f_{j,0}(1)(\delta_{2ij} - \theta_i\theta_j) + \Phi(1)f_{j,0}(1)\delta_{2j}\overline{g}_i + \Phi(1)f_{i,0}(1)\delta_{2i}\overline{g}_j \\ &+ \sigma_1^2\sigma_{ij}\sum_{q_1=1}^{\infty} f_{i,q_1}(1)f_{j,q_1}(1) + \sigma_1^2\sigma_{ij}\sum_{q_1=1}^{\infty} m_{i,q_1}(1)m_{j,q_1}(1) \\ &+ \theta_j\theta_i\sum_{q=1}^{\infty} f_{i,q}(1)m_{j,q}(1) + \theta_i\theta_j\sum_{q=1}^{\infty} f_{j,q}(1)m_{i,q}(1),\end{aligned} \tag{3.20}$$

where $\overline{g}_i = \int_0^1 g_i(\tau)d\tau$. For $d_i > 1$ but $d_j = 1$,

$$\Omega_{ij} = \sigma_1^2\Phi(1)^2Q_{ij} + \Phi(1)f_{j,0}(1)\delta_{2j}\overline{g}_i. \tag{3.21}$$

Finally, when $d_i > 1$ and $d_j > 1$,

$$\Omega_{ij} = \sigma_1^2\Phi(1)^2Q_{ij}. \tag{3.22}$$

Note that the endogenous correlation between the error terms induces bias in the simple OLS estimator. As the strong trends dominate the time series, the endogeneity bias diminishes to zero when $n \to \infty$. To investigate the bias and the asymptotic distribution under different orders of trends, we study the univariate regression and introduce the following Corollary.

Corollary 3.3.1. Under Assumptions 3.2.1 to 3.2.4, let $k = 1$. As $n \to \infty$, we have

$$\sqrt{n^d}\left(\widehat{\beta}_{ols} - \beta - B_n\right) \to_D \mathcal{N}(0, Q^{-1}\Omega Q^{-1}), \tag{3.23}$$

where $B_n = n^{1-d}\widehat{Q}b$.

Also note that when $d = 1$, $B_n = \widehat{Q}b$, and it does not vanish when the sample size tends to infinity. Therefore, the simple OLS estimator is biased and inconsistent. While in the case of $d > 1$, the bias term is negligible that $B_n = o_p(1)$ and the simple OLS estimator is super-consistent when $n \to \infty$. The speed of convergence is faster than the usual $\sqrt{n}$ rate because the trend goes to infinity when the sample size tends to infinity. Hence, in terms of estimation, the endogeneity does not affect the consistency of the estimators, though in finite sample, the bias can hardly be neglected. However, inference is substantially affected by endogeneity since $\sqrt{n^d}B_n = n^{1-\frac{d}{2}}\widehat{Q}b$ is not decaying with sample size n when $1 < d \leq 2$. Even when $d > 2$, the limit distribution can also be distorted significantly in finite sample. We formally address these properties as follows.

Corollary 3.3.2. As $n \to \infty$, when $d = 1$, the OLS estimator is inconsistent. However, when $d > 1$, the OLS estimator is consistent. Specifically

- For $d = 1$, the estimator is inconsistent with bias Qb. Hence, we can not apply the bias-correction method. [1]

- For $1 < d < 2$, the estimator is biased but consistent, and the bias diminishes to zero at the rate of n^{1-d}.

- For $d = 2$, the estimator is super-consistent. However, $n(\widehat{\beta}_{ols} - \beta)$ converges to a distribution that is not centered around zero. Specifically, the limit distribution becomes

$$n(\widehat{\beta}_{ols} - \beta) \longrightarrow_D \mathcal{N}(B, Q^{-1}\Omega Q^{-1}), \tag{3.24}$$

  where $B = Q^{-1}b$ is a non-zero constant when there exists endogeneity.

- For $d > 2$, the potential bias term satisfies $\sqrt{n^d}B_n = o_P(1)$, which is negligible to the limit distribution that is always regarded as $O_P(1)$. Therefore, the simple OLS

[1] This is the weak trend condition, and we should use the control function approach introduced in Chapter 2.

estimator is unbiased and consistent.

$$\sqrt{n^d}(\widehat{\beta}_{ols} - \beta) \longrightarrow_D \mathcal{N}(0, Q^{-1}\Omega Q^{-1}), \tag{3.25}$$

In the next section, we propose a bias-corrected estimator to adjust for the bias in the simple OLS estimator when $d_i > 1$ for $i = 1, 2, ..., k$, without the need to know the exact order of the trending components.

3.4 The bias-corrected estimator

Since the simple OLS estimator is always consistent when $d > 1$, the bias in the OLS estimator can therefore be estimated consistently. We propose a bias-corrected estimator as follows.

$$\widehat{\beta}_{bc} = \widehat{\beta}_{ols} - \widehat{Bias} = \widehat{\beta}_{ols} - \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \sum_{t=1}^{n} \widehat{v_t}\widehat{e_t}, \tag{3.26}$$

where $\widehat{v_t} = x_t - \widehat{g}(\tau_t), \widehat{e_t} = y_t - x_t'\widehat{\beta}_{ols}$, and $\widehat{g}(\tau_t) = (\widehat{g_1}(\tau_t), ..., \widehat{g_k}(\tau_t))'$ that $\widehat{g_i}(\tau_t)$ is the non-parametric estimates[2] of the time trend $\widetilde{g_i}(\tau_t)$.

Remark 3.4.1. Fortunately, we do not need to know the value of d_i when we do the bias-correction for $\widehat{\beta}_{ols}$ as d_i is not involved in the estimator of the bias term.

Theorem 3.4.1. Under Assumptions 3.2.1 to 3.2.4, the bias-corrected estimator is unbiased and consistent.

$$D\left(\widehat{\beta}_{bc} - \beta\right) \longrightarrow_D \mathcal{N}(0, Q^{-1}\Omega Q^{-1}). \tag{3.27}$$

The availability of the bias-correction method depends on the consistency of the first-stage OLS estimator and this condition is similar to that in Phillips and Hansen (1990). In fact, we show that

$$\widehat{\beta}_{bc} - \beta = \widehat{\beta}_{ols} - \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} n\widehat{b} - \beta$$

[2]It is not a problem to write $\widehat{g}(\tau_t)$ instead of $\widehat{g}(t)$ because for estimation purposes, they are equivalent.

$$
\begin{aligned}
&= \underbrace{\widehat{\beta}_{ols} - \left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} nb - \beta}_{} + \underbrace{\left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} n(b - \widehat{b})}_{} \\
&= S_1(n) + B_2(n),
\end{aligned} \tag{3.28}
$$

where $S_1(n)$ converges to the limit distribution in (3.27) according to (3.19), while $B_2(n)$ is the potential bias term that $B_2(n) = O_p(n^{\frac{1}{2}-d})$ and $\sqrt{n^d}B_2(n) = O_p(n^{\frac{1-d}{2}})$, which is always $o_p(1)$ when $d > 1$. Hence regardless of the value of d, the endogeneity bias can always be ignored in the bias-corrected estimator $\widehat{\beta}_{bc}$ and the asymptotic distribution is always centered around zero when $n \to \infty$.

3.5 Estimation of the trending parameter

Although we do not need to know d_i in the bias-correction procedure, we do need to approximate its value when we estimate the variance-covariance matrix. An imperfect way of estimating d_i is introduced as follows.

Note that as $\widehat{Q}_i$ converges in probability to a constant value Q_i, and

$$\sum_{t=1}^{n} x_{it}^2 = Q_i n^{d_i}. \tag{3.29}$$

We take the logarithms of both sides, and yield

$$\log \sum_{t=1}^{n} x_{it}^2 = \log Q_i + d_i \log n. \tag{3.30}$$

Therefore,

$$d_i = \frac{\log \sum_{t=1}^{n} x_{it}^2}{\log n} + \underbrace{\frac{\log Q_i}{\log n}}_{o_p(1)}. \tag{3.31}$$

Since $1/\log n \to 0$ as $n \to \infty$, we define a consistent estimator for d_i as

$$\widehat{d_i} = \frac{\log \sum_{t=1}^{n} x_{it}^2}{\log n}, \tag{3.32}$$

for $i = 1, 2, ..., k$.

Remark 3.5.1. The problem in the estimator of d_i is that $\log n$ goes to infinity very slowly with n, so that the second term in (3.31) can be quite large in finite sample, resulting a relatively quite large bias in $\widehat{d_i}$.

Chapter 4

Numerical evidence

4.1 Overview

In this chapter, we first present some simulated examples to compare the OLS estimator and the proposed estimator in this thesis. In the weak trending regression case, we find the presence of persistent biases in the simple OLS estimator due to the problem of endogeneity. By using the control function approach, the estimators for the coefficients in the augmented semiparametric partially linear model become unbiased and consistent. While in the strong trending regression case, the simple OLS estimator itself is consistent when the sample size tends to infinity. In the finite sample case, however, the bias can be substantially large for the OLS estimator. On the other hand, the bias-corrected estimators are unbiased and consistent as they significantly reduce the bias.

We then show an empirical example to demonstrate the applicability of the models as well as the estimation procedures. We consider the linear regression of the logarithm of aggregate personal consumption on the logarithm of the aggregate personal disposable income and the real interest rate. The result reveals how personal consumption reacts to the changes in personal disposable income and real interest rate. We also find a nonlinear relationship between the error terms that induces the problem of endogeneity.

4.2 Simulated examples

4.2.1 The weak trending regression model

In this subsection, consider the time series data generated from the trending regression equations as

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + e_t, \tag{4.1}$$

$$x_{1t} = g_1(\tau_t) + v_{1t}, \tag{4.2}$$

$$x_{2t} = g_2(\tau_t) + v_{2t}, \tag{4.3}$$

for $t = 1, 2, ..., n$, and $\tau_t = t/n$. In the process, we let $\beta_1 = 0.7, \beta_2 = 0.5$. The time trends are bounded weak trends $g_1(\tau_t) = 3 - 4(\tau_t - 0.5)^2$ and $g_2(\tau_t) = 2 + 0.7\sin(2\pi\tau_t)$. The error term e_t is correlated with v_{1t} and v_{2t} that $e_t = 1.5v_{1t} + v_{2t} + u_t$. Meanwhile, v_{1t}, v_{2t} and u_t follow stationary AR(1) processes $v_{it} = 0.2v_{i,t-1} + \eta_{it}$ and $u_t = 0.2u_{t-1} + \epsilon_t$ where $\epsilon_t, \eta_{it} \overset{i.i.d.}{\sim} N(0, 0.2^2)$ for $i = 1, 2$.

We examine the performances of the simple OLS estimator for β_1 and β_2 in (4.1) and the proposed estimator (2.23) for the semiparametric partially linear model

$$y_t = x_{1t}\beta_1 + x_{2t}\beta_2 + \lambda(v_{1t}, v_{2t}) + u_t, \tag{4.4}$$

where $\lambda(\cdot)$ is an unknown nonparametric control function, u_t is the error term independent with $x_{1t}, x_{2t}, v_{1t}, v_{2t}$.

The time series are simulated independently for $N_B = 5,000$ times, and the estimation procedures are conducted each time. We denote the OLS estimator and the estimator in equation (2.23) as $\widehat{\beta}_{i,p}^{ols}$ and $\widehat{\beta}_{i,p}^{control}$ respectively for $i = 1, 2$ and $p = 1, 2, ..., N_B$. The sample sizes are chosen as 250, 600 and 1000 respectively.

To show the properties of the two estimators, we calculate the averages of the biases, the standard deviations as well as the root mean squared errors for the two estimators in Table 4.1. Specifically, they are computed by the formulas below.

$$Bias_i = \frac{1}{N_B}\sum_{p=1}^{N_B}\widehat{\beta}_{i,p} - \beta, \tag{4.5}$$

$$Std_i = \sqrt{\frac{1}{N_B - 1}\sum_{p=1}^{N_B}\left(\widehat{\beta}_{i,p} - \frac{1}{N_B}\sum_{p=1}^{N_B}\widehat{\beta}_{i,p}\right)^2}, \tag{4.6}$$

$$RMSE_i = \sqrt{Bias_i^2 + Std_i^2}, \tag{4.7}$$

for $i = 1,2$, and $\widehat{\beta}_{i,p}$ is replaced by $\widehat{\beta}_{i,p}^{ols}$ and $\widehat{\beta}_{i,p}^{control}$ respectively for the two estimators.

Table 4.1: Simulation results for the weak trending regression with endogeneity.

		OLS			Control function		
	n	250	600	1000	250	600	1000
$\widehat{\beta}_1$	Bias	0.4776	0.4755	0.4787	0.0105	0.0088	0.0028
	Std	0.0663	0.0430	0.0340	0.1132	0.0764	0.0577
	RMSE	0.4822	0.4775	0.4799	0.1137	0.0769	0.0578
$\widehat{\beta}_2$	Bias	0.1420	0.1446	0.1462	-0.0324	-0.0307	-0.0298
	Std	0.0506	0.0307	0.0242	0.0719	0.0428	0.0343
	RMSE	0.1507	0.1479	0.1482	0.0789	0.0527	0.0454

According to the discussion in Chapter 2, since the weak trends do not dominate the stationary error terms, endogeneity causes persistent biases in the simple OLS estimators for the coefficients. Therefore, as expected, a non-diminishing positive bias is seen in the simple OLS estimates of β_1 (≈ 0.47) and β_2(≈ 0.14). This result reconciles with the theoretical conclusion that the OLS estimators are inconsistent in the weak trending regression with endogeneity.

On the other hand, by applying the control function approach, we fix the problem of endogeneity, and the control function extends the linear regression model to a semiparametric partially linear model as (4.4). In Chapter 2, we have shown that the proposed estimators for β_1 and β_2 converge to their true values consistently. The biases are negligible, and the standard deviations decrease at the rate of $1/\sqrt{n}$ as $n \to \infty$. Therefore, the control function approach successfully adjusts for the endogeneity bias and yields unbiased and consistent estimates of the coefficients.

4.2.2 The strong trending regression model

In this subsection, we investigate the behaviors of the simple OLS estimator and the bias-corrected estimator in the strong trending regression models with endogeneity. We also show the improvement in terms of statistical inference when the endogeneity bias has been adjusted. We consider a univariate regression

$$y_t = x_t\beta + e_t, \tag{4.8}$$

$$x_t = g(t) + v_t, \tag{4.9}$$

where we let $\beta = 0.5$ and consider two kinds of trends in the regressors

- Example 1: $g(t) = 0.1\sqrt{t}$;
- Example 2: $g(t) = 0.01t$.

In the first example, as $t \to \infty$, the trend term goes to infinity with diminishing speed. The magnitude parameter $d = 2$ and according to our previous discussion in Chapter 3, the OLS estimator is consistent, but the limit distribution of $n(\widehat{\beta}_{ols} - \beta)$ is not centered around zero. While in the second example, there is a linear time trend in the regressor, hence the trending parameter $d = 3$. The OLS estimator is unbiased and consistent theoretically.

The error terms e_t and v_t follow AR(1) processes as $e_t = 0.2e_{t-1} + \epsilon_t$ and $v_t = 0.2v_{t-1} + \eta_t$, where $(\epsilon_t, \eta_t)' \overset{i.i.d.}{\sim} N(0, \Sigma)$ and

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 0.3 \end{pmatrix}, \tag{4.10}$$

where the off-diagonal element is nonzero. Hence, e_t and v_t are correlated, causing endogeneity in the regression model.

We generate the data and obtain the simulation results independently for $N_B = 5,000$ times with sample size $n = 300, 600$ and 900 respectively. In each replication, we compute the estimates by using the simple OLS and the bias-correction methods. Table 4.2 shows the biases, standard deviations and the root mean squared errors for both estimators $\widehat{\beta}_p^{ols}$ and $\widehat{\beta}_p^{bc}$ under different cases of trends and sample sizes.

Table 4.2: Simulation results for the strong trending regression with endogeneity.

		OLS			Bias-correction		
	n	Bias	Std	RMSE	Bias	Std	RMSE
	300	0.2851	0.0412	0.2880	0.0581	0.0522	0.0781
Example 1	600	0.1563	0.0245	0.1582	0.0174	0.0285	0.0334
	900	0.1076	0.0171	0.1090	0.0082	0.0191	0.0208
	300	0.1541	0.0344	0.1579	0.0179	0.0403	0.0441
Example 2	600	0.0418	0.0141	0.0442	0.0017	0.0148	0.0149
	900	0.0189	0.0078	0.0204	0.0004	0.0081	0.0080

According to the main results in Chapter 3, the strong trend in x_t dominates the stationary error term v_t, and therefore the OLS estimation of the coefficient β is consistent for both examples regardless of the existence of endogeneity. As the sample size increases, a sequence of decreasing biases is seen in Table 4.2 for the OLS estimator in both examples. However, when the sample size is relatively small, the bias can still be substantially large. It is obvious in Table 4.2 that as an improvement, the bias-correction method significantly reduces the biases, and the RMSEs for the bias-corrected estimator are only a quarter of those for the simple OLS estimators.

The endogeneity issue not only affects the estimation accuracy, but also severely distorts the statistical inferences of the coefficients even for the very strong trend conditions. Suppose that we are interested in testing the hypothesis

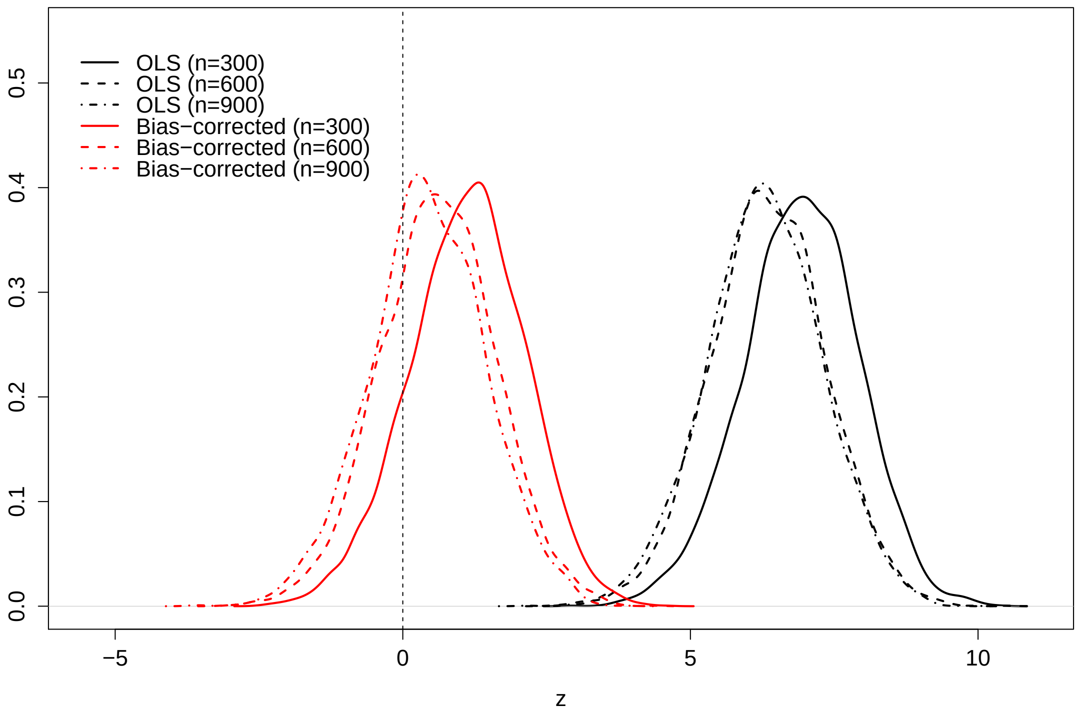
$$\mathbb{H}_0 : \beta = 0.5 \quad \text{v.s.} \quad \mathbb{H}_1 : \beta \neq 0.5. \tag{4.11}$$

In the following context, we show the nonparametrically estimated distributions of the t-statistics for the two estimators. The t statistics are computed as follows.

$$t_{i,p} = \frac{\widehat{\beta}_p - 0.5}{Std_i}, \tag{4.12}$$

for the two estimators $\widehat{\beta}_p^{ols}$ and $\widehat{\beta}_p^{bc}$ and $p = 1, 2, ..., N_B$. Figure 4.1 and 4.2 are presented

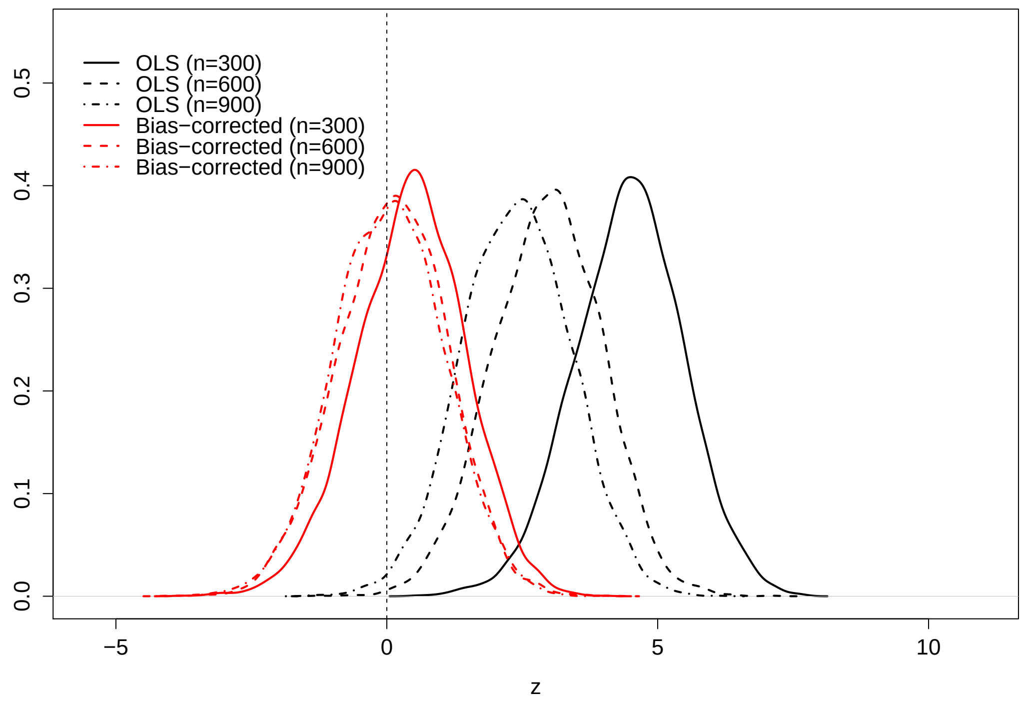
Figure 4.1: The distributions of the t-statistics for the OLS and the bias-corrected estimators when $d = 2$.



respectively for the two examples of trends. In each graph, black lines represent the estimated distributions for the t-statistics based on the OLS estimator, while the red lines shows the estimated distributions for t-distributions of the bias-corrected estimator with sample size 300 (solid line), 600 (dash line) and 900 (dot-dash line) respectively.

Figure 4.1 shows that when the order of the trending magnitude is 2, the distributions of the t-statistics for the OLS estimators are not centered around 0. In other words, the distributions are persistently biased (the black lines) regardless of the sample size. When it comes to the bias-corrected estimators, however, the means of the distributions of the t-statistics are moving towards 0 as the sample size tends to infinity. Hence, bias-correction is of critical importance for hypothesis testing as it significantly reduces the probability of making the *type-one error*, which is very high (close to 100%) for the OLS estimators.

Figure 4.2: The distributions of the t-statistics for the OLS and the bias-corrected estimators when $d = 3$.



In the second example, the order of the trending magnitude is 3 for the linear time trend. Therefore, the OLS estimator is consistent when $n \to \infty$ and the potential bias caused by endogeneity is proportional to $o_P(1)$, which is negligible to the limit distribution. Consequently, in Figure 4.2, all the centers of the t statistics distributions are moving towards 0 for both OLS and bias-corrected estimators. When the sample size is small, the mean of the t-statistics for the OLS estimator starts from the location which is quite far away from 0. Thus, the bias in the distribution of the t-statistic may not be negligible. On the other hand, in terms of unbiasedness and consistency, the bias-correction method gives a much better estimator whose distribution of the t-statistics is centered close to 0 even when the sample size is small.

Table 4.3: The probability of making the *Type I error*.

	Example 1($d = 1$)		Example 2($d = 2$)	
n	OLS	Bias-correction	OLS	Bias-correction
300	0.998	0.184	0.991	0.071
600	1.000	0.089	0.836	0.048
900	1.000	0.074	0.663	0.051

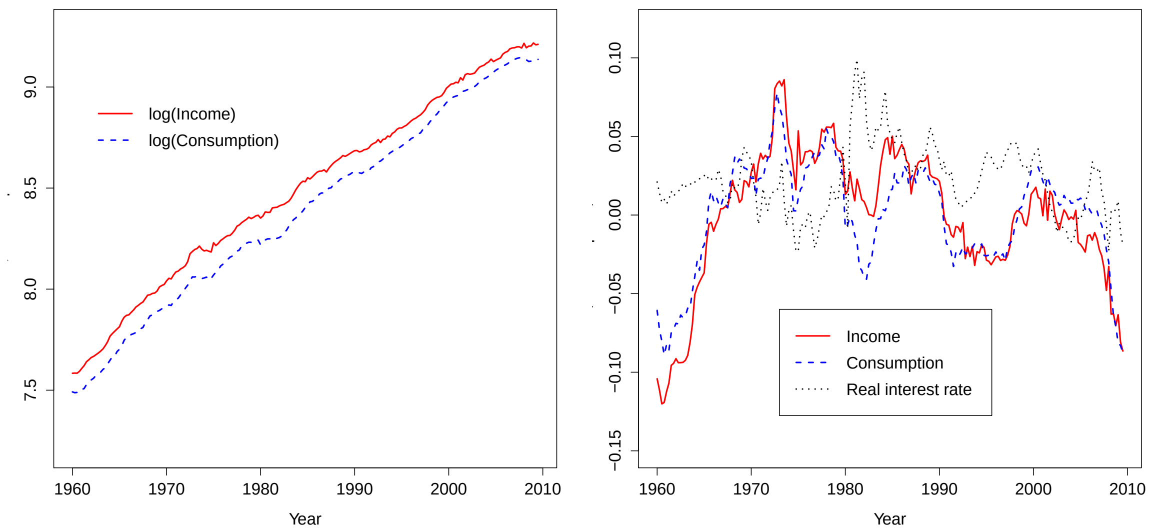
The adjustments for the endogenous bias is critical to statistical inferences of the coefficients. In the simulation, the real value of β is 0.5. Therefore, we should not reject $\mathbb{H}_0$. Table 4.3 shows the proportion of the t-statistic that is greater than the critical value[1] under 5% significance level. Due to the endogenous correlation, the sizes of the t-test based on the simple OLS estimator exhibit severe distortions in both examples. While with relatively larger sample size, the probability of making the *Type I error* converges to the normal 0.05 for the bias-corrected estimators. To summarize, even though the OLS estimator is consistent when the trending parameter $d > 3$, the inferences are not reliable when endogeneity is present in the regression model. The bias-corrected estimator performs much better in terms of estimations and inferences of the coefficients.

[1] Here we choose the 97.5% quantile of the t-distribution with degree of freedom $n-1$.

4.3 Empirical example

In this section, we explore the relationship between the quarterly data[2] of the U.S. aggregate personal disposable income, the aggregate personal consumption expenditure on non-durable goods and services and the real interest rate from 1960Q1 to 2009Q3.

Figure 4.3: The data and the removal of linear trends



(a) Log of income and consumption. (b) Data used in the regression model.

The logarithms of aggregate personal disposable income and aggregate personal consumption expenditure are plotted in Figure 4.3a. We estimate and remove the linear time trends[3] in both time series and denote the residuals as i_t and c_t for income and consumption respectively. Figure 4.3b shows the graphs of c_t and i_t as well as the real interest rate r_t. These three time series are usually believed to be pure random walk processes as the null hypothesis of unit root cannot be rejected in the ADF unit

[2]The data can be downloaded from http://www.bea.gov.

[3]We should remove the linear trends as they are caused by the average growth rates of income and consumption. However, we are interested in the deviations to such average level, which would cause nonlinear trends in the long-run.

root test. Based on the arguments in this thesis, it is also reasonable to assume that these three sequences c_t, i_t and r_t are nonlinear trend-stationary time series that

$$i_t = g_1(t) + v_{1t}, \tag{4.13}$$

$$r_t = g_2(t) + v_{2t}, \tag{4.14}$$

where $g_i(t)$ are nonparametric functions of time trends for $i = 1, 2$.

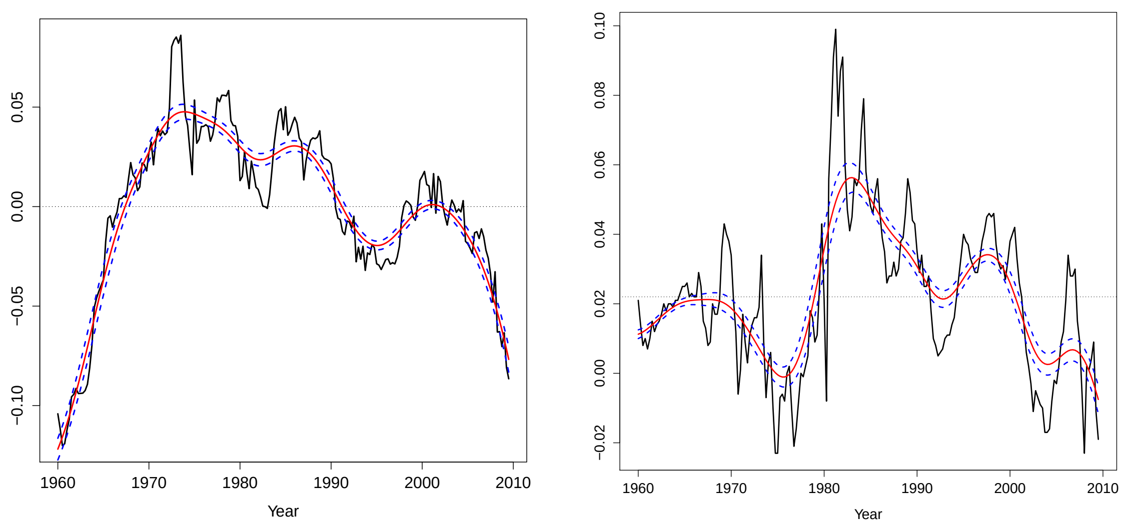
For the nonparametric estimation purposes, the nonlinear trends can be standardized on $[0, 1]$ with given sample size n. i.e., we write the DGP of the regressors as

$$i_t = g_1(\tau_t) + v_{1t}, \tag{4.15}$$

$$r_t = g_2(\tau_t) + v_{2t}, \tag{4.16}$$

where $\tau_t = t/n$. Since the trend functions are continuous and differentiable, one can estimate the trends using nonparametric kernel methods. Note that $\widehat{g}(\tau_t)$ is the estimated value of $\widehat{g_i}(t)$. In this thesis, the two functions $g_1(\tau_t)$ and $g_2(\tau_t)$ are estimated using nonparametric local linear kernel methods as in Figure 4.4.

Figure 4.4: The explanatory variables and their estimated trends.



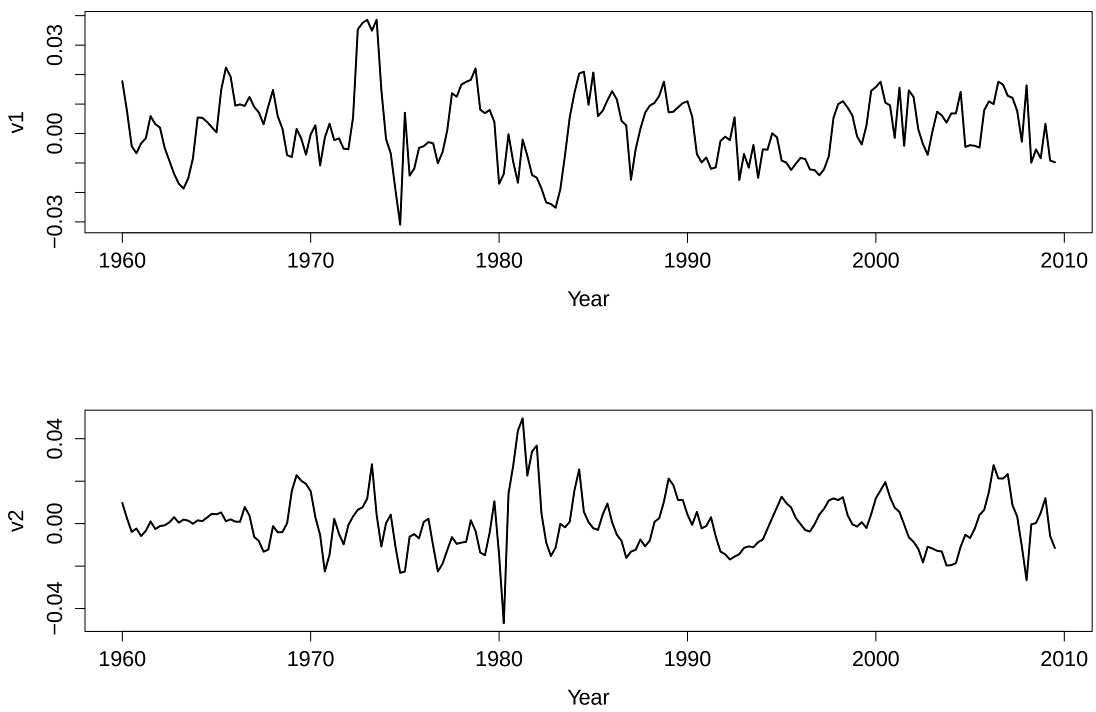
(a) i_t and its estimated trend.

(b) r_t and its estimated trend.

The real solid lines are the estimated trends, and the blue dashed lines are their 90% confidence bands. Since the horizontal zeros-line (dot line) is not entirely contained in the 90% confidence band, it indicates that the trends exist significantly.

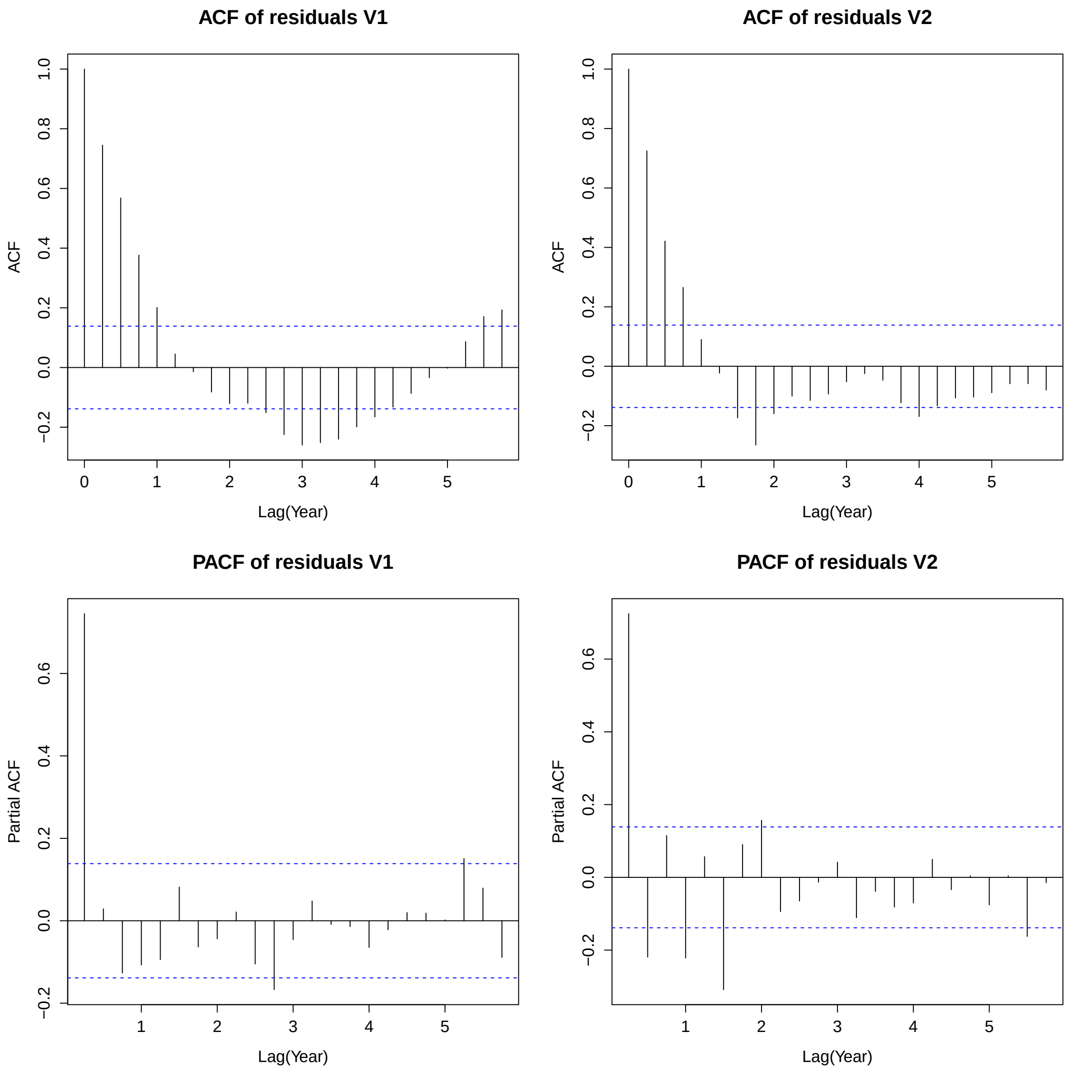
We then examine the stationarity of the two residual sequences $\widehat{v_1}_t = i_t - \widehat{g_1}(\tau_t)$ and $\widehat{v_2}_t = r_t - \widehat{g_2}(\tau_t)$. By visual inspection, the two residual sequences have stable means as they fluctuate steadily around zero in Figure 4.5.

Figure 4.5: The residuals of $\widehat{v_1}_t$ and $\widehat{v_2}_t$.



Meanwhile, Figure 4.6 presents the autocorrelation functions (ACF) and the partial autocorrelation functions (PACF) of the two residual sequences. The ACF and PACF decay to zero very quickly, suggesting very weak serial dependence in the residuals. Further, the p-values of the *Augmented Dickey-Fuller test* with respect to the two residual sequences are smaller than 5%. Therefore, the residuals of $\widehat{v_1}_t$ and $\widehat{v_2}_t$ are stationary time series, and hence the data generating process (4.13) and (4.14) for i_t and c_t are reasonable.

Figure 4.6: The ACF and PACF of the estimated residuals $\widehat{v}_{1t}$ and $\widehat{v}_{2t}$.



To reveal the relationship between the aggregate personal disposable income, the personal consumption expenditure and the real interest rate, we consider the linear regression[4] model[5]

$$c_t = i_t\beta + r_t\gamma + e_t, \tag{4.17}$$

where the regressors are assumed to follow (4.13) and (4.14). Intuitively, consumers

[4]It is equivalent to estimate $c_t = \alpha + \delta t + i_t\beta + r_t\gamma + e_t$ if we do not remove linear trends in c_t and i_t.

[5]We subtract each time series with its mean and therefore the intercept term is ignored.

spend more if there were an increase in the personal disposable income or a decline in the real interest rate. Hence, we would expect a positive value of $\widehat{\beta}$ and a negative value of $\widehat{\gamma}$. The problem of endogeneity arises as the error terms v_{1t} and v_{2t} are possibly correlated with e_t. Hence, the simple OLS estimation of the coefficients are not reliable. In the following two subsections, the regression is conducted under the assumptions of weak and strong trends respectively.

4.3.1 The weak trending regression model

In this subsection, we assume that the time trends in i_t and r_t are bounded functions of τ_t in (4.15) and (4.16). Therefore, i_t and r_t contain weak trends and the simple OLS estimator is biased and inconsistent. To deal with endogeneity, suppose that the endogenous correlation can be expressed by the nonparametric control function

$$e_t = \lambda(v_{1t}, v_{2t}) + u_t, \tag{4.18}$$

where u_t is uncorrelated with v_{1t} and v_{2t}. Hence, replacing e_t in the linear regression model, we have

$$c_t = i_t\beta + r_t\gamma + \lambda(v_{1t}, v_{2t}) + u_t, \tag{4.19}$$

in which the nonparametric control function captures the endogenous correlation without the risk of misspecification. The problem of endogeneity disappears since u_t is assumed to be uncorrelated with i_t, r_t, v_{1t}, and v_{2t}.

Table 4.4: Estimated coefficients in the weak trending regression model.

	OLS	Control function
$\widehat{\beta}$	0.7541^{***} (0.0232)	0.7718^{***} (0.0266)
$\widehat{\gamma}$	-0.2438^{***} (0.0435)	-0.3663^{***} (0.0662)

The estimates of the coefficients are summarized in Table 4.4. They are significant at the 1% significance level. The estimated coefficients have the correct signs as expected that higher income leads to higher consumption (positive β), while higher real interest rate encourages people to save more and spend less (negative γ). Meanwhile, due to the issue of endogeneity, the simple OLS estimates underestimate the elasticity of income and real interest rate to consumption.

Since $\widehat{\beta}$ and $\widehat{\gamma}$ are unbiased and consistent estimators for β and γ by using the control function approach, we can recover the control function $e_t = \lambda(v_{1t}, v_{2t})$ using the residuals $\widehat{e_t}, \widehat{v}_{1t}, \widehat{v}_{2t}$.

Figure 4.7: The local linear kernel estimation of the control function

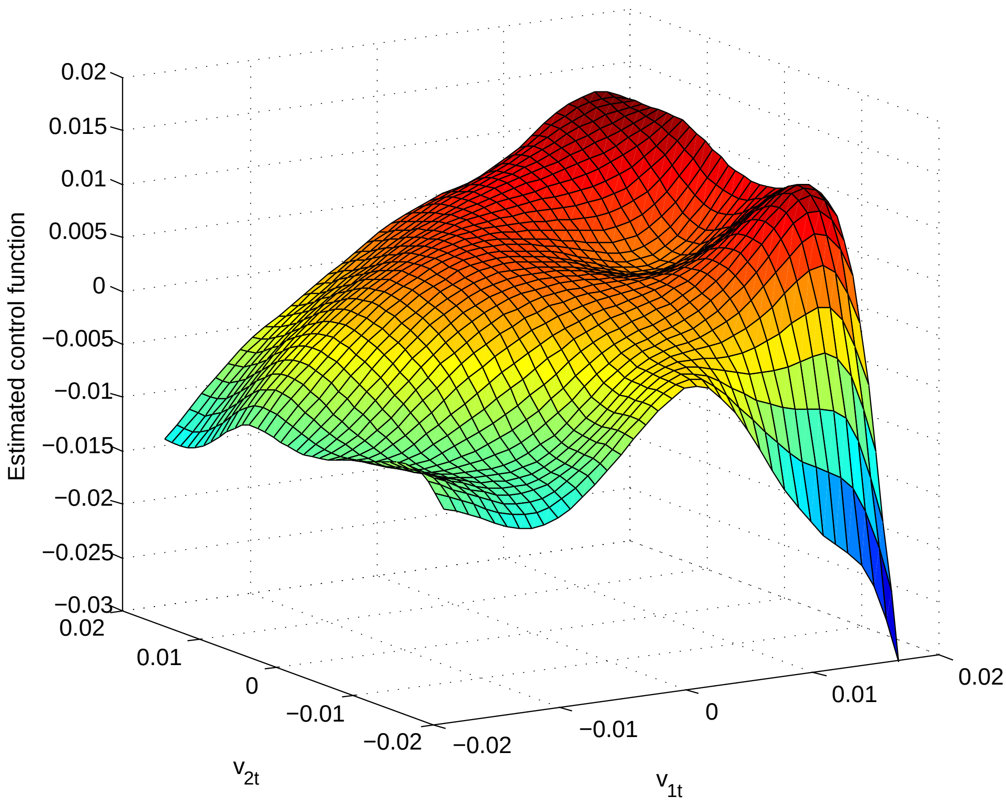
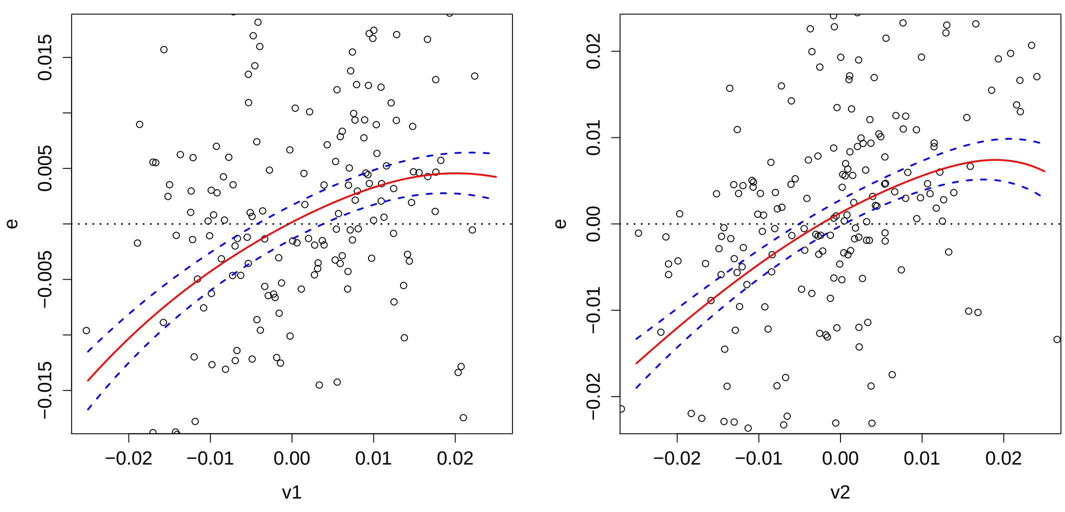


Figure 4.7 is the local linear kernel estimation of the control function, which is nonlinear and can hardly be expressed as additive functions of v_{1t} and v_{2t}. Therefore, if we specify incorrect parametric forms for the control function, for example, $\lambda(v_{1t}, v_{2t}) = \rho_1 v_{1t} + \rho_2 v_{2t}$, such misspecification would lead to inconsistent estimates of

β and γ.

Figure 4.8: The estimated conditional means of the control function



(a) $m_1(v_1) = \mathrm{E}[\lambda(v_{1t}, v_{2t})|v_{1t} = v_1]$ (b) $m_2(v_2) = \mathrm{E}[\lambda(v_{1t}, v_{2t})|v_{2t} = v_2]$

In addition, Figure 4.8 presents the nonparametric kernel estimates of the expectations of the control function conditional on v_{1t} and v_{2t}, respectively. The 2-dimensional graphs provide much convenience to the determination of the significance of the control functions with respect to v_{1t} and v_{2t}. The 90% confidence bands (the blue dashed lines) show that both regressors are endogenously correlated with the error term in the regression model (1.2). Also, the form of such correlation is not linear as what we have observed in the graph of $\widehat{\lambda}(\widehat{v}_{1t}, \widehat{v}_{2t})$.

4.3.2 The strong trending regression model

In practice, the logarithm of aggregate income is usually believed to be a random walk process with drift, i.e.,

$$i_t = \omega + i_{t-1} + \nu_{1t}, \tag{4.20}$$

for some constant $\omega > 0$. The positive drift generates the upward trend and a pure random walk is then left behind as z_t.

$$i_t = i_0 + \omega t + z_t, \tag{4.21}$$

where $z_t = \sum_{s=1}^{t} v_{1s}$. The order of magnitude is 2 for the pure random walk process that $\sum_{t=1}^{n} z_t^2 = O_P(n^2)$. As discussed in the introduction, it is difficult to distinguish it from a nonlinear trend-stationary process. Therefore, it is also reasonable to assume that in addition to the linear upward trend, there still exists a time trend that is weaker than the linear time trend, but stronger than a weak trend (therefore $1 < d < 3$). Specifically, the data generating process of log income can be written as

$$i_t = i_0 + w_1 t + g_1(t) + v_{1t}. \tag{4.22}$$

where v_{1t} is stationary and $g_1(t)$ is a deterministic nonlinear time trend with magnitude order $1 < d < 3$. Since the strong linear trend ωt is caused by the average growth rate, which is not our concern, we remove the linear trend and focus on

$$i_t = g_1(t) + v_{1t}, \tag{4.23}$$

where $g_1(t)$ is a strong trend. Similarly, the real interest rate r_t can also be written as

$$r_t = g_2(t) + v_{2t}, \tag{4.24}$$

in which $g_2(t)$ is a strong trend. Therefore, we consider the regression model[6] as follows.

$$c_t = i_t\beta + r_t\gamma + e_t, \tag{4.25}$$

where i_t and r_t contain strong trends $g_1(t)$ and $g_2(t)$.

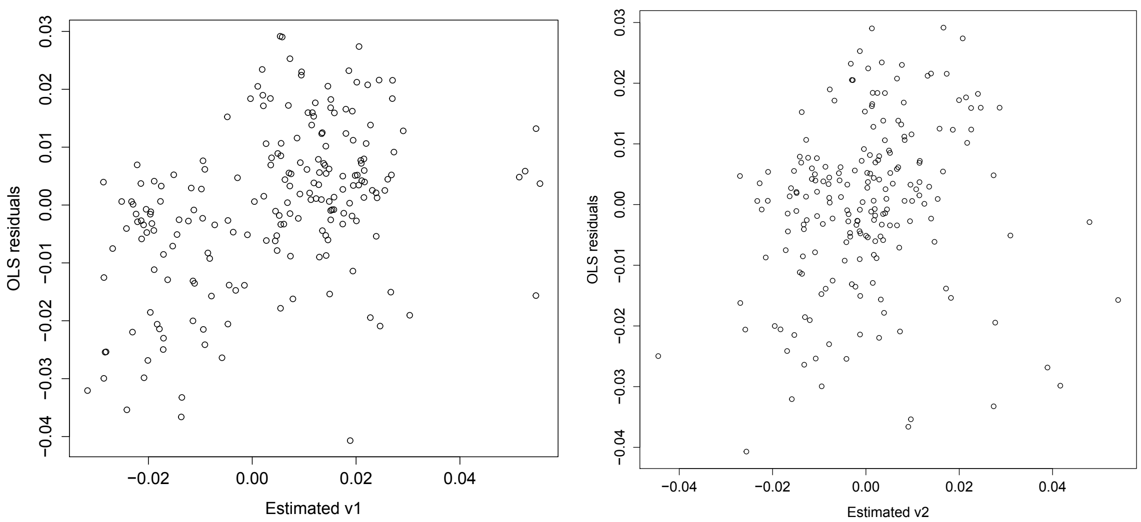
Table 4.5: Estimated coefficients in the strong trending regression model.

	OLS	Bias-correction	s.e.
$\widehat{\beta}$	0.7541	0.7233	0.0231***
$\widehat{\gamma}$	-0.2438	-0.3059	0.0434***

[6]Again, the data has been demeaned and therefore the intercept term can be ignored.

Table 4.5 shows the coefficients in (4.25) estimated by the simple OLS as well as the bias-correction method. For the estimated values of β and γ, the difference between the two methods are approximately 1.5 standard errors for each coefficient.

Figure 4.9: Scatter plots of residuals



(a) $\widehat{v}_{1t}$ and the estimated OLS residuals. (b) $\widehat{v}_{2t}$ and the estimated OLS residuals.

Figure 4.9 shows the scatter plots of the OLS residuals versus the estimated values of v_{1t} and v_{2t} respectively. It is clear that the OLS residuals are correlated with the innovations in i_t and r_t. Specifically, the OLS estimates are positively biased as $\widehat{e_t}$ is positively correlated with $\widehat{v}_{1t}$ and $\widehat{v}_{2t}$.

Chapter 5

Conclusions and further discussion

5.1 Conclusions

We first discussed the differences between the nonlinear trend-stationary process and the unit root process in term of statistical properties and empirical interpretations. We also explained the difficulties to statistically discriminate between the two kinds of data generating processes. The model we studied in this thesis reveals the co-trending relationship between the nonstationary time series with nonlinear deterministic trends, and it is regarded as an analogy to the co-integration model with integrated time series.

To deal with the problem of endogeneity, we adopted two methods to correct for the endogeneity bias in the OLS estimator of the regression models with weak and strong trending time series respectively. We showed that when the regressors have weak trends, the simple OLS estimator is biased and inconsistent as the trending component does not dominate the stationary error term. A nonparametric control function approach is employed to fix the problem by extending the linear regression to a semi-parametric partially linear model. Based on the generating process of our regressors, the usual identifiability condition for the partially linear model is not satisfied. However, we showed that the conventional estimator is still unbiased and consistent in that the sample version of the identifiability condition can be satisfied. Simulation results verified the consistency of the estimator for the partially linear model with comparison to the simple OLS estimator for the linear regression.

While in the second case, we assumed that all the regressors have strong trends, which dominate the stationary error terms. Therefore, the endogeneity bias diminishes, and the simple OLS estimator is consistent when the sample size goes to infinity. However, statistical inferences are affected by the issue of endogeneity. We defined a magnitude parameter that describes the strength of the trends in the regressors, and developed the asymptotic results for different values of such trending parameter. We also found that the estimation bias could be substantially large when the sample size is small. To adjust for the endogeneity, we proposed the bias-corrected estimator based on the fact that the initial OLS estimator is consistent. The asymptotic results as well as the simulation results show that the bias-corrected estimator is unbiased and consistent. Moreover, the size distortion in the hypothesis test is alleviated by using the bias-correction method.

We applied our model and estimation method to study the regression between the logarithm of aggregate personal consumption, the logarithm of the aggregate personal disposable income and the real interest rate. The regression equation represents the long-run behavior of consumption to income and real interest rate. However, it is affected by the problem of endogeneity because income and consumption belong to a system of simultaneous equations. In the two scenarios of trends, we applied the nonparametric control function approach and the bias-correction method respectively. Both methods showed corrections to the OLS estimator, and the nonlinear endogenous correlation between the error terms are found according to the graph of the estimated control function and the scatter plot of the residuals.

One major finding of this thesis is that the econometric properties of the estimators with the trending time series involved are different from those with stationary time series. The main reason is that for the nonstationary time series, we have additional information of trends contained in the data. Of course, this thesis is a very primary study of this topic, and there is much more to be explored in the future.

5.2 Further discussion

Before we discuss the future research directions on the basis of the results in this thesis, let us first review some of the regression models with deterministic and stochastic trends in the next two sections.

5.2.1 Regression models with deterministic trends

We first summarise the models that mainly focus on explaining the variations in the trending time series and panel data. The main feature is that these models have non-linear and nonparametric deterministic time trends included. For example, Gao and Hawthorne (2006) investigate the climate time series data using a semi-parametric model

$$y_t = g(t/n) + \sum_{p=1}^{k} \beta_p x_{pt} + e_t, \tag{5.1}$$

where y_t is a nonstationary trending time series, $g(t/n)$ is a nonparametric function of $\tau_t = t/n$ that captures the nonlinear time trend in the temperature series y_t. The regressors $x_{1t},...,x_{pt}$ are stationary covariates that explain the variations in y_t around the time trend with constant coefficients. The nonparametric form of $g(\cdot)$ allows the data to speak for themselves so that the trend term is free from misspecification. A simple test shows that the estimated trend $\widehat{g}(\cdot)$ should be nonlinear rather than linear.

Later in Cai (2007), the author proposes a varying-coefficient trending time series model formulated as

$$y_t = g(t/n) + \sum_{p=1}^{k} \beta_p(t/n) x_{pt} + e_t, \tag{5.2}$$

where the error term is allowed to be serially correlated. Compared with (5.1), the time-varying coefficient model is more adaptive to the empirical data as it is able to capture the dynamic changes in the relationships between y_t and the regressors.

In Liang and Li (2012), the authors constructed a functional coefficient regression model as

$$y_t = x_t'\beta_1(z_t) + t\beta_2(z_t) + u_t, \tag{5.3}$$

for $t = 1, 2, ..., n$, where z_t is a scalar, and x_t is stationary. Instead of allowing the coefficients to change over time, the authors let them depend on a scalar variable z_t so that we are able to identify the reasons that caused such changes in the trends.

In panel data models, nonlinear time trends are also taken into account to capture the trending feature. Robinson (2012) considers a nonparametric trending regression with cross-sectional dependence. The model is formulated as

$$y_{it} = \alpha_i + \beta_t + e_{it}, \tag{5.4}$$

where e_{it} is allowed to be correlated and heteroscedastic over the cross section.

Chen et al. (2012) study a semi-parametric trending panel data model with cross-sectional dependence

$$y_{it} = x_{it}'\beta + f(t/n) + \alpha_i + e_{it}, \tag{5.5}$$

$$x_{it} = g(t/n) + x_i + v_{it}. \tag{5.6}$$

By incorporating the unknown nonlinear deterministic trends $f(t/n)$ and $g(t/n)$, the model is capable of accommodating a wide range of nonstationary time series.

There are two major improving directions in the above models. First issue is that the regressors in all these models except Chen et al. (2012) need to be stationary. Hence, they can not reveal the co-trending relationship that should be expressed in the form of regressions between trending time series. The second is that none of the above models consider the possible endogeneity issue in the regressions.

5.2.2 Co-integration models

In the thesis, our model deals with nonstationary time series with deterministic trends rather than stochastic trends. Nonlinear trend-stationary is regarded as an alternative way of modeling the nonstationary time series data. Therefore, the co-integration models with unit root time series can be borrowed in a parallel way. i.e., we can consider the same model, but use the data generating process (1.3).

Since the seminal paper by Engle and Granger (1987), co-integration models have been quite popular as it reflects the long-run equilibrium between the variables with stochastic trends. This thesis is motivated by Phillips and Hansen (1990) that studied the endogeneity issue in the co-integration model. Recently, co-integration models with functional coefficients are studied in Xiao (2009a). The model takes the form as

$$y_t = x_t'\beta(z_t) + u_t, \tag{5.7}$$

where x_t is an I(1) process, $\{z_t, u_t\}$ are stationary processes. The coefficient $\beta(\cdot)$ represents the varying relationship between x_t and y_t, which depends on the market or macroeconomic conditions expressed as z_t. Xiao (2009b) studies the quantile co-integration model, where the cointegrating coefficients are computed with respect to different quantiles. He also proposed the quantile-varying coefficient co-integration models, in which the coefficient varies smoothly over the quantiles.

Wang and Phillips (2009) investigated the nonparametric co-integration model

$$y_t = m(x_t) + e_t, \tag{5.8}$$

$$x_t = x_{t-1} + v_t, \tag{5.9}$$

which avoids potential model misspecification. They considered the endogeneity problem that e_t and v_t are correlated. Dong and Gao (2014) considered the same model and developed a specification test using the method of series expansions.

A general functional coefficient nonstationary regression model is investigated in Gao and Phillips (2013). The model studies the non- and semi-parametric co-integrations of the form

$$y_t = x_t'\beta(z_t, u_t) + e_t, \tag{5.10}$$

$$x_t = x_{t-1} + \mu_t, \tag{5.11}$$

$$z_t = z_{t-1} + \nu_t, \tag{5.12}$$

where $\{u_t, e_t, \mu_t, \nu_t\}$ are stationary time series. Under certain conditions, their model is able to incorporate several useful models, including the additive nonparametric regression and the partially linear models with integrated processes.

5.2.3 Some future research directions

Based on the above discussions on deterministic and stochastic trending time series models, the current work in this thesis can be further developed in following directions.

(1) [Mixture of weak and strong deterministic trends] In the thesis, we only consider the case where all the trends are either weak or strong. While in practice, the regressors may have weak and strong trends simultaneously. i.e.,

$$y_t = \sum_{i=1}^{p} x_{it}\beta_i + \sum_{j=1}^{q} z_{jt}\gamma_j + e_t, \tag{5.13}$$

where

$$x_{it} = g_i(t) + v_{it}, \tag{5.14}$$

$$z_{jt} = h_j(t) + v_{jt}, \tag{5.15}$$

in which $g_i(t)$ are weak trends and $h_j(t)$ are strong trends for $i = 1, 2, ..., p$ and $j = 1, 2, ..., q$.

(2) [Time-varying coefficient models] The varying coefficient model is able to show the dynamic changes in the relationships between the economic variables. It is more adaptive to the empirical applications compared with the constant coefficient model in this thesis. We can consider the problem of endogeneity in the varying coefficient model formulated as

$$y_t = x_t'\beta_t + e_t, \tag{5.16}$$

$$x_t = g(t) + v_t, \tag{5.17}$$

where e_t and v_t are endogenously correlated. In particular, when the first element of x_t is assumed to be 1 constantly, then the first coefficient β_{1t} represents the trend in y_t that is not explained by the regressors.

(3) [Nonparametric regression] The linear regression model discussed in this thesis is misspecified if the true relationship between the variables is not linear. To solve this problem, we should consider the nonparametric regression model with trending time series of the form

$$y_t = m(x_t) + e_t, \tag{5.18}$$

$$x_t = g(t) + v_t, \tag{5.19}$$

where $m(\cdot)$ is a continuous and differentiable function, and e_t and v_t are allowed to be correlated.

5.2.4 Empirical applications

Since most of the economic data contain trends, the trending time series regressions can be widely applied in the empirical applications.

(1) [Permanent income hypothesis] Since Hall (1978), the time series of income and consumption are usually assumed to be a random walk process based on the assumption of rational expectations. Therefore, permanent income is the same as current income. The permanent income hypothesis is then verified if consumption tracks income perfectly. However, the conclusion was carried out by the regression with the differenced data (i.e., the growth rates), and the detrending process may induce spurious cycles in the residuals that lead to unreliable inferences. For references, see Mankiw and Shapiro (1985, 1986), Campbell (1987), Campbell and Mankiw (1990), Han and Ogaki (1997), Fernández-Villaverde and Krueger (2007).

In fact, the mean of the first difference of log income (i.e., the growth rate) may not be a constant over time as shown in Figure 5.1. There have been periods with high growth rates and other periods with low growth rates. Therefore, the drift term in the random walk process is smoothly varying instead of a constant value over time. Specifically,

$$x_t = a_t + x_{t-1} + v_t, \tag{5.20}$$

where a_t is a term of low-frequency information. In Figure 5.1, the red dashed line is the nonparametrically estimated drift term $\widehat{a_t}$. Intuitively, when $a_t = a$, the constant drift generates a linear time trend. However, when a_t is time-varying, it generates nonlinear time trends as

$$g(t) = g(0) + \sum_{s=1}^{t} a_s, \tag{5.21}$$

where $g(0)$ is the starting point of the nonlinear time trend. Removing the deterministic trend $g(t)$, we obtain the 'detrended' component of x_t as

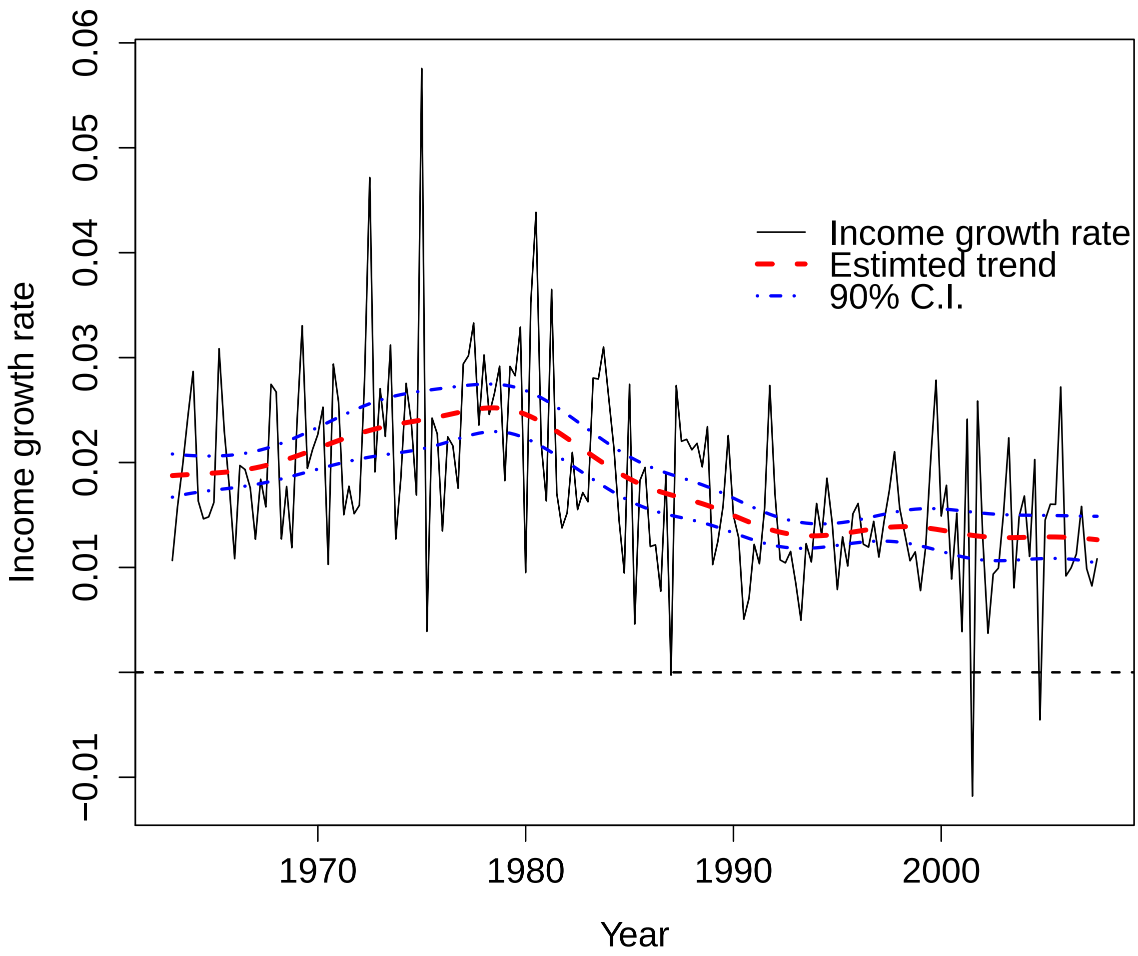
$$\widetilde{x_t} = x_t - \widehat{g}(t) = x_t - g(0) - \sum_{s=1}^{t} \widehat{a_s}, \tag{5.22}$$

where $\widetilde{x_t}$ is found to be stationary. Therefore, the log income is a nonlinear trend-stationary process and an MA unit root was generated in Δx_t after differentiation.

Generally, since the nonlinear time trend can always capture the nonlinearity and nonstationarity in x_t regardless of the existence of unit root, we can then verify the permanent income hypothesis by regressing the log aggregate personal consumption over the log aggregate personal income.

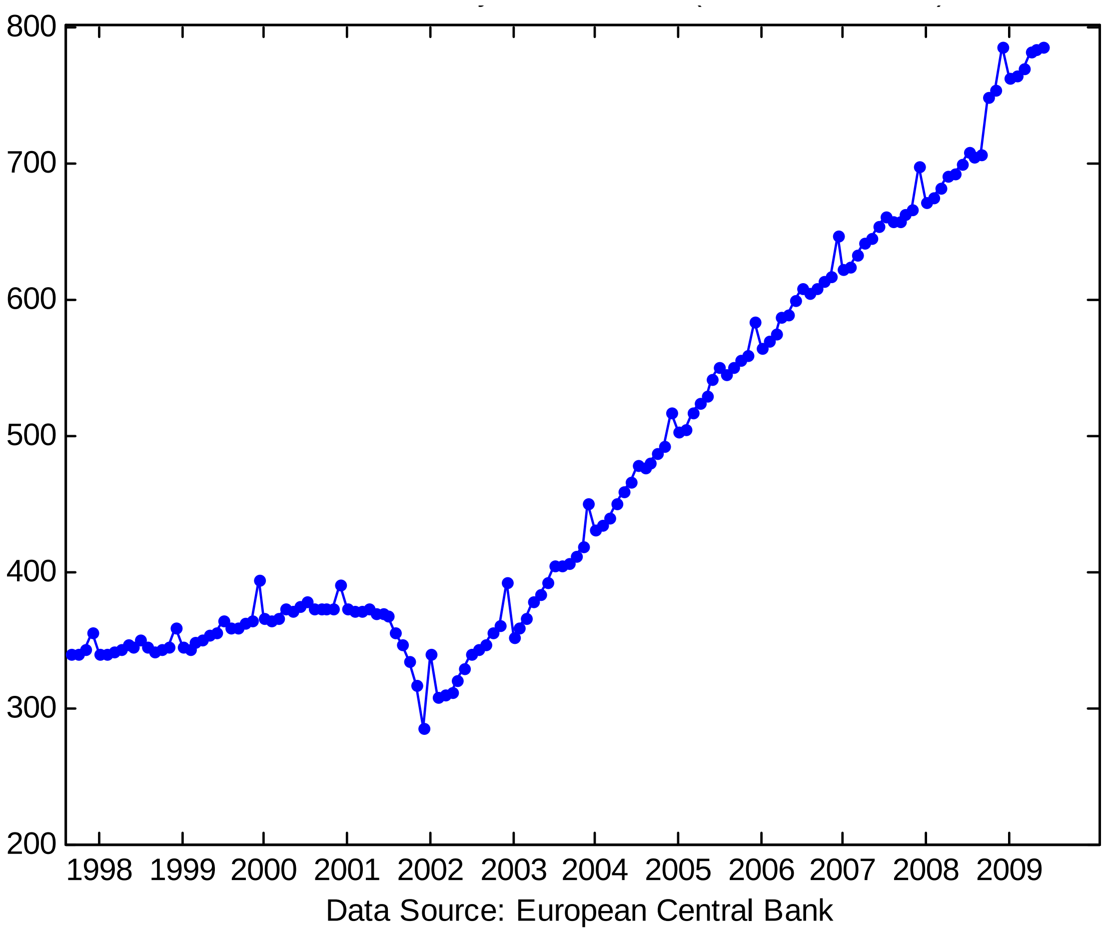
(2) [Breaks, bubbles and cycles] In White and Granger (2011), the authors discussed the breaks, bubbles and cycles in economics, which are highly related to the characteristics of trends involved in the time series. Therefore, trending time series models can be used to detect and analyze the structural breaks, economic bubbles and business cycles. Figure 5.2 is the example used in White and Granger (2011) that illustrates the structural break in the time series of the amount of currency in circulation for the Euro Area. There is an obvious break at 2002. Before the break point, the trend is weak, however, it becomes a strong linear trend after the break. Intuitively, there is an intrinsic linkage between the occurrence of a trend break and the change of the trending parameter d on the two sides of the break point. The reason

Figure 5.1: The growth rate of aggregate personal income



is that trend break (or the place where a bubble starts) implies changes of the trend type and therefore will cause the variation in the magnitude parameter that represents the strength of the trend. Recently, the problem of bubble detection has been quite popular, see Phillips et al. (2011) and Phillips et al. (2015), in which they developed the sup augmented Dickey-Fuller test. In short, the quantitative analysis of these bubble phenomena depends on our understanding of the trend behaviors as well as the econometric tools that can deal with them effectively.

Figure 5.2: Currency in Circulation: an example of break in White and Granger (2011)



Appendix A

A constructed IV approach

In the first year of my PhD candidature, I believed that the conventional estimation method for the semiparametric partially linear model cannot be used because the identifiability condition is not satisfied for our model. Therefore, under particular assumptions, I proposed a method to estimate the coefficients in the linear regression model (1.2) by constructing functions of v_t in (1.3) as instrumental variables. Here I explain the method in this Appendix just for the completeness of my PhD thesis.

A.1 Identification and assumptions

Recall that our model is

$$y_t = x_t'\beta + e_t, \tag{A.1}$$

$$x_t = g(\tau_t) + v_t, \tag{A.2}$$

$$e_t = \lambda(v_t) + u_t. \tag{A.3}$$

where x_t is a k-dimensional vector of trending regressors as equation (1.3). We first introduce some assumptions for the error terms in the model before we estimate the parameters.

Assumption A.1.1. Let (u_t, v_t) be zero-mean stationary and $\alpha-$mixing innovations with mixing coefficient $\alpha(k)$ satisfying $\sum_{k=1}^{\infty} \alpha^{\frac{\delta}{2+\delta}}(k) < \infty$ for some $\delta > 0$, where δ is chosen such that $E[|u_t|^{2+\delta}] < \infty$ and $E[\|v_t\|^{2+\delta}] < \infty$. u_t and v_t are uncorrelated, and $E[u_t^2] = \sigma_u^2 > 0$, $E[v_t v_t'] = \Omega_v$, which is a symmetric positive definite matrix.

Assumption A.1.2. Let $\lambda(\cdot)$ be a continuous function defined on $\mathbb{R}^k$ to $\mathbb{R}^1$. There always exists some continuous function $\pi(v_t)$ that satisfies $E[\pi(v_t)\lambda(v_t)] = 0$, and $\Sigma_v = E[\pi(v_t)]\int_0^1 g(\tau)d\tau + E[\pi(v_t)v_t']$ is a positive definite matrix.

Assumption A.1.3. Let $K(\cdot)$ be a symmetric and continuous probability density function with $\int K^2(u)du < \infty$, $\int uK(u)du = 0$, $\int uu'K(u)du < \infty$. Let h be the bandwidth, which satisfies $h \to 0$, $nh \to \infty$, and $nh^2 \to \infty$ as $n \to \infty$.

A.1.1 Parameter estimation

Consider equation (A.1) in the model, the error term u_t can be written as

$$u_t = y_t - x_t'\beta - \lambda(v_t). \tag{A.4}$$

Since we have assumed that u_t and v_t are uncorrelated, the following orthogonal condition holds.

$$\mathrm{E}[\pi(v_t)u_t] = 0. \tag{A.5}$$

Substituting u_t by (A.4), we have

$$\mathrm{E}[\pi(v_t)y_t] = \mathrm{E}[\pi(v_t)x_t']\beta + \mathrm{E}[\pi(v_t)\lambda(v_t)]. \tag{A.6}$$

Assumption A.1.2 suggested that $\mathrm{E}[\pi(v_t)\lambda(v_t)] = 0$, therefore

$$\mathrm{E}[\pi(v_t)y_t] = \mathrm{E}[\pi(v_t)x_t']\beta. \tag{A.7}$$

Hence, β can be explicitly expressed provided that the inverse $\mathrm{E}[\pi(v_t)x_t']^{-1}$ exists. In fact,

$$\mathrm{E}[\pi(v_t)x_t'] = \mathrm{E}[\pi(v_t)(g(\tau_t) + v_t)'] = \mathrm{E}[\pi(v_t)]\int_0^1 g(\tau)d\tau + \mathrm{E}[\pi(v_t)v_t'] = \Sigma_v, \tag{A.8}$$

where Σ_v is assumed to be an invertible positive definite matrix by Assumption A.1.2. We obtain an equation which can be used to estimate β as follows.

$$\beta = \mathrm{E}[\pi(v_t)x_t']^{-1}\mathrm{E}[\pi(v_t)y_t]. \tag{A.9}$$

By the method of moments, β can be estimated by the sample analogue of (A.9).

$$\widehat{\beta_0} = \Big(\sum_{t=1}^{T}\pi(v_t)x_t'\Big)^{-1}\Big(\sum_{t=1}^{T}\pi(v_t)y_t\Big). \tag{A.10}$$

Since v_t is not observable, the estimator $\widehat{\beta_0}$ is not feasible. By equation (A.2), we construct an alternative feasible estimator

$$\widehat{\beta} = \Big(\sum_{t=1}^{T}\pi(\widehat{v_t})x_t'\Big)^{-1}\Big(\sum_{t=1}^{T}\pi(\widehat{v_t})y_t\Big). \tag{A.11}$$

where

$$\widehat{v_t} = x_t - \widehat{g}(\tau_t). \tag{A.12}$$

and $\widehat{g}(\tau_t)$ is the nonparametric kernel estimator for the trend term $g(\tau_t)$.

A.2 Nonparametric estimation

There are various kinds of methods to estimate the trend term. As discussed in the introduction, misspecified parametric forms may cause inconsistent estimations. In this thesis, we employ the nonparametric kernel methods. For instance, we use the nonparametric local constant method to estimate $g(\tau)$ at some point $\tau \in (0,1)$ that

$$\widehat{g}(\tau) = \sum_{s=1}^{n} w_{ns}(\tau)x_s, \tag{A.13}$$

where

$$w_{ns}(\tau) = \frac{K\left(\frac{\tau_s-\tau}{h}\right)}{\sum_{p=1}^{n} K\left(\frac{\tau_p-\tau}{h}\right)}. \tag{A.14}$$

The control function $\lambda(v_t)$ can also be estimated by kernel methods. Once the trend term and coefficients of the linear component are properly estimated, the control function can be approximated by the methods in Fan and Gijbels (1996).

A.3 Asymptotic properties

Theorem A.3.1. Let Assumptions A.1.1 and A.1.3 hold, as $n \to \infty$,

$$\sqrt{nh}\left(\widehat{g}(\tau) - g(\tau) - \frac{h^2}{2} g''(\tau)\mu_2\right) \xrightarrow{p} N(0, \Omega_v \kappa_2), \tag{A.15}$$

for $\tau \in [0,1]$ and $\mu_2 = \int_{-\infty}^{\infty} u^2 K(u) du$, $\kappa_2 = \int_{-\infty}^{\infty} K^2(u) du$.

Theorem A.3.2. Let Assumptions A.1.1 to A.1.3 hold, we have

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, \Omega). \tag{A.16}$$

where $\Omega = \Sigma_v^{-1} \Gamma \Sigma_v^{-1}$, Γ is the long run variance-covariance matrix of $\xi_t = \pi(v_t) e_t$ that

$$\Gamma = \sum_{j=-\infty}^{\infty} E[\xi_t \xi_{t-j}]. \tag{A.17}$$

The long-run variance can be consistently estimated by

$$\widehat{\Gamma} = \sum_{j=-p}^{p} \widehat{\Gamma}_j k(j), \tag{A.18}$$

where $p = [\sqrt{n}]^-$, $\widehat{v_t} = x_t - \widehat{g}(\tau_t)$, $k(j)$ is a weight function, $\widehat{e_t} = y_t - x_t \widehat{\beta}$, and

$$\widehat{\Gamma_j} = \frac{\sum_{t=j}^{n} \widehat{v_t} \widehat{e_t} \widehat{e}'_{t-j} \widehat{v}'_{t-j}}{n}. \tag{A.19}$$

A.4 Simulation results

A.4.1 I.I.D. error terms

The data generating process follows (A.1) to (A.3) and has explicit forms as $g(\tau_t) = 7 + 2\tau_t$, $\lambda(v_t) = 0.5 v_t$, and $y_t = 0.6 x_t + e_t$, where $u_t, v_t \overset{i.i.d.}{\sim} N(0,1)$. In this example, we first let $\pi(v_t) = \frac{1}{1+v_t^2}$.

Repeat=500	$\widehat{\beta}$			$\widehat{\beta}_{ols}$		
Sample size	300	600	1000	300	600	1000
Bias(×1000)	-0.0215	0.2431	0.0866	7.8000	7.9000	7.8000
Std	0.0086	0.0062	0.0046	0.0079	0.0058	0.0043
RMSE	0.0086	0.0062	0.0046	0.0111	0.0098	0.0089
$\bar{\widehat{\Omega}}$	0.0196	0.0197	0.0196	0.0195	0.0196	0.0196
$Std/\sqrt{\widehat{\Omega}}$	0.0616	0.0442	0.0329	0.0561	0.0411	0.0305

The next table shows the simulation results based on another case of the trend function. In particular, when $g(\tau_t) = 2\sin(3.14\tau_t)$, the IV estimator is still consistent.

Repeat=500	$\widehat{\beta}$			$\widehat{\beta}_{ols}$		
Sample size	300	600	1000	300	600	1000
Bias	-0.0007	-0.0042	-0.0006	0.1663	0.1644	0.1658
Std	0.0547	0.0357	0.0299	0.0354	0.0235	0.0196
RMSE	0.0547	0.0360	0.0299	0.1700	0.1660	0.1669
$\bar{\widehat{\Omega}}$	0.7826	0.7822	0.7778	0.3613	0.3613	0.3613
$Std/\sqrt{\widehat{\Omega}}$	0.0618	0.0404	0.0339	0.0589	0.0391	0.0327

In the next table, we show the estimation results if we choose $\pi(v_t) = e^{-v_t^2}$ as an alternative IV. When $g(\tau_t) = 7 + 2\tau_t$, the simulation results are

Repeat=500	$\widehat{\beta}$			$\widehat{\beta}_{ols}$		
Sample size	300	600	1000	300	600	1000
Bias(×1000)	0.0069	-0.2682	0.1334	7.7000	7.5000	7.6000
Std	0.0096	0.0068	0.0049	0.0083	0.0058	0.0042
RMSE	0.0096	0.0068	0.0049	0.0113	0.0095	0.0087
$\bar{\widehat{\Omega}}$	0.0219	0.0220	0.0219	0.0190	0.0190	0.0190
$Std/\sqrt{\widehat{\Omega}}$	0.0646	0.0457	0.0334	0.0600	0.0421	0.0304

Again, when $g(\tau_t) = 2\sin(3.14\tau_t)$, the results are shown in the following table.

Repeat=500	$\widehat{\beta}$			$\widehat{\beta}_{ols}$		
Sample size	300	600	1000	300	600	1000
Bias	0.0000	0.0024	-0.0016	0.1650	0.1649	0.1670
Std	0.0577	0.0417	0.0311	0.0344	0.0261	0.0190
RMSE	0.0577	0.0418	0.0312	0.1686	0.1670	0.1681
$\bar{\bar{\widehat{\Omega}}}$	0.8886	0.8756	0.8803	0.3834	0.3834	0.3834
$Std/\sqrt{\widehat{\Omega}}$	0.0612	0.0466	0.0332	0.0555	0.0421	0.0306

A.4.2 Autocorrelated error terms

When the innovations follow AR(1) process, i.e. other DGPs remain unchanged, u_t and v_t are generated by $u_t = \rho u_{t-1} + \xi_t$ and $v_t = \rho v_{t-1} + \epsilon_t$, where $\xi_t, \epsilon_t \overset{i.i.d.}{\sim} N(0,1)$.

(1) $\rho = 0.2, g(\tau_t) = 2\sin(3.14\tau_t), \pi(v_t) = e^{-v_t^2}$, Cross-validation: leave 5 out.

Repeat=500	$\widehat{\beta}$			$\widehat{\beta}_{ols}$		
Sample size	300	600	1000	300	600	1000
Bias	-0.0142	-0.0078	-0.0056	0.1691	0.1723	0.1708
Std	0.0727	0.0485	0.0371	0.0400	0.0301	0.0221
RMSE	0.0741	0.0491	0.0375	0.1737	0.1749	0.1722
$\bar{\bar{\widehat{\Omega}}}$	1.1498	1.1362	1.1404	0.8517	0.8480	0.8559
$Std/\sqrt{\widehat{\Omega}}$	0.0678	0.0455	0.0347	0.0433	0.0326	0.0239

(2) $\rho = 0.2, g(\tau_t) = 2\sin(3.14\tau_t), \pi(v_t) = \frac{1}{1+v_t^2}$, Cross-validation: leave 5 out.

Repeat=500	$\widehat{\beta}$			$\widehat{\beta}_{ols}$		
Sample size	300	600	1000	300	600	1000
Bias	-0.0168	-0.0072	-0.0038	0.1691	0.1696	0.1722
Std	0.0669	0.0492	0.0366	0.0407	0.0289	0.0215
RMSE	0.0690	0.0498	0.0368	0.1739	0.1720	0.1735
$\widehat{\bar{\Omega}}$	1.1087	1.0154	1.0531	0.8869	0.8440	0.8690
$Std/\sqrt{\widehat{\Omega}}$	0.0636	0.0489	0.0357	0.0432	0.0315	0.0231

(3) $\rho = 0.6, g(\tau_t) = 2\sin(3.14\tau_t), \pi(v_t) = e^{-v_t^2}$, Cross-validation: leave 9 out.

Repeat=500	$\widehat{\beta}$			$\widehat{\beta}_{ols}$		
Sample size	300	600	1000	300	600	1000
Bias	-0.0183	-0.0171	-0.0178	0.2231	0.2200	0.2170
Std	0.1429	0.0941	0.0783	0.0672	0.0483	0.0389
RMSE	0.1441	0.0956	0.0803	0.2330	0.2253	0.2204
$\widehat{\bar{\Omega}}$	3.4662	3.4428	3.3339	1.6787	1.7021	1.7295
$Std/\sqrt{\widehat{\Omega}}$	0.0768	0.0507	0.0429	0.0519	0.0370	0.0296

(4) $\rho = 0.6, g(\tau_t) = 2\sin(3.14\tau_t), \pi(v_t) = \frac{1}{1+v_t^2}$, Cross-validation: leave 9 out.

Repeat=500	$\widehat{\beta}$			$\widehat{\beta}_{ols}$		
Sample size	300	600	1000	300	600	1000
Bias	-0.0277	-0.0181	-0.0138	0.2169	0.2195	0.2163
Std	0.1404	0.0987	0.0750	0.0681	0.0497	0.0399
RMSE	0.1431	0.1003	0.0763	0.2273	0.2251	0.2199
$\widehat{\bar{\Omega}}$	3.3738	3.1418	3.0823	1.6917	1.7254	1.7198
$Std/\sqrt{\widehat{\Omega}}$	0.0764	0.0557	0.0427	0.0524	0.0379	0.0304

Remark A.4.1. All the results in the above tables show that our estimator is unbiased and consistent, while the simple OLS estimator is biased and inconsistent. Meanwhile, the convergence rate of our estimator is $\sqrt{n}$.

A.5 Application

We consider the U.S. personal consumption expenditure and personal income quarterly data from 1947 to 2009. In our model, x_t represents the logarithm of personal income, while y_t represents the logarithm of personal consumption.

$$y_t = x_t\beta + e_t, \tag{A.20}$$

$$x_t = g(\tau_t) + v_t, \tag{A.21}$$

$$e_t = \lambda(v_t) + u_t, \tag{A.22}$$

which is equivalent to estimate

$$y_t = x_t\beta + \lambda(v_t) + u_t, \tag{A.23}$$

$$x_t = g(\tau_t) + v_t, \tag{A.24}$$

The estimation involves the choice of $\pi(v_t)$, hence the subsequent analysis follows two schemes depending on the characteristics of $\lambda(v_t)$.

A.5.1 Case I: $\lambda(v_t)$ is an odd function of v_t

Suppose that $\lambda(v_t)$ is an odd function of v_t. For example, the error terms e_t and v_t follow a joint normal distribution. Let $\pi(v_t) = (1, v_t^2)'$, $X_t = (1, x_t)'$, $\gamma = (\alpha, \beta)'$, we have

$$y_t = X_t'\gamma + \lambda(v_t) + u_t. \tag{A.25}$$

Multiply both sides of the equation by $\pi(v_t)$, and take expectation,

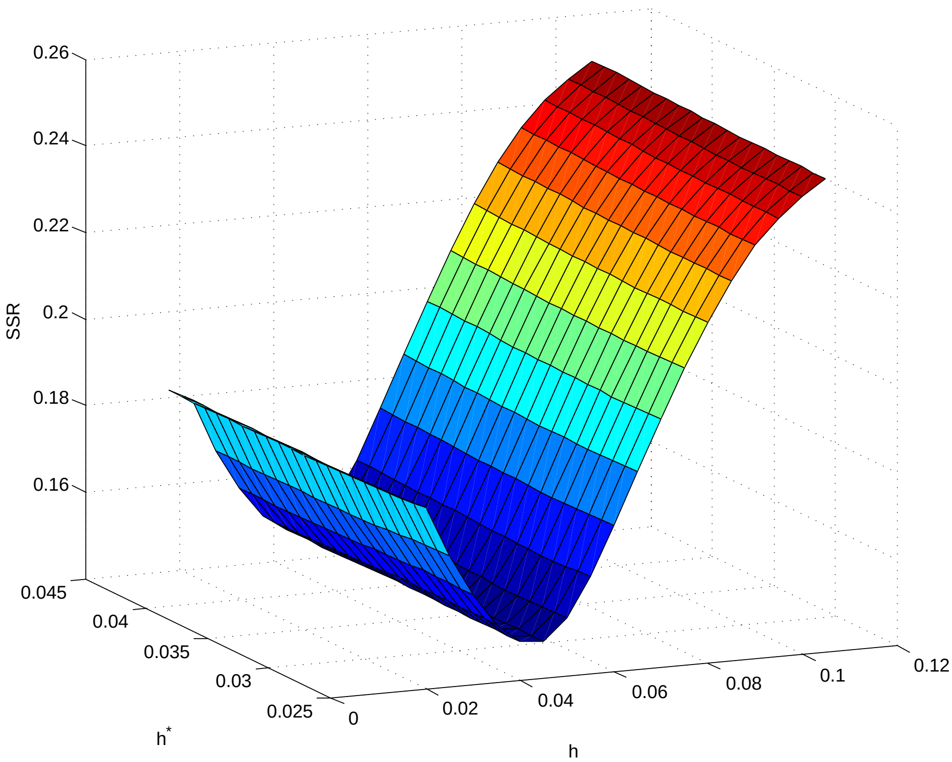
$$\mathrm{E}\begin{pmatrix} y_t \\ v_t^2 y_t \end{pmatrix} = \mathrm{E}\begin{pmatrix} 1 & x_t^2 \\ v_t^2 & v_t^2 x_t \end{pmatrix}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \mathrm{E}\begin{pmatrix} \lambda(v_t) \\ v_t^2\lambda(v_t) \end{pmatrix} + \mathrm{E}\begin{pmatrix} u_t \\ v_t^2 u_t \end{pmatrix}. \tag{A.26}$$

Because $\lambda(v_t)$ is an odd function of v_t, it is straightforward to obtain $\mathrm{E}(\lambda(v_t)) = 0$ and $\mathrm{E}(v_t^2\lambda(v_t)) = 0$. Meanwhile, the parameters are identified only if the matrix in front of the coefficients is positive definite. Therefore, if $\mathrm{E}(v_t^2)\int_0^1 g(\tau)d\tau - (\mathrm{E}(v_t^2))^2 > 0$,

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \begin{pmatrix} 1 & n^{-1}\sum_{t=1}^n v_t^2 \\ n^{-1}\sum_{t=1}^n v_t^2 & n^{-1}\sum_{t=1}^n v_t^2 x_t \end{pmatrix}^{-1}\begin{pmatrix} n^{-1}\sum_{t=1}^n y_t \\ n^{-1}\sum_{t=1}^n v_t^2 y_t \end{pmatrix}. \tag{A.27}$$

To make the estimators feasible, we substitute v_t by its consistent estimator $\widehat{v_t} = x_t - \widehat{g}(\tau_t, h)$. Then estimate $\lambda(v_t)$ using local linear kernel estimation with bandwidth chosen as h^* and obtain the error terms $\widehat{u_t}$. Figure (A.1) shows the sum of squared residuals $\sum_{t=1}^{T} \widehat{u}_t^2$ under different bandwidth pairs of (h, h^*), where the smallest value of SSR is obtained at an interior point. For fixed h^*, the SSR and bandwidth form a U-shape

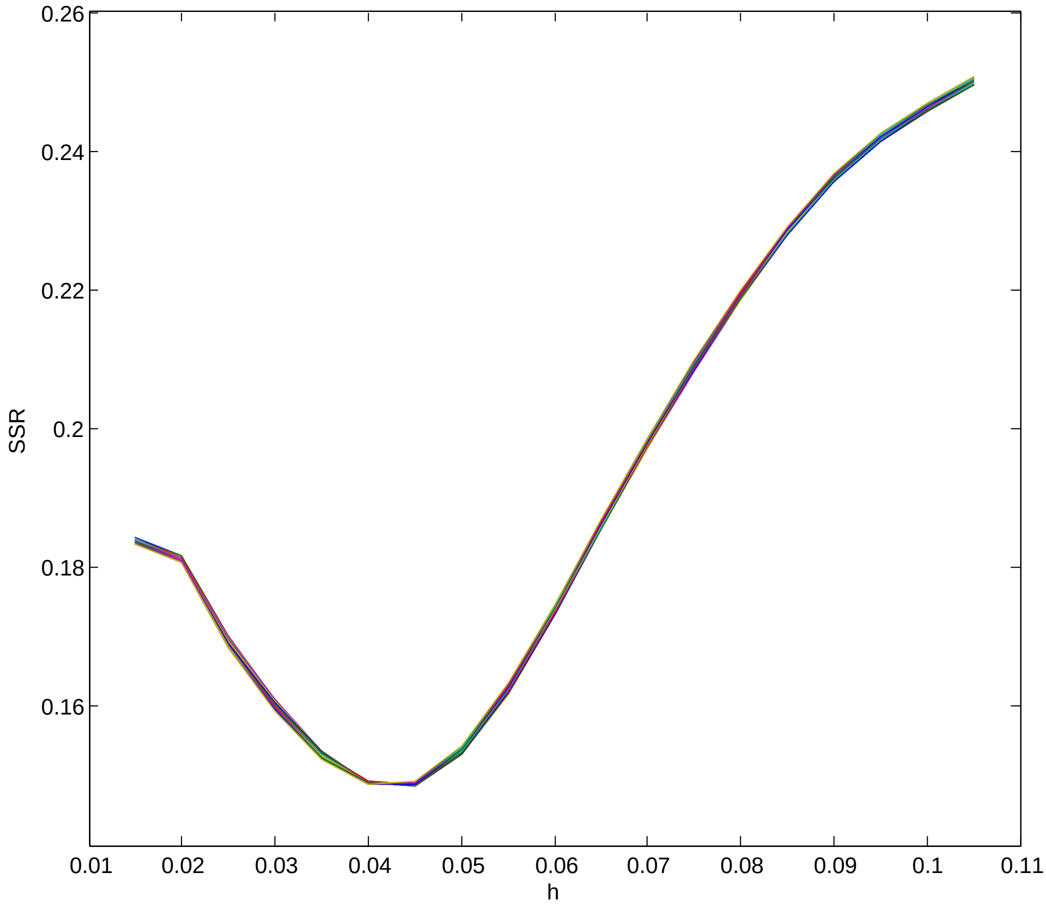
Figure A.1: Sum of squared residuals for the bandwidth selection



functional relation as shown in Figure (A.2).

Therefore, the optimal bandwidth for h is 0.045, under this bandwidth, the SSR under different h^* is shown in Figure (A.3). At the lowest point, SSR is minimized by choosing $h^* = 0.029$. The parameters can be estimated as $\widehat{\alpha} = -0.1994, \widehat{\beta} = 1.0118$ under selected bandwidths. The functional form of $\lambda(v_t)$ can be estimated by local linear estimation as the odd function shown in Figure (A.4). The p-values of ADF-test for the residual series $\{\widehat{e_t}\}$ and $\{\widehat{v_t}\}$ are 0.0001 and 0.0012 respectively.

Figure A.2: Sum of squared residuals for fixed h^*



A.5.2 Case II: $\lambda(v_t)$ is an even function of v_t

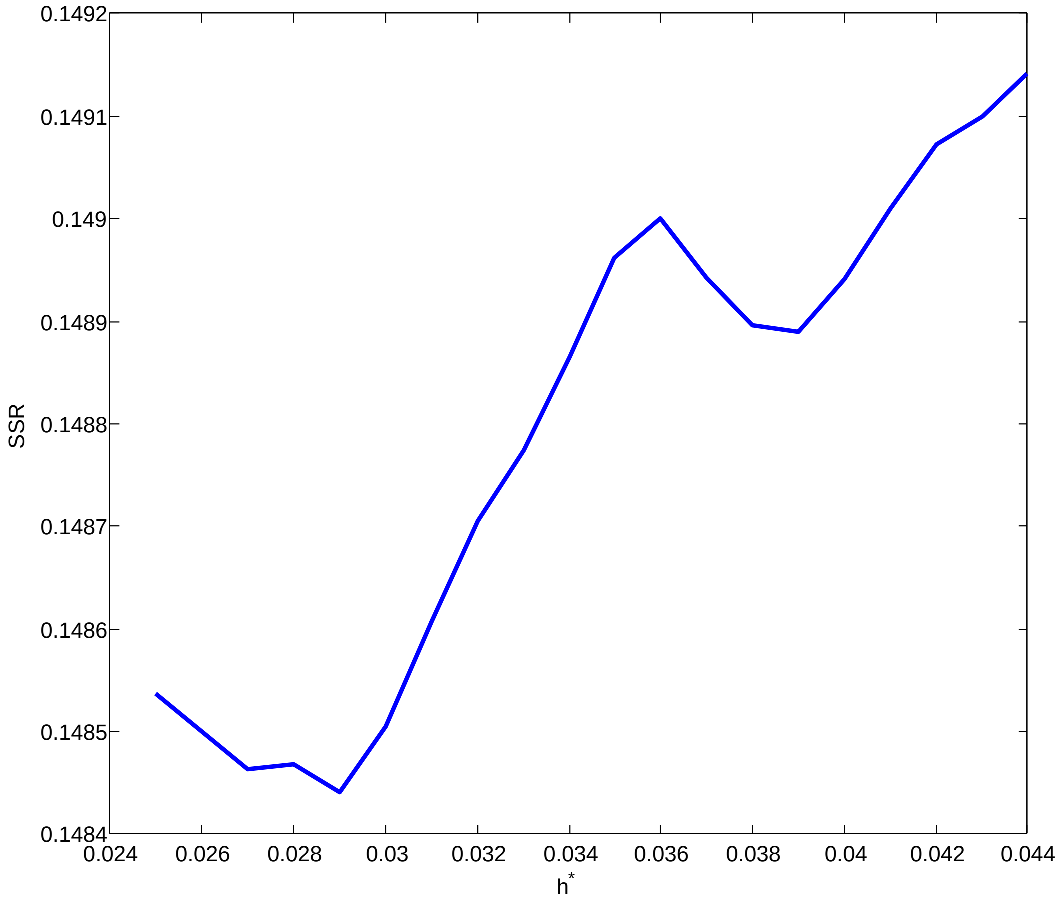
If the control function is symmetric about the y-axis, the assumption $E(\lambda(v_t)) = 0$ can be relaxed and the intercept term in the model can be eliminated and included in the control function. Consequently,

$$y_t = x_t\beta + \lambda(v_t) + u_t. \tag{A.28}$$

Multiply both sides by $\pi(v_t) = v_t$, and take expectation.

$$\mathrm{E}(v_t y_t) = \mathrm{E}(v_t x_t)\beta + \mathrm{E}(v_t\lambda(v_t)) + E(v_t u_t). \tag{A.29}$$

Figure A.3: Sum of squared residuals for fixed h^*



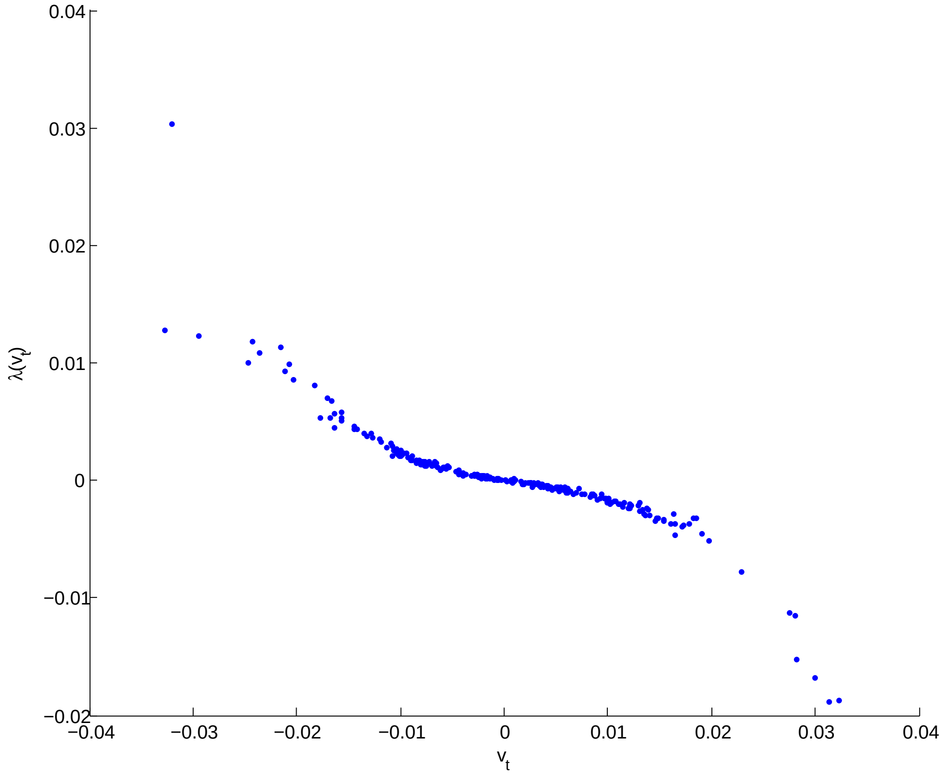
The parameter β can be estimated by

$$\widehat{\beta} = \left(\sum_{t=1}^{T} \widehat{v}_t x_t \right)^{-1} \left(\sum_{t=1}^{T} \widehat{v}_t y_t \right), \tag{A.30}$$

where $v_t = x_t - \widehat{g}(\tau_t, h)$. The bandwidths are selected by minimizing the sum of squared residuals of the model as $h = 0.095, h^* = 0.016$, which can be seen from figure (A.5), figure (A.6) and figure (A.7).

The coefficient β is estimated as $\widehat{\beta} = 0.9865$, and the estimated functional structure of $\lambda(v_t)$ is shown in Figure (A.8). The p-value for the ADF-test for the residual series $\{\widehat{e}_t\}$ and $\{\widehat{u}_t\}$ are 0.0076 and 0.0022 respectively. Figure (A.9) shows the raw data of income and consumption, together with the estimated trend term and fitted value of consumption.

Figure A.4: Estimated control function $\widehat{\lambda}(v_t)$.



A.6 Conclusions

This appendix introduced the estimation methods for the trending regressions with endogeneity by constructing a function that is orthogonal to the control function. The assumption that $E(\pi(v_t)\lambda(v_t)) = 0$ holds for some continuous function $\pi(v_t)$ was quite important for identifying and estimating the coefficient.

There are several aspects to improve the model in the future. Since $\lambda(v_t)$ is unable to capture the time trend, and it is likely that the time trend in y_t cannot be fully explained by x_t. It is reasonable to add a trend component in the regression as the following model.

$$y_t = x_t'\beta + f(\tau_t) + \lambda(v_t) + u_t, \tag{A.31}$$

$$x_t = g(\tau_t) + v_t, \tag{A.32}$$

where $f(\tau_t)$ captures the remaining trend in y_t that can not be explained by x_t.

Figure A.5: Sum of Squared Residuals under different bandwidth pairs.

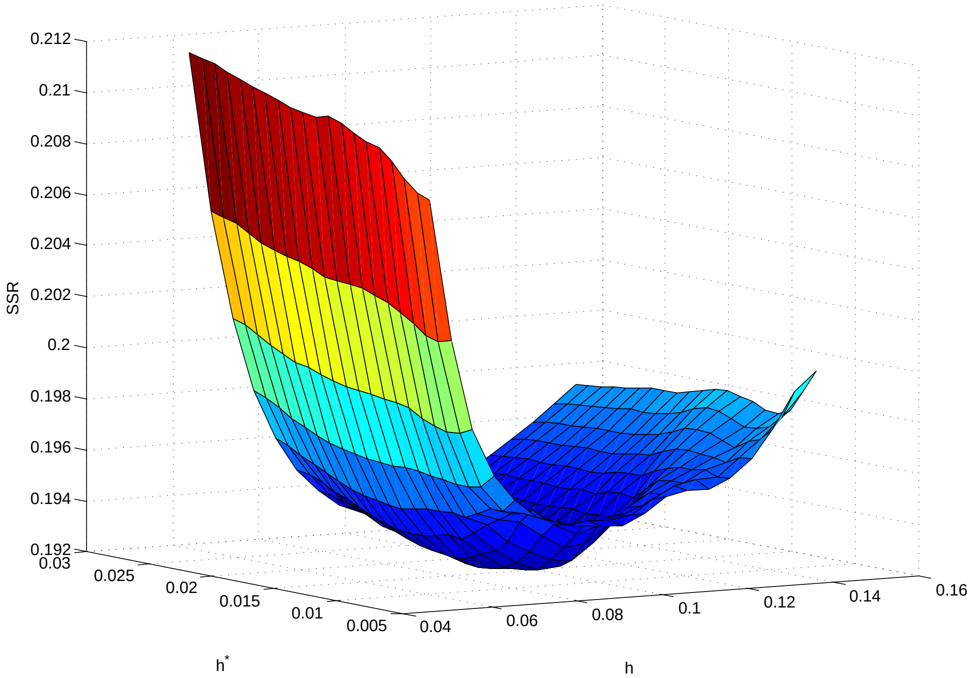


Figure A.6: Sum of Squared Residuals under different bandwidth of h for fixed h^*.

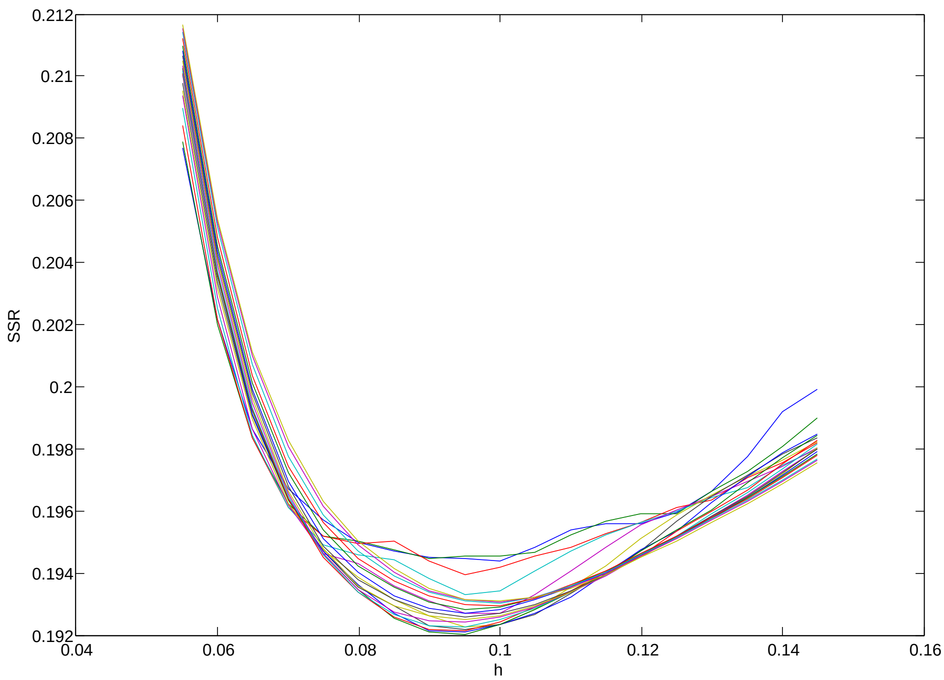


Figure A.7: Sum of Squared Residuals when $h = 0.095$.

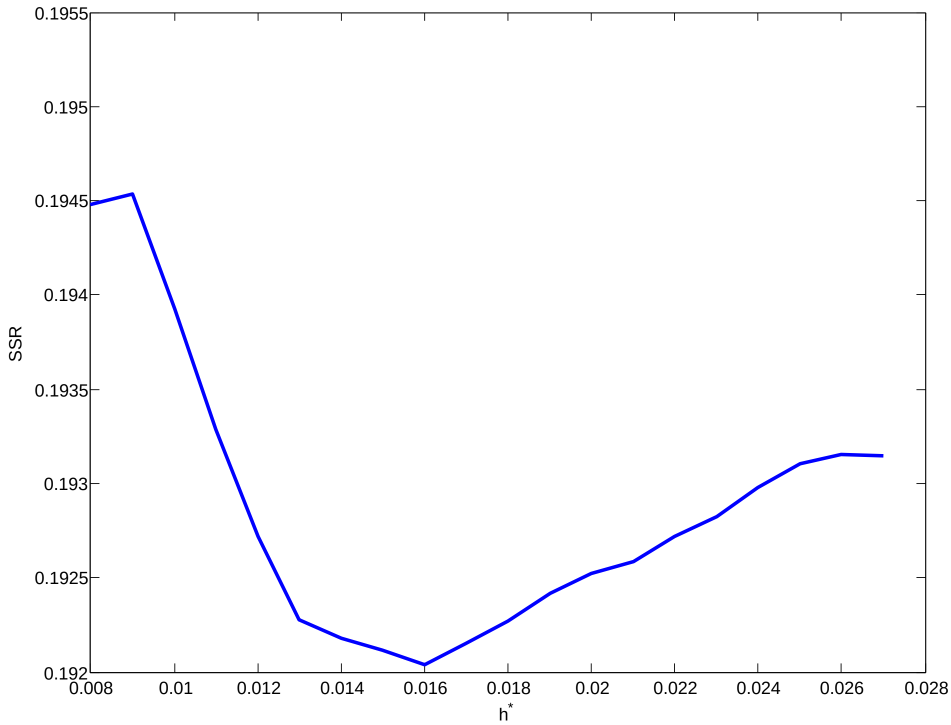


Figure A.8: Estimated control function $\widehat{\lambda}(v_t)$.

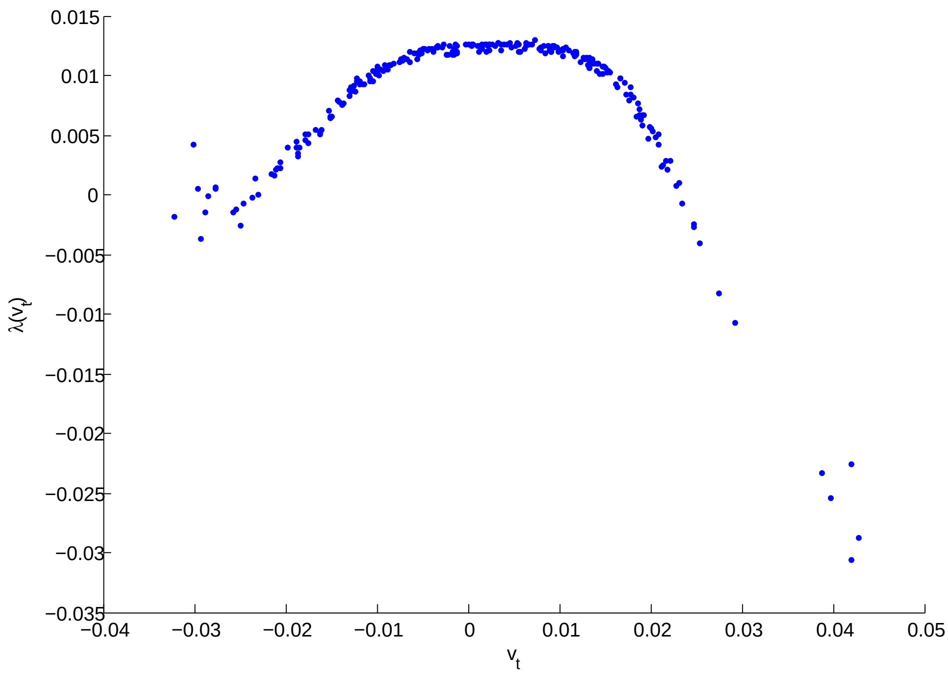
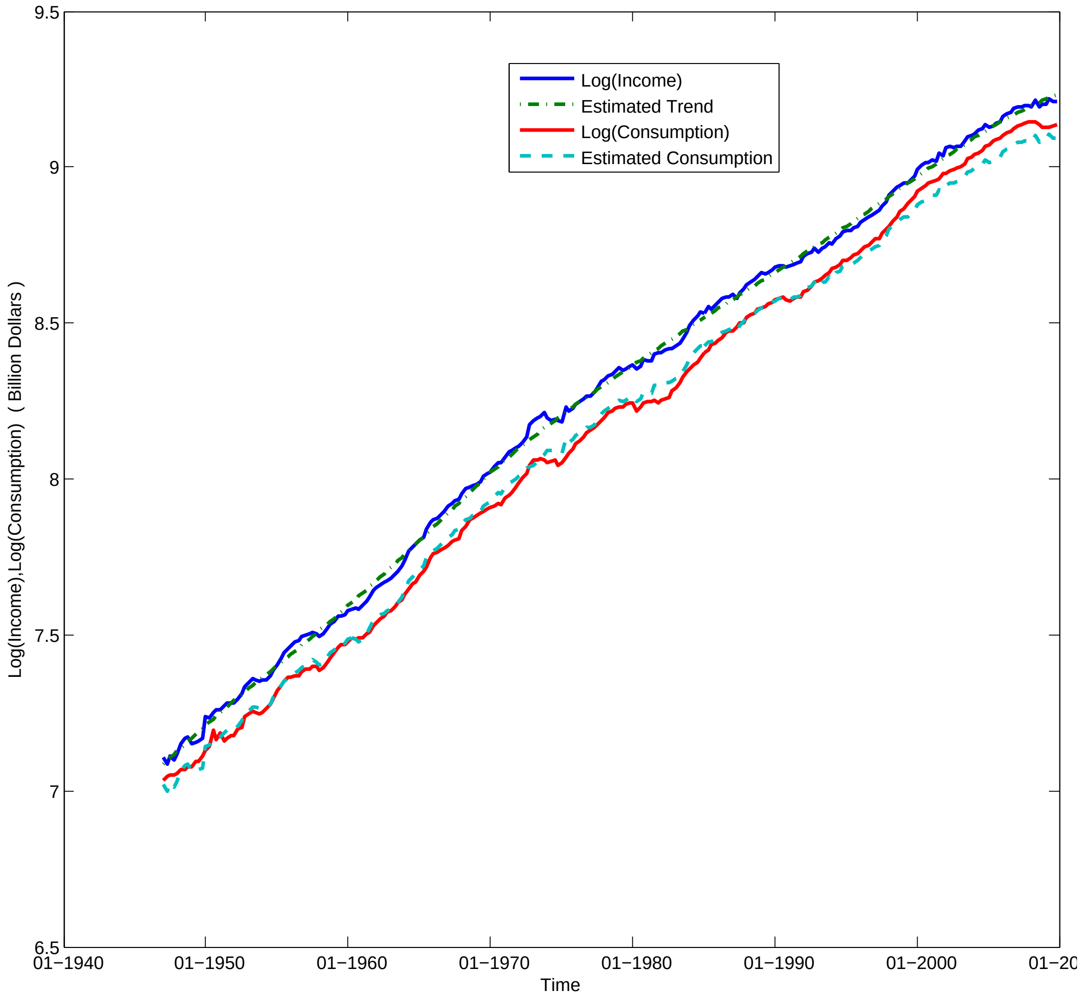


Figure A.9: Income, Consumption, Trend Term, and Fitted Value



Considering structural breaks is another direction, because the relationship between x_t and y_t may vary over time. As in the example we discussed in the thesis, β represents the elasticity between income and consumption. This elasticity, however, may be different in recent years when compared to that 30 years ago. Following this idea, a threshold can be added to the model we discussed. For $t < c$,

$$y_t = x_t'\beta_1 + \lambda_1(v_t) + u_t, \tag{A.33}$$

$$x_t = g_1(\tau_t) + v_t. \tag{A.34}$$

For $t \geq c$,

$$y_t = x_t'\beta_2 + \lambda_2(v_t) + u_t, \tag{A.35}$$

$$x_t = g_2(\tau_t) + v_t. \tag{A.36}$$

The critical issues are to test the existence of structural break(s) and detect the time point at which the structural break occurred.

Appendix B

Proofs of the Theorems

The following Appendices are organized as follows. In Appendix B, I proved the main Theorems in Chapter 2 and 3 for the weak/strong trending regressions by introducing several important Lemmas. Appendix C are the detailed proofs for the Lemmas used in Appendix B. In Section 2 for the weak trend case, I used a special assumption on the density functions of v_t to address it weak dependence across time. That special assumption is verified by a stationary AR(1) process in Appendix D. The proofs of the Theorems in Appendix A are given in Appendix E.

B.1 Proofs of the Theorems in Chapter 2

Some useful Lemmas

We first introduce the following Lemmas that would provide much convenience to the proofs of the Theorems in the thesis.

Lemma B.1.1. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$, let $\overline{g} = \int_0^1 g(\tau)d\tau$, $\overline{g}_i = \int_0^1 g_i(\tau)d\tau$, $\widetilde{g}_i(\tau_t) = g_i(\tau_t) - \sum_{s=1}^n w_{ns}(t)g_i(\tau_s)$, $\widetilde{x}_{it} = x_{it} - \sum_{s=1}^n w_{ns}(t)x_{is}$,

$$M_1(i,j) = \frac{1}{n}\sum_{t=1}^{n}\Big(g_i(\tau_t) - \overline{g}_{i,n}\Big)\Big(g_j(\tau_t) - \overline{g}_{j,n}\Big) \longrightarrow \int_0^1 (g_i(\tau) - \overline{g}_i)(g_j(\tau) - \overline{g}_j)d\tau, \quad \text{(B.1)}$$

$$M_2(i,j) = \frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n} - \sum_{s=1}^{n} w_{ns}(t)g_i(\tau_s)\right)\left(\overline{g}_{j,n} - \sum_{s=1}^{n} w_{ns}(t)g_j(\tau_s)\right) \longrightarrow_P 0, \tag{B.2}$$

$$M_{12}(i,j) = \frac{1}{n}\sum_{t=1}^{n}\left(g_i(\tau_t) - \overline{g}_{i,n}\right)\left(\overline{g}_{j,n} - \sum_{s=1}^{n} w_{ns}(t)g_j(\tau_s)\right) \longrightarrow_P 0, \tag{B.3}$$

$$M_{21}(i,j) = \frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n} - \sum_{s=1}^{n} w_{ns}(t)g_i(\tau_s)\right)\left(g_j(\tau_t) - \overline{g}_{j,n}\right) \longrightarrow_P 0, \tag{B.4}$$

$$S_2(i,j) = \frac{1}{n}\sum_{t=1}^{n}\left(v_{it} - \sum_{s=1}^{n} w_{ns}(t)v_{is}\right)\left(v_{jt} - \sum_{s=1}^{n} w_{ns}(t)v_{js}\right) \longrightarrow_P 0. \tag{B.5}$$

Lemma B.1.2. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$, let $\zeta(v) = \left(\partial\lambda/\partial v_1, \ldots, \partial\lambda/\partial v_k\right)'$ be the first order derivative of $\lambda(v)$ with respect to vector v, then

$$I_{1n} = \left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{v}_t\zeta'(v_t)\widetilde{v}_t\right\| = o_p(1), \quad I_{2n} = \left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{g}(\tau_t)\zeta'(v_t)\widetilde{v}_t\right\| = o_p(1). \tag{B.6}$$

Lemma B.1.3. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$,

$$\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t\overline{u}_t\right\| = o_p(1), \tag{B.7}$$

where $\overline{u}_t = \sum_{s=1}^{n} w_{ns}(t)u_s$.

Lemma B.1.4. Under Assumptions 2.3.1 to 2.3.6, as $n \to \infty$,

$$\left\|\mathfrak{D}_1(n)\right\| = \left\|\frac{1}{n}\sum_{t=1}^{n}(\widehat{x}_t - \widetilde{x}_t)(\widehat{x}_t - \widetilde{x}_t)'\right\| = o_p(1). \tag{B.8}$$

Lemma B.1.5. Let Assumptions 2.3.3 to 2.3.6 hold. As $n \to \infty$, we have

$$\left\|\frac{\widehat{X}'\widehat{e} - \widetilde{X}'\widetilde{e}}{\sqrt{n}}\right\| = o_p(1). \tag{B.9}$$

Proof of Theorem 2.3.1: Recall that x_t is a k-dimensional vector of trending time series sequence, hence $\widetilde{\Sigma}_n$ is a $k \times k$ matrix. To prove $\widetilde{\Sigma}_n \longrightarrow_P Q$, it suffices to show that $\widetilde{\Sigma}_n(i,j) \longrightarrow_P Q(i,j)$, for $i,j = 1,\ldots,k$. Note that $\widetilde{x}_{it} = x_{it} - \sum_{s=1}^{n} w_{ns}(t)x_{is}$ and $x_{it} = g_i(\tau_t) + v_{it}$, therefore,

$$\widetilde{\Sigma}_n(i,j) = \frac{1}{n}\sum_{t=1}^{n}\widetilde{x}_{it}\widetilde{x}_{jt} = \frac{1}{n}\sum_{t=1}^{n}\left(x_{it} - \sum_{s=1}^{n} w_{ns}(t)x_{is}\right)\left(x_{jt} - \sum_{s=1}^{n} w_{ns}(t)x_{js}\right)$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{t=1}^{n}\left(g_i(\tau_t) - \sum_{s=1}^{n} w_{ns}(t) g_i(\tau_s) + v_{it} - \sum_{s=1}^{n} w_{ns}(t) v_{is}\right) \\
&\qquad \times \left(g_j(\tau_t) - \sum_{s=1}^{n} w_{ns}(t) g_j(\tau_s) + v_{jt} - \sum_{s=1}^{n} w_{ns}(t) v_{js}\right) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(g_i(\tau_t) - \sum_{s=1}^{n} w_{ns}(t) g_i(\tau_s)\right)\left(g_j(\tau_t) - \sum_{s=1}^{n} w_{ns}(t) g_j(\tau_s)\right) \\
&\quad + \frac{1}{n}\sum_{t=1}^{n}\left(v_{it} - \sum_{s=1}^{n} w_{ns}(t) v_{is}\right)\left(v_{jt} - \sum_{s=1}^{n} w_{ns}(t) v_{js}\right) \\
&\quad + \frac{1}{n}\sum_{t=1}^{n}\left(g_i(\tau_t) - \sum_{s=1}^{n} w_{ns}(t) g_i(\tau_s)\right)\left(v_{jt} - \sum_{s=1}^{n} w_{ns}(t) v_{js}\right) \\
&\quad + \frac{1}{n}\sum_{t=1}^{n}\left(v_{it} - \sum_{s=1}^{n} w_{ns}(t) v_{is}\right)\left(g_j(\tau_t) - \sum_{s=1}^{n} w_{ns}(t) g_j(\tau_s)\right) \\
&\triangleq \frac{1}{n}\sum_{t=1}^{n}\widetilde{g}_i(\tau_t)\widetilde{g}_j(\tau_t) + \frac{1}{n}\sum_{t=1}^{n}\widetilde{v}_{it}\widetilde{v}_{jt} + \frac{1}{n}\sum_{t=1}^{n}\widetilde{g}_i(\tau_t)\widetilde{v}_{jt} + \frac{1}{n}\sum_{t=1}^{n}\widetilde{v}_{it}\widetilde{g}_j(\tau_t) \\
&\triangleq S_1(i,j) + S_2(i,j) + S_{12}(i,j) + S_{21}(i,j).
\end{aligned} \tag{B.10}
$$

In the above equations, we define

$$
\begin{aligned}
S_1(i,j) &= \frac{1}{n}\sum_{t=1}^{n}\widetilde{g}_i(\tau_t)\widetilde{g}_j(\tau_t), \quad S_2(i,j) = \frac{1}{n}\sum_{t=1}^{n}\widetilde{v}_{it}\widetilde{v}_{jt}, \\
S_{12}(i,j) &= \frac{1}{n}\sum_{t=1}^{n}\widetilde{g}_i(\tau_t)\widetilde{v}_{jt}, \quad S_{21}(i,j) = \frac{1}{n}\sum_{t=1}^{n}\widetilde{v}_{it}\widetilde{g}_j(\tau_t),
\end{aligned} \tag{B.11}
$$

where $\widetilde{v}_{it} = v_{it} - \sum_{s=1}^{n} w_{ns}(t) v_{is}$, for $i,j = 1,2,\ldots,k$, $t = 1,2,\ldots,n$. Let $\overline{g}_{i,n} = n^{-1}\sum_{t=1}^{n} g_i(\tau_t)$ for $i = 1,2,\ldots,k$. Therefore, $\overline{g}_{i,n}$ denotes the sample average of the trend component of x_i. We further decompose $S_1(i,j)$ as

$$
\begin{aligned}
S_1(i,j) &= \frac{1}{n}\sum_{t=1}^{n}\left(g_i(\tau_t) - \sum_{s=1}^{n} w_{ns}(t) g_i(\tau_s)\right)\left(g_j(\tau_t) - \sum_{s=1}^{n} w_{ns}(t) g_j(\tau_s)\right) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(g_i(\tau_t) - \overline{g}_{i,n} + \overline{g}_{i,n} - \sum_{s=1}^{n} w_{ns}(t) g_i(\tau_s)\right)\left(g_j(\tau_t) - \overline{g}_{j,n} + \overline{g}_{j,n} - \sum_{s=1}^{n} w_{ns}(t) g_j(\tau_s)\right) \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(g_i(\tau_t) - \overline{g}_{i,n}\right)\left(g_j(\tau_t) - \overline{g}_{j,n}\right) + \frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n} - \sum_{s=1}^{n} w_{ns}(t) g_i(\tau_s)\right)\left(\overline{g}_{j,n} - \sum_{s=1}^{n} w_{ns}(t) g_j(\tau_s)\right)
\end{aligned}
$$

$$+\frac{1}{n}\sum_{t=1}^{n}\Big(g_i(\tau_t)-\overline{g}_{i,n}\Big)\left(\overline{g}_{j,n}-\sum_{s=1}^{n}w_{ns}(t)g_j(\tau_s)\right)+\frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n}-\sum_{s=1}^{n}w_{ns}(t)g_i(\tau_s)\right)\Big(g_j(\tau_t)-\overline{g}_{j,n}\Big)$$

$$\triangleq M_1(i,j)+M_2(i,j)+M_{12}(i,j)+M_{21}(i,j), \tag{B.12}$$

where

$$M_1(i,j)=\frac{1}{n}\sum_{t=1}^{n}\Big(g_i(\tau_t)-\overline{g}_{i,n}\Big)\Big(g_j(\tau_t)-\overline{g}_{j,n}\Big), \tag{B.13}$$

$$M_2(i,j)=\frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n}-\sum_{s=1}^{n}w_{ns}(t)g_i(\tau_s)\right)\left(\overline{g}_{j,n}-\sum_{s=1}^{n}w_{ns}(t)g_j(\tau_s)\right), \tag{B.14}$$

$$M_{12}(i,j)=\frac{1}{n}\sum_{t=1}^{n}\Big(g_i(\tau_t)-\overline{g}_{i,n}\Big)\left(\overline{g}_{j,n}-\sum_{s=1}^{n}w_{ns}(t)g_j(\tau_s)\right), \tag{B.15}$$

$$M_{21}(i,j)=\frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n}-\sum_{s=1}^{n}w_{ns}(t)g_i(\tau_s)\right)\Big(g_j(\tau_t)-\overline{g}_{j,n}\Big). \tag{B.16}$$

Therefore, by Lemma B.1.1,

$$S_1(i,j)=M_1(i,j)+M_2(i,j)+M_{12}(i,j)+M_{21}(i,j)\longrightarrow_P\int_0^1(g_i(\tau)-\overline{g}_i)(g_j(\tau)-\overline{g}_j)d\tau, \tag{B.17}$$

$$S_2(i,j)\longrightarrow_P 0. \tag{B.18}$$

Meanwhile, by Cauchy-Schwarz inequality, given (B.17) and (B.18),

$$\begin{aligned}|S_{12}(i,j)|=&\left|n^{-1}\sum_{t=1}^{n}\left(g_i(\tau_t)-\sum_{s=1}^{n}w_{ns}(t)g_i(\tau_s)\right)\left(v_{jt}-\sum_{s=1}^{n}w_{ns}(t)v_{js}\right)\right|\\ \le&\left|n^{-1}\sum_{t=1}^{n}\left(g_i(\tau_t)-\sum_{s=1}^{n}w_{ns}(t)g_i(\tau_s)\right)^2\right|^{1/2}\left|n^{-1}\sum_{t=1}^{n}\left(v_{jt}-\sum_{s=1}^{n}w_{ns}(t)v_{js}\right)^2\right|^{1/2}\\ =&|S_1(i,i)|^{1/2}|S_2(j,j)|^{1/2}=O_p(1)o_p(1)=o_p(1).\end{aligned} \tag{B.19}$$

The same result holds for $S_{21}(i,j)$. Hence, for $i,j=1,\ldots,k$,

$$\frac{1}{n}\sum_{t=1}^{n}\widetilde{x}_{it}\widetilde{x}_{jt}\longrightarrow_P\int_0^1(g_i(\tau)-\overline{g}_i)(g_j(\tau)-\overline{g}_j)d\tau, \tag{B.20}$$

i.e., $\widetilde{\Sigma}_n(i,j)\longrightarrow_P Q(i,j)$. Therefore, the convergence of every element in $\widetilde{\Sigma}_n$ yields the convergence of the whole matrix that as $n\to\infty$, $\widetilde{\Sigma}_n\longrightarrow_P Q$. ■

Proof of Lemma 2.3.1: Note that

$$\sqrt{n}B_n = \left(\frac{1}{n}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{\lambda}(v_t)\right). \tag{B.21}$$

Theorem 2.3.1 shows that $n^{-1}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t' \longrightarrow_P Q$, hence $\left(n^{-1}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t'\right)^{-1} \longrightarrow_P Q^{-1}$, which is $O_p(1)$. Without loss of generality, we assume that $k = 1$. Our objective becomes showing that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{\lambda}(v_t) = o_p(1). \tag{B.22}$$

Note that $\widetilde{x}_t$ can be written as

$$\begin{aligned}\widetilde{x}_t &= x_t - \sum_{s=1}^{n} w_{ns}(t)x_s = v_t + g(\tau_t) - \sum_{s=1}^{n} w_{ns}(t)v_s - \sum_{s=1}^{n} w_{ns}(t)g(\tau_s)\\ &= v_t - \sum_{s=1}^{n} w_{ns}(t)v_s + g(\tau_t) - \sum_{s=1}^{n} w_{ns}(t)g(\tau_s) = \widetilde{v}_t + \widetilde{g}(\tau_t),\end{aligned} \tag{B.23}$$

where $\widetilde{v}_t = v_t - \sum_{s=1}^{n} w_{ns}(t)v_s$ and $\widetilde{g}(\tau_t) = g(\tau_t) - \sum_{s=1}^{n} w_{ns}(t)g(\tau_s)$.
Meanwhile, $\widetilde{\lambda}(v_t)$ can be written as

$$\widetilde{\lambda}(v_t) = \lambda(v_t) - \sum_{s=1}^{n} w_{ns}(t)\lambda(v_s) = \sum_{s=1}^{n} w_{ns}(t)\Big(\lambda(v_t) - \lambda(v_s)\Big). \tag{B.24}$$

Apply the Taylor expansion for $\lambda(v_s)$,

$$\lambda(v_s) = \lambda(v_t) + \lambda^{(1)}(v_t)(v_s - v_t) + \frac{1}{2}\lambda^{(2)}(v_t^*)(v_s - v_t)^2, \tag{B.25}$$

where v_t^* lies between v_t and v_s, and $\lambda^{(2)}(v_t^*) < \infty$ for any t. Therefore,

$$\begin{aligned}\widetilde{\lambda}(v_t) &= \sum_{s=1}^{n} w_{ns}(t)\Big(\lambda(v_t) - \lambda(v_s)\Big) = \sum_{s=1}^{n} w_{ns}(t)\left(\lambda^{(1)}(v_t)(v_t - v_s) - \frac{1}{2}\lambda^{(2)}(v_t^*)(v_s - v_t)^2\right)\\ &= \lambda^{(1)}(v_t)\sum_{s=1}^{n} w_{ns}(t)(v_t - v_s) - \frac{1}{2}\lambda^{(2)}(v_t^*)\sum_{s=1}^{n} w_{ns}(t)(v_t - v_s)^2\\ &= \lambda^{(1)}(v_t)\widetilde{v}_t - \frac{1}{2}\lambda^{(2)}(v_t^*)\sum_{s=1}^{n} w_{ns}(t)(v_s - v_t)^2.\end{aligned} \tag{B.26}$$

Substitute $\widetilde{x}_t$ in (B.22) with (B.23) and $\widetilde{\lambda}(v_t)$ with its leading term in (B.26), denote the first order derivative of $\lambda(v_t)$ as $\zeta(v_t)$,

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{\lambda}(v_t) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{v}_t\widetilde{\lambda}(v_t) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{g}(\tau_t)\widetilde{\lambda}(v_t)$$

$$
\begin{aligned}
&=\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\zeta(v_t)\widetilde{v}_t^2+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{g}(\tau_t)\zeta(v_t)\widetilde{v}_t\right)\Big(1+o_p(1)\Big)\\
&=(I_{1n}+I_{2n})\Big(1+o_p(1)\Big),
\end{aligned}
\tag{B.27}
$$

where we defined $I_{1n}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\zeta(v_t)\widetilde{v}_t^2$ and $I_{2n}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{g}(\tau_t)\zeta(v_t)\widetilde{v}_t$. Therefore, we complete the proof provided that Lemma B.1.2 holds. The proof of Lemma B.1.2 is provided in Appendix B. ■

Proof of Theorem 2.3.2: Substituting $\widetilde{y}_t$ in $\widetilde{\beta}$ using $\widetilde{y}_t=\widetilde{x}_t'\beta+\widetilde{\lambda}(v_t)+\widetilde{u}_t$, we have

$$
\begin{aligned}
\widetilde{\beta}&=\left(\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t'\right)^{-1}\left(\sum_{t=1}^{n}\widetilde{x}_t\left(\widetilde{x}_t'\beta+\widetilde{\lambda}(v_t)+\widetilde{u}_t\right)\right)\\
&=\beta+\left(\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t'\right)^{-1}\left(\sum_{t=1}^{n}\widetilde{x}_t\widetilde{\lambda}(v_t)\right)+\left(\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t'\right)^{-1}\left(\sum_{t=1}^{n}\widetilde{x}_t\widetilde{u}_t\right)\\
&=\beta+B_n+\left(\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t'\right)^{-1}\left(\sum_{t=1}^{n}\widetilde{x}_t\widetilde{u}_t\right),
\end{aligned}
\tag{B.28}
$$

where $B_n=\left(\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t'\right)^{-1}\left(\sum_{t=1}^{n}\widetilde{x}_t\widetilde{\lambda}(v_t)\right)$. Therefore,

$$
\sqrt{n}\left(\widetilde{\beta}-\beta-B_n\right)=\left(\frac{1}{n}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{u}_t\right).
\tag{B.29}
$$

In the previous proofs of Theorem 2.3.1, we have shown that

$$
\frac{1}{n}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t'\longrightarrow_P Q,
\tag{B.30}
$$

where Q is assumed to be positive definite, therefore invertible. For the latter part,

$$
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{u}_t=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t(u_t-\overline{u}_t)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_tu_t+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t\overline{u}_t,
\tag{B.31}
$$

where $\overline{u}_t=\sum_{s=1}^{n}w_{ns}(t)u_s$. Under Assumptions 2.3.3 to 2.3.6, since u_t is mixing and independent with v_t, by Central Limit Theorem (CLT) for mixing processes (see Fan and Yao (2003), Theorem 2.21), as $n\to\infty$,

$$
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_tu_t\longrightarrow_D N(0,Q\Lambda_u),
\tag{B.32}
$$

where Λ_u is the long-run variances of u_t that

$$\Lambda_u = \mathrm{E}\left[u_t u_t\right] + 2\sum_{j=1}^{\infty} \mathrm{E}\left[u_t u_{t-j}\right]. \tag{B.33}$$

Meanwhile, by Lemma B.1.3, we have

$$\left\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \widetilde{x}_t \overline{u}_t\right\| = o_p(1). \tag{B.34}$$

Then, by the Slutsky Theorem, we have

$$\sqrt{n}\left(\widetilde{\beta} - \beta - B_n\right) \longrightarrow_D N(0, \Omega), \tag{B.35}$$

where $\Omega = Q^{-1}\Lambda_u$. By Lemma 2.3.1, the bias term B_n is negligible that $\sqrt{n}B_n = o_p(1)$ as $n \to \infty$. Thus, we can ignore the potential bias term and yield

$$\sqrt{n}(\widetilde{\beta} - \beta) \longrightarrow_D N(0, \Omega). \quad \blacksquare \tag{B.36}$$

Proof of Lemma 2.3.2 and Theorem 2.3.3: We focus on the difference between $\widetilde{\Sigma}_n$ and $\widehat{\Sigma}_n$. Write

$$\begin{aligned}
\widehat{\Sigma}_n =& \frac{1}{n}\sum_{t=1}^{n} \widehat{x}_t \widehat{x}_t' = \frac{1}{n}\sum_{t=1}^{n} (\widehat{x}_t - \widetilde{x}_t + \widetilde{x}_t)(\widehat{x}_t - \widetilde{x}_t + \widetilde{x}_t)' \\
=& \frac{1}{n}\sum_{t=1}^{n} (\widehat{x}_t - \widetilde{x}_t)(\widehat{x}_t - \widetilde{x}_t)' + \frac{1}{n}\sum_{t=1}^{n} (\widehat{x}_t - \widetilde{x}_t)\widetilde{x}_t' + \frac{1}{n}\sum_{t=1}^{n} \widetilde{x}_t(\widehat{x}_t - \widetilde{x}_t)' + \frac{1}{n}\sum_{t=1}^{n} \widetilde{x}_t \widetilde{x}_t',
\end{aligned} \tag{B.37}$$

where the last term is $\widetilde{\Sigma}_n$. Therefore,

$$\begin{aligned}
\widehat{\Sigma}_n - \widetilde{\Sigma}_n =& \frac{1}{n}\sum_{t=1}^{n} (\widehat{x}_t - \widetilde{x}_t)(\widehat{x}_t - \widetilde{x}_t)' + \frac{1}{n}\sum_{t=1}^{n} (\widehat{x}_t - \widetilde{x}_t)\widetilde{x}_t' + \frac{1}{n}\sum_{t=1}^{n} \widetilde{x}_t(\widehat{x}_t - \widetilde{x}_t)' \\
=& \mathcal{D}_1(n) + \mathcal{D}_2(n) + \mathcal{D}_3(n),
\end{aligned} \tag{B.38}$$

where $\mathcal{D}_1(n) = n^{-1}\sum_{t=1}^{n} (\widehat{x}_t - \widetilde{x}_t)(\widehat{x}_t - \widetilde{x}_t)'$, $\mathcal{D}_2(n) = n^{-1}\sum_{t=1}^{n} (\widehat{x}_t - \widetilde{x}_t)\widetilde{x}_t'$, and $\mathcal{D}_3(n) = n^{-1}\sum_{t=1}^{n} \widetilde{x}_t(\widehat{x}_t - \widetilde{x}_t)'$. We control the difference by

$$\left\|\widehat{\Sigma}_n - \widetilde{\Sigma}_n\right\| = \left\|\mathcal{D}_1(n) + \mathcal{D}_2(n) + \mathcal{D}_3(n)\right\| \leq \|\mathcal{D}_1(n)\| + \|\mathcal{D}_2(n)\| + \|\mathcal{D}_3(n)\|. \tag{B.39}$$

Hence, $\|\widehat{\Sigma}_n - \widetilde{\Sigma}_n\|$ converges to zeros in probability if $\|\mathfrak{D}_l(n)\| \longrightarrow_P 0$ for $l = 1,2,3$. By Cauchy-Schwarz inequality,

$$\begin{aligned}\|\mathfrak{D}_2(n)\| = &\left\| \frac{1}{n}\sum_{t=1}^{n}(\widehat{x}_t - \widetilde{x}_t)\widetilde{x}_t' \right\| \leq \left\| \frac{1}{n}\sum_{t=1}^{n}(\widehat{x}_t - \widetilde{x}_t)(\widehat{x}_t - \widetilde{x}_t)' \right\|^{1/2} \left\| \frac{1}{n}\sum_{t=1}^{n}\widetilde{x}_t\widetilde{x}_t' \right\|^{1/2}\\ =&\|\mathfrak{D}_1(n)\|^{1/2}\|\widetilde{\Sigma}_n\|^{1/2}.\end{aligned} \tag{B.40}$$

By Lemma B.1.4, $\|\mathfrak{D}_1(n)\| = o_p(1)$, and we have shown that the limit of $\|\widetilde{\Sigma}_n\|$ is finite, therefore, $\|\mathfrak{D}_2(n)\| = o_p(1)$. We can also show $\|\mathfrak{D}_3(n)\| = o_p(1)$ using the same method. Thus $\|\widehat{\Sigma}_n - \widetilde{\Sigma}_n\| \rightarrow_P 0$ as $n \rightarrow \infty$. Therefore, we complete the proof of Lemma 2.3.2 and Theorem 2.3.3. ■

Proof of Lemma 2.3.3 and Theorem 3.4.1: Note that

$$\sqrt{n}(\widehat{\beta} - \beta) = \sqrt{n}(\widehat{\beta} - \widetilde{\beta} + \widetilde{\beta} - \beta) = \sqrt{n}(\widehat{\beta} - \widetilde{\beta}) + \sqrt{n}(\widetilde{\beta} - \beta), \tag{B.41}$$

where we have shown (B.36). Therefore, we can complete the proof once Lemma 2.3.3 is proved in which

$$\left\|\sqrt{n}(\widehat{\beta} - \widetilde{\beta})\right\| = o_p(1). \tag{B.42}$$

Note that $\widehat{Y} = \widehat{X}\beta + \widehat{\lambda}(V) + \widehat{U}$, $\widetilde{Y} = \widetilde{X}\beta + \widetilde{\lambda}(V) + \widetilde{U}$, $\widehat{e} = \widehat{\lambda}(\widehat{V}) + \widehat{U}$ and $\widetilde{e} = \widetilde{\lambda}(V) + \widetilde{U}$. Here, $\widehat{Y} = (\widehat{y}_1, ..., \widehat{y}_n)'$, $\widehat{X} = (\widehat{x}_1, ..., \widehat{x}_n)'$, $\widehat{\lambda}(V) = (\widehat{\lambda}(v_1), ..., \widehat{\lambda}(v_n))'$, $\widehat{U} = (\widehat{u}_1, ..., \widehat{u}_n)'$, $\widehat{e} = (\widehat{e}_1, ..., \widehat{e}_n)'$ and $\widehat{y}_t = y_t - \widehat{\mathrm{E}}_h[y_t|\widehat{v}_t]$, $\widehat{x}_t = x_t - \widehat{\mathrm{E}}_h[x_t|\widehat{v}_t]$, $\widehat{\lambda}(v_t) = \lambda(v_t) - \widehat{\mathrm{E}}_h[\lambda(v_t)|\widehat{v}_t]$, $\widehat{u}_t = u_t - \widehat{\mathrm{E}}_h[u_t|\widehat{v}_t]$, $\widehat{e}_t = e_t - \widehat{\mathrm{E}}_h[e_t|\widehat{v}_t]$.

$$\begin{aligned}\widehat{\beta} - \widetilde{\beta} =&\widehat{\beta} - \beta - (\widetilde{\beta} - \beta) = (\widehat{X}'\widehat{X})^{-1}(\widehat{X}'\widehat{Y}) - \beta - ((\widetilde{X}'\widetilde{X})^{-1}(\widetilde{X}'\widetilde{Y}) - \beta)\\ =&(\widehat{X}'\widehat{X})^{-1}(\widehat{X}'\widehat{e}) - (\widetilde{X}'\widetilde{X})^{-1}(\widetilde{X}'\widetilde{e})\\ =&(\widehat{X}'\widehat{X})^{-1}(\widehat{X}'\widehat{e}) - (\widehat{X}'\widehat{X})^{-1}(\widetilde{X}'\widetilde{e}) + (\widehat{X}'\widehat{X})^{-1}(\widetilde{X}'\widetilde{e}) - (\widetilde{X}'\widetilde{X})^{-1}(\widetilde{X}'\widetilde{e})\\ =&(\widehat{X}'\widehat{X})^{-1}(\widehat{X}'\widehat{e} - \widetilde{X}'\widetilde{e}) + \left((\widehat{X}'\widehat{X})^{-1} - (\widetilde{X}'\widetilde{X})^{-1}\right)(\widetilde{X}'\widetilde{e}).\end{aligned} \tag{B.43}$$

Then our goal is to show that the norm of the following equation is $o_p(1)$.

$$\sqrt{n}\left(\widehat{\beta} - \widetilde{\beta}\right) = \left(\frac{\widehat{X}'\widehat{X}}{n}\right)^{-1}\left(\frac{\widehat{X}'\widehat{e} - \widetilde{X}'\widetilde{e}}{\sqrt{n}}\right) + \left(\left(\frac{\widehat{X}'\widehat{X}}{n}\right)^{-1} - \left(\frac{\widetilde{X}'\widetilde{X}}{n}\right)^{-1}\right)\left(\frac{\widetilde{X}'\widetilde{e}}{\sqrt{n}}\right)$$

$$\triangleq DB_1 + DB_2. \tag{B.44}$$

We first examine DB_1. By Lemma B.1.5, $\left\|(\widehat{X}'\widehat{e} - \widetilde{X}'\widetilde{e})/\sqrt{n}\right\| = o_p(1)$ and as shown in the first Theorem, $\widehat{\Sigma}_n = \frac{\widehat{X}'\widehat{X}}{n} \rightarrow_P Q$, where Q is positive definite, hence $\widehat{\Sigma}_n^{-1} \rightarrow_P Q^{-1}$. Then we have

$$DB_1 = O_p(1)o_p(1) = o_p(1). \tag{B.45}$$

We then examine DB_2. By the second step in the proofs of Theorem 2.3.3, $\|\widehat{\Sigma}_n - \widetilde{\Sigma}_n\| \rightarrow_P 0$, therefore,

$$\left(\frac{\widehat{X}'\widehat{X}}{n}\right)^{-1} - \left(\frac{\widetilde{X}'\widetilde{X}}{n}\right)^{-1} = \left(\frac{\widehat{X}'\widehat{X}}{n}\right)^{-1}\left[\frac{\widetilde{X}'\widetilde{X}}{n} - \frac{\widehat{X}'\widehat{X}}{n}\right]\left(\frac{\widetilde{X}'\widetilde{X}}{n}\right)^{-1} = o_p(1). \tag{B.46}$$

Meanwhile, in the first step, we have shown that $\widetilde{X}'\widetilde{e}/\sqrt{n} = O_p(1)$. Therefore,

$$DB_2 = o_p(1). \tag{B.47}$$

To summarize, equations (B.45) and (B.47) imply (B.42). Then, together with (B.36), we are able to complete the proof. ■

B.2 Proofs of the Theorems in Chapter 3

Proof of Theorem 3.2.1: The estimator $\widehat{Q}$ is defined as $D^{-1}\boldsymbol{X}'\boldsymbol{X}D^{-1}$, then let $d_{ij} = (d_i + d_j)/2 > 1$ and

$$\begin{aligned}\widehat{Q}_{ij} =& n^{-\frac{d_i+d_j}{2}} \sum_{t=1}^{n} x_{it}x_{jt} = n^{-\frac{d_i+d_j}{2}} \sum_{t=1}^{n} (g_i(t) + v_{it})(g_j(t) + v_{jt}) \\ =& \frac{1}{n^{d_{ij}}} \sum_{t=1}^{n} g_i(t)g_j(t) + \frac{1}{n^{d_{ij}}} \sum_{t=1}^{n} g_i(t)v_{jt} + \frac{1}{n^{d_{ij}}} \sum_{t=1}^{n} g_j(t)v_{it} + \frac{1}{n^{d_{ij}}} \sum_{t=1}^{n} v_{it}v_{jt},\end{aligned} \tag{B.48}$$

for $i, j = 1, 2, ..., K$. The first term is deterministic and by Assumption 3.2.1, as $n \rightarrow \infty$,

$$\frac{1}{n^{d_{ij}}} \sum_{t=1}^{n} g_i(t)g_j(t) \longrightarrow Q_{ij}. \tag{B.49}$$

For the last term,

$$\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n} v_{it}v_{jt} = \frac{1}{n^{d_{ij}-1}}\left(\frac{1}{n}\sum_{t=1}^{n} v_{it}v_{jt}\right), \tag{B.50}$$

where $d_{ij}-1>0$ and $n^{-1}\sum_{t=1}^{n} v_{it}v_{jt} = O_P(1)$. In fact, by Cauchy-Schwarz Inequality,

$$\left|\frac{1}{n}\sum_{t=1}^{n} v_{it}v_{jt}\right| \le \left(\frac{1}{n}\sum_{t=1}^{n} v_{it}^2\right)^{1/2}\left(\frac{1}{n}\sum_{t=1}^{n} v_{jt}^2\right)^{1/2}, \tag{B.51}$$

where

$$\begin{aligned}\mathrm{E}\left[\frac{1}{n}\sum_{t=1}^{n} v_{it}^2\right] =& \mathrm{E}\left[\frac{1}{n}\sum_{t=1}^{n}\left(\sum_{s=0}^{\infty}\psi_{s,i}\eta_{i,t-s}\right)^2\right] = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{\infty}\sum_{l=0}^{\infty}\psi_{s,i}\psi_{l,i}\mathrm{E}\left[\eta_{i,t-s}\eta_{i,t-l}\right]\\ =&\frac{1}{n}\sum_{t=1}^{n}\sum_{s=0}^{\infty}\psi_{s,i}^2\mathrm{E}\left[\eta_{i,t-s}^2\right] = \sigma_{ii}\sum_{s=0}^{\infty}\psi_{s,i}^2 < \infty.\end{aligned} \tag{B.52}$$

Therefore, as $n\to\infty$, the last term

$$\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n} v_{it}v_{jt} = O_P(n^{1-d_{ij}}) = o_P(1). \tag{B.53}$$

The second and third terms are similar, so we only study one of them. By Cauchy-Schwarz Inequality, it follows that

$$\begin{aligned}\left|\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n} g_i(t)v_{jt}\right| =& \left|\frac{1}{n}\sum_{t=1}^{n}\left(\frac{g_i(t)}{n^{\frac{d_i-1}{2}}}\right)\left(\frac{v_{jt}}{n^{\frac{d_j-1}{2}}}\right)\right|\\ \le& \left|\frac{1}{n}\sum_{t=1}^{n}\left(\frac{g_i(t)}{n^{\frac{d_i-1}{2}}}\right)^2\right|^{1/2}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\frac{v_{jt}}{n^{\frac{d_j-1}{2}}}\right)^2\right|^{1/2}\\ =& \left|\frac{1}{n^{d_i}}\sum_{t=1}^{n} g_i(t)^2\right|^{1/2}\left|\frac{1}{n^{d_j}}\sum_{t=1}^{n} v_{jt}^2\right|^{1/2},\end{aligned} \tag{B.54}$$

in which $n^{-d_i}\sum_{t=1}^{n} g_i(t)^2 \longrightarrow Q_{ii} < \infty$ and $n^{-d_j}\sum_{t=1}^{n} v_{jt}^2 = O_P(n^{1-d_j}) = o_P(1)$. Therefore, the second term is $o_P(1)$, so is the third term. To summarize,

$$\widehat{Q}_{ij} \longrightarrow_P Q_{ij}, \tag{B.55}$$

for all $i,j=1,2,...,k$. Therefore, $\widehat{Q} \longrightarrow_P Q$ as $n\to\infty$, in which Q is assumed to be a positive definite matrix.

Proof of Theorem 3.3.1: Recall that the OLS estimator is defined as

$$\widehat{\beta}_{ols} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'y\right). \tag{B.56}$$

Therefore,

$$D\left(\widehat{\beta}_{ols} - \beta\right) = \left(D^{-1}\boldsymbol{X}'\boldsymbol{X}D^{-1}\right)^{-1}\left(D^{-1}\boldsymbol{X}'\boldsymbol{e}\right). \tag{B.57}$$

We have already proved in Theorem 3.2.1 that $D^{-1}\boldsymbol{X}'\boldsymbol{X}D^{-1} \longrightarrow_P Q$ as $n \to \infty$, where Q is a positive definite matrix.

Let $\boldsymbol{b} = \mathrm{E}[e_t\boldsymbol{v}_t]$ and $\boldsymbol{b} = (b_1, \ldots, b_k)'$ where for $i = 1, \ldots, k$,

$$b_i = \mathrm{E}[e_t v_{it}] = \mathrm{E}\left[\sum_{s=0}^{\infty}\phi_s\epsilon_{t-s}\sum_{l=0}^{\infty}\psi_{l,i}\eta_{i,t-l}\right] = \sum_{s=0}^{\infty}\phi_s\psi_{s,i}\mathrm{E}\left[\epsilon_{t-s}\eta_{i,t-s}\right] = \theta_i\sum_{s=0}^{\infty}\phi_s\psi_{s,i}, \tag{B.58}$$

where $\theta_i = \mathrm{E}[\epsilon_t\eta_{it}]$, which is the i^{th} element of the $k \times 1$ vector $\boldsymbol{\Theta}$. We then focus on $D^{-1}\boldsymbol{X}'\boldsymbol{e}$. Denote

$$D^{-1}\left(\boldsymbol{X}'\boldsymbol{e} - n\boldsymbol{b}\right) \triangleq \boldsymbol{Z}_2. \tag{B.59}$$

Then the i^{th} element of the $k \times 1$ vector $\boldsymbol{Z}_2$ is

$$\begin{aligned}
\boldsymbol{Z}_{2,i} =& n^{-\frac{d_i}{2}}\sum_{t=1}^{n}(x_{it}e_t - b_i) = n^{-\frac{d_i}{2}}\sum_{t=1}^{n}g_i(t)e_t + n^{-\frac{d_i}{2}}\sum_{t=1}^{n}(v_{it}e_t - b_i)\\
=& n^{-\frac{d_i}{2}}\sum_{t=1}^{n}g_i(t)e_t + n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\left(\sum_{s=0}^{\infty}\sum_{l=0}^{\infty}\phi_s\psi_{l,i}\epsilon_{t-s}\eta_{i,t-l} - \sum_{s=0}^{\infty}\phi_s\psi_{s,i}\theta_i\right)\\
=& n^{-\frac{d_i}{2}}\sum_{t=1}^{n}g_i(t)\sum_{s=0}^{\infty}\phi_s\epsilon_{t-s} + n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{s=0}^{\infty}\phi_s\psi_{s,i}\left(\epsilon_{t-s}\eta_{i,t-s} - \theta_i\right)\\
&+ n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{s=0}^{\infty}\sum_{l=s+1}^{\infty}\phi_s\psi_{l,i}\epsilon_{t-s}\eta_{i,t-l} + n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{l=0}^{\infty}\sum_{s=l+1}^{\infty}\phi_s\psi_{l,i}\epsilon_{t-s}\eta_{i,t-l}\\
\triangleq& P_1(n) + P_2(n) + P_3(n) + P_4(n),
\end{aligned} \tag{B.60}$$

for $i = 1, \ldots, K$, where the last three terms correspond to $s = l$, $s < l$ and $s > l$ respectively. Then for $P_1(n)$, as $e_t = \Phi(L)\epsilon_t$, by Beveridge-Nelson(BN) decomposition, we have

$$P_1(n) = n^{-\frac{d_i}{2}}\sum_{t=1}^{n}g_i(t)\Phi(L)\epsilon_t = n^{-\frac{d_i}{2}}\sum_{t=1}^{n}g_i(t)\Phi(1)\epsilon_t - n^{-\frac{d_i}{2}}\sum_{t=1}^{n}g_i(t)(1-L)\widetilde{\Phi}(L)\epsilon_t$$

$$=n^{-\frac{d_i}{2}}\sum_{t=1}^{n}g_i(t)\Phi(1)\epsilon_t+n^{-\frac{d_i}{2}}\sum_{t=1}^{n}g_i(t)(\widetilde{\epsilon}_{t-1}-\widetilde{\epsilon}_t), \tag{B.61}$$

where $\widetilde{\epsilon}_t=\widetilde{\Phi}(L)\epsilon_t$ as defined in the beginning of the Appendix. The second term is negligible provided the following Lemma holds.

Lemma B.2.1. Under Assumptions 3.2.1 to 3.2.4, as $n\to\infty$,

$$n^{-\frac{d_i}{2}}\sum_{t=1}^{n}g_i(t)(\widetilde{\epsilon}_{t-1}-\widetilde{\epsilon}_t)=o_P(1).$$

For $P_2(n)$, let $f_{i,0}(L)=\sum_{s=0}^{\infty}\phi_s\psi_{s,i}L^s$. By BN decomposition, we have

$$\begin{aligned}P_2(n)=&n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{s=0}^{\infty}\phi_s\psi_{s,i}\left(\epsilon_{t-s}\eta_{i,t-s}-\theta_i\right)\\
=&n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{s=0}^{\infty}\phi_s\psi_{s,i}L^s\left(\epsilon_t\eta_{it}-\theta_i\right)=n^{-\frac{d_i}{2}}\sum_{t=1}^{n}f_{i,0}(L)\left(\epsilon_t\eta_{it}-\theta_i\right)\\
=&n^{-\frac{d_i}{2}}\sum_{t=1}^{n}f_{i,0}(1)\left(\epsilon_t\eta_{it}-\theta_i\right)-n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\widetilde{f}_{i,0}(L)(1-L)\left(\epsilon_t\eta_{it}-\theta_i\right)\\
=&n^{-\frac{d_i}{2}}f_{i,0}(1)\sum_{t=1}^{n}\left(\epsilon_t\eta_{it}-\theta_i\right)+n^{-\frac{d_i}{2}}\widetilde{\epsilon\eta}_{i,n}^{f}-n^{-\frac{d_i}{2}}\widetilde{\epsilon\eta}_{i,0}^{f},\end{aligned} \tag{B.62}$$

where we denote $\widetilde{\epsilon\eta}_{i,t}^{f}=\widetilde{f}_{i,0}(L)(\epsilon_t\eta_{it}-\theta_i)$. The last two terms in (B.62) are negligible according to the following lemma.

Lemma B.2.2. Under Assumptions 3.2.1 to 3.2.4, as $n\to\infty$, for $t=0$ and $t=n$,

$$n^{-\frac{d_i}{2}}\widetilde{\epsilon\eta}_{it}^{f}\longrightarrow_P 0. \tag{B.63}$$

For $P_3(n)$, let $f_{i,q}(L)=\sum_{s=0}^{\infty}\phi_s\psi_{s+q,i}L^s$. By the BN decompositon,

$$\begin{aligned}P_3(n)=&n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{s=0}^{\infty}\sum_{l=s+1}^{\infty}\phi_s\psi_{l,i}\epsilon_{t-s}\eta_{i,t-l}=n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{s=0}^{\infty}\sum_{q=1}^{\infty}\phi_s\psi_{s+q,i}\epsilon_{t-s}\eta_{i,t-s-q}\\
=&n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{q=1}^{\infty}\left(\sum_{s=0}^{\infty}\phi_s\psi_{s+q,i}L^s\right)\epsilon_t\eta_{i,t-q}=n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{q=1}^{\infty}f_{i,q}(L)\epsilon_t\eta_{i,t-q}\\
=&n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{q=1}^{\infty}f_{i,q}(1)\epsilon_t\eta_{i,t-q}-n^{-\frac{d_i}{2}}\sum_{t=1}^{n}\sum_{q=1}^{\infty}\widetilde{f}_{i,q}(L)(1-L)\epsilon_t\eta_{i,t-q}\end{aligned}$$

$$
\begin{aligned}
&= n^{-\frac{d_i}{2}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} f_{i,q}(1)\epsilon_t \eta_{i,t-q} - n^{-\frac{d_i}{2}} \sum_{t=1}^{n} (1-L) B^f_{i,t} \\
&= n^{-\frac{d_i}{2}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} f_{i,q}(1)\epsilon_t \eta_{i,t-q} + n^{-\frac{d_i}{2}} B^f_{i,0} - n^{-\frac{d_i}{2}} B^f_{i,n},
\end{aligned} \tag{B.64}
$$

where $B^f_{i,t} = \sum_{q=1}^{\infty} \widetilde{f}_{i,q}(L)\epsilon_t \eta_{i,t-q}$. The last two terms are negligible provided the following lemma holds.

Lemma B.2.3. Under Assumptions 3.2.1 to 3.2.4, as $n \to \infty$, for $t = 0$ and $t = n$,

$$
n^{-\frac{d_i}{2}} B^f_{i,t} \longrightarrow_P 0. \tag{B.65}
$$

Similarly, for $P_4(n)$, let $m_{i,q}(L) = \sum_{l=0}^{\infty} \phi_{q+l} \psi_{l,i} L^l$. By the BN decomposition,

$$
\begin{aligned}
P_4(n) &= n^{-\frac{d_i}{2}} \sum_{t=1}^{n} \sum_{l=0}^{\infty} \sum_{s=l+1}^{\infty} \phi_s \psi_{l,i} \epsilon_{t-s} \eta_{i,t-l} = n^{-\frac{d_i}{2}} \sum_{t=1}^{n} \sum_{l=0}^{\infty} \sum_{q=1}^{\infty} \phi_{q+l} \psi_{l,i} \epsilon_{t-q-l} \eta_{i,t-l} \\
&= n^{-\frac{d_i}{2}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} \left(\sum_{l=0}^{\infty} \phi_{q+l} \psi_{l,i} L^l \right) \epsilon_{t-q} \eta_{i,t} = n^{-\frac{d_i}{2}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} m_{i,q}(L) \epsilon_{t-q} \eta_{i,t} \\
&= n^{-\frac{d_i}{2}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} m_{i,q}(1) \epsilon_{t-q} \eta_{i,t} - n^{-\frac{d_i}{2}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} \widetilde{m}_{i,q}(L)(1-L) \epsilon_{t-q} \eta_{i,t} \\
&= n^{-\frac{d_i}{2}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} m_{i,q}(1) \epsilon_{t-q} \eta_{i,t} - n^{-\frac{d_i}{2}} \sum_{t=1}^{n} (1-L) B^m_{i,t} \\
&= n^{-\frac{d_i}{2}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} m_{i,q}(1) \epsilon_{t-q} \eta_{i,t} + n^{-\frac{d_i}{2}} B^m_{i,0} - n^{-\frac{d_i}{2}} B^m_{i,n},
\end{aligned} \tag{B.66}
$$

where $B^m_{i,t} = \sum_{q=1}^{\infty} \widetilde{m}_{i,q}(L)\epsilon_{t-q} \eta_{i,t}$. According to the following lemma, we can ignore the last two terms in (B.66).

Lemma B.2.4. Under Assumptions 3.2.1 to 3.2.4, as $n \to \infty$, for $t = 0$ and $t = n$,

$$
n^{-\frac{d_i}{2}} B^m_{i,t} \longrightarrow_P 0. \tag{B.67}
$$

To summarize, we can ignore the negligible terms in $P_i(n)$ for $i = 1, 2, 3, 4$. Thus, equation (B.60) can be written as

$$
\boldsymbol{Z}_{2,i} = \sum_{t=1}^{n} M^i_{nt} + o_P(1), \tag{B.68}
$$

where

$$M_{nt}^{i} = n^{-\frac{d_i}{2}} \left(g_i(t)\Phi(1)\epsilon_t + f_{i,0}(1)\left(\epsilon_t\eta_{it} - \theta_i\right) + \sum_{q=1}^{\infty} f_{i,q}(1)\epsilon_t\eta_{i,t-q} + \sum_{q=1}^{\infty} m_{i,q}(1)\epsilon_{t-q}\eta_{i,t} \right). \tag{B.69}$$

We denote $M_{nt} = (M_{nt}^1, \ldots, M_{nt}^K)'$, then

$$\boldsymbol{Z}_2 = \sum_{t=1}^{n} M_{nt} + o_P(1). \tag{B.70}$$

M_{nt} is a martingale difference sequence(m.d.s.) suggested by the equation asfollows.

$$\mathrm{E}\left[M_{nt} | F_{t-1}\right] = 0, \tag{B.71}$$

where F_{t-1} is the filtration that $F_{t-1} = \{\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \eta_{t-1}, \eta_{t-2}, \ldots\}$. We are able to apply the Central Limit Theorem for martingale difference sequence(m.d.s) given the following Lemma holds.

Lemma B.2.5. For any $K \times 1$ vector $a = (a_1, \ldots, a_K)'$ with $\|a\| = 1$,

$$\sum_{t=1}^{n} \mathrm{E}\left[\left(a' M_{nt}\right)^2 \middle| F_{t-1}\right] \longrightarrow_P a'\Omega a, \tag{B.72}$$

and

$$\sum_{t=1}^{n} \mathrm{E}\left[\left(a' M_{nt}\right)^4 \middle| F_{t-1}\right] \longrightarrow_P 0, \tag{B.73}$$

where Ω is a $K \times K$ positive definite matrix $\Omega_{ij} = \sigma_1^2 \Phi(1)^2 Q_{ij}$.

Remark B.2.1. Once the two conditions are satisfied, we can apply the CLT for $a'M_{nt}$, which is any linear combination of the elements in M_{nt}. Thus the vector $\sum_{t=1}^{n} M_{nt}$ converges in distribution to multivariate Gaussian.

$$\sum_{t=1}^{n} M_{nt} \longrightarrow_D N(0, \Omega). \tag{B.74}$$

Therefore, by equation (B.70), we have

$$\boldsymbol{Z}_2 = D^{-1}(\boldsymbol{X}'\boldsymbol{e} - n\boldsymbol{b}) \longrightarrow_D N(0, \Omega), \tag{B.75}$$

with $\Omega_{ij} = \sigma_1^2 \Phi(1)^2 Q_{ij}$. As $n \to \infty$, since $\widehat{Q} = D^{-1}\boldsymbol{X}'\boldsymbol{X}D^{-1} \longrightarrow_P Q$, we have

$$D\left(\widehat{\beta}_{ols} - \beta\right) - \widehat{Q}^{-1}D^{-1}n\boldsymbol{b} = (D^{-1}\boldsymbol{X}'\boldsymbol{X}D^{-1})^{-1}(D^{-1}(\boldsymbol{X}'e - n\boldsymbol{b}))$$
$$\longrightarrow_D N(0, Q^{-1}\Omega Q^{-1}), \tag{B.76}$$

i.e.,

$$D\left(\widehat{\beta}_{ols} - \beta - D^{-1}\widehat{Q}^{-1}D^{-1}n\boldsymbol{b}\right) \longrightarrow_D N(0, Q^{-1}\Omega Q^{-1}), \tag{B.77}$$

Thus we complete the proof for this Theorem. ■

Appendix C

Proofs of the Lemmas in Appendix B

C.1 Proofs of the Lemmas in Appendix B.1

Proof of Lemma B.1.1:

Equation (B.1) in Lemma B.1.1 Recall that $\tau_t = t/n$ and

$$M_1(i,j) = \frac{1}{n}\sum_{t=1}^{n}\Big(g_i(\tau_t) - \overline{g}_{i,n}\Big)\Big(g_j(\tau_t) - \overline{g}_{j,n}\Big), \tag{C.1}$$

where $\overline{g}_{i,n} = n^{-1}\sum_{t=1}^{n} g_i(\tau_t)$. Since $g(\cdot)$ is a continuous differentiable function of $\tau \in [0,1]$, under Assumption 2.3.1, we have

$$\overline{g}_{i,n} = \frac{1}{n}\sum_{t=1}^{n} g_i\left(\frac{t}{n}\right) \longrightarrow \int_0^1 g_i(\tau)d\tau = \overline{g}_i, \tag{C.2}$$

as the Riemann sum converges to its integral limit. The same argument applies to $M_1(i,j)$, that

$$\frac{1}{n}\sum_{t=1}^{n}\Big(g_i(\tau_t) - \overline{g}_{i,n}\Big)\Big(g_j(\tau_t) - \overline{g}_{j,n}\Big) \longrightarrow \int_0^1 (g_i(\tau) - \overline{g}_i)(g_j(\tau) - \overline{g}_j)d\tau = Q(i,j). \tag{C.3}$$

Therefore, as $n \to \infty$, $M_1(i,j) \longrightarrow Q(i,j)$, for $i,j = 1,\ldots,k$. ■

Equation (B.2) in Lemma B.1.1 Applying the Cauchy-Schwarz inequality, we have

$$|M_2(i,j)| = \left|\frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n} - \sum_{s=1}^{n} w_{ns}(t) g_i(\tau_s)\right)\left(\overline{g}_{j,n} - \sum_{s=1}^{n} w_{ns}(t) g_j(\tau_s)\right)\right|$$

$$\leq \left|\frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n}-\sum_{s=1}^{n}w_{ns}(t)g_i(\tau_s)\right)^2\right|^{1/2}\left|\frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{j,n}-\sum_{s=1}^{n}w_{ns}(t)g_j(\tau_s)\right)^2\right|^{1/2}. \tag{C.4}$$

Hence we are able to show that $M_2(i,j)=o_p(1)$ if as $n\to\infty$,

$$\frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n}-\sum_{s=1}^{n}w_{ns}(t)g_i(\tau_s)\right)^2 = o_p(1), \tag{C.5}$$

for any $i=1,2,...,k$. Since (C.5) is always positive, we only need to show that

$$\mathrm{E}\left[\frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n}-\sum_{s=1}^{n}w_{ns}(t)g_i(\tau_s)\right)^2\right]=o(1). \tag{C.6}$$

To prove the above result, we first write the equation as

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left(\overline{g}_{i,n}-\sum_{s=1}^{n}w_{ns}(t)g_i(\tau_s)\right)^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{s=1}^{n}w_{ns}(t)\Big(g_i(\tau_s)-\overline{g}_{i,n}\Big)\right)^2\\
=&\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}w_{ns}(t)^2\Big(g_i(\tau_s)-\overline{g}_{i,n}\Big)^2\\
&+\frac{1}{n}\sum_{t=1}^{n}\sum_{s_1=1}^{n}\sum_{\substack{s_2=1,\\ s_2\neq s_1}}^{n}w_{ns_1}(t)w_{ns_2}(t)\Big(g_i(\tau_{s_1})-\overline{g}_{i,n}\Big)\Big(g_i(\tau_{s_2})-\overline{g}_{i,n}\Big).
\end{aligned} \tag{C.7}$$

For simplicity and without loss of generality, we assume $k=1$. Since $w_{ns}(t)=\frac{K\left(\frac{v_s-v_t}{h}\right)}{nh\widehat{f}(v_t)}$, and for the kernel density estimator, we have $\widehat{f}(v)=f(v)+o_p(1)$, which is commonly applied in the literature. Let $\xi_{s,t}=\frac{K\left(\frac{v_s-v_t}{h}\right)}{hf(v_t)}$. Therefore, it is equivalent to show that

$$\mathrm{E}[M_{2,i}(n)]=o(1), \tag{C.8}$$

for $i=1,2$, where

$$M_{2,1}(n)=\frac{1}{n^3}\sum_{t=1}^{n}\sum_{s=1}^{n}\xi_{s,t}^2\Big(g(\tau_s)-\overline{g}_n\Big)^2, \tag{C.9}$$

and

$$M_{2,2}(n)=\frac{1}{n^3}\sum_{t=1}^{n}\sum_{s_1=1}^{n}\sum_{\substack{s_2=1,\\ s_2\neq s_1}}^{n}\xi_{s_1,t}\xi_{s_2,t}\Big(g(\tau_{s_1})-\overline{g}_n\Big)\Big(g(\tau_{s_2})-\overline{g}_n\Big). \tag{C.10}$$

Note that

$$\begin{aligned}
\mathrm{E}[\xi_{s,t}^2] =&\mathrm{E}\left[\frac{K^2\left(\frac{v_s-v_t}{h}\right)}{h^2 f(v_t)^2}\right] = \iint \frac{K^2\left(\frac{v_s-v_t}{h}\right)}{h^2 f(v_t)^2} f_{s,t}(v_s,v_t)dv_s dv_t \\
=&\frac{1}{h^2}\iint \frac{K^2(w) f_{s,t}(z+wh,z)}{f(z)^2} hdwdz = \frac{1+o(h)}{h}\int K^2(w)dw \int \frac{f_{s,t}(z,z)}{f(z)^2}dz. \quad (C.11)
\end{aligned}$$

Therefore, $\mathrm{E}[M_{2,1}(n)] = O((nh)^{-1}) = o(1)$ given that $\int wK^2(w)dw = 0$, $\int K^2(w)dw < \infty$ and

$$\max_{s,t}\int \frac{f_{s,t}(z,z)}{f(z)^2}dz < \infty. \quad (C.12)$$

Meanwhile, Let Θ_3 denote the condition that $t \neq s_1 \neq s_2$, then

$$\begin{aligned}
&\mathrm{E}\left[\frac{1}{n^3}\sum_{\Theta_3}\xi_{s_1,t}\xi_{s_2,t}\left(g(\tau_{s_1})-\overline{g}_n\right)\left(g(\tau_{s_2})-\overline{g}_n\right)\right] \\
=&\frac{1}{n^3}\sum_{\Theta_3}\mathrm{E}\left[\xi_{s_1,t}\xi_{s_2,t}\right]\left(g(\tau_{s_1})-\overline{g}_n\right)\left(g(\tau_{s_2})-\overline{g}_n\right). \quad (C.13)
\end{aligned}$$

Note that by change of variables, we have

$$\begin{aligned}
&\mathrm{E}\left[\xi_{s_1,t}\xi_{s_2,t}\right] = \iiint \frac{K\left(\frac{v_{s_1}-v_t}{h}\right)K\left(\frac{v_{s_2}-v_t}{h}\right)}{h^2 f(v_t^2)} f(v_{s_1},v_{s_2},v_t)dv_{s_1}dv_{s_2}dv_t \\
=&\iiint \frac{K(w_1)K(w_2)}{h^2 f(z)^2} f_{s_1,s_2,t}(z+w_1h, z+w_2h, z)h^2 dw_1 dw_2 dz \\
=&(1+o(1))\left(\int K(w_1)dw_1\right)^2 \int \frac{f_{s_1,s_2,t}(z,z,z)}{f(z)^2}dz \\
=&(1+o(1))\int \frac{f_{s_1,s_2,t}(z,z,z)-f(z)^3+f(z)^3}{f(z)^2}dz \\
=&(1+o(1))\Big(D_3(s_1,s_2,t)+1\Big), \quad (C.14)
\end{aligned}$$

where

$$D_3(s_1,s_2,t) = \int \frac{f_{s_1,s_2,t}(z,z,z)-f(z)^3}{f(z)^2}dz. \quad (C.15)$$

Then,

$$\begin{aligned}
&\frac{1}{n^3}\sum_{\Theta_3}\mathrm{E}\left[\xi_{s_1,t}\xi_{s_2,t}\right]\left(g(\tau_{s_1})-\overline{g}_n\right)\left(g(\tau_{s_2})-\overline{g}_n\right) \\
=&\frac{(1+o(1))}{n^3}\sum_{\Theta_3}\Big(D_3(s_1,s_2,t)+1\Big)\left(g(\tau_{s_1})-\overline{g}_n\right)\left(g(\tau_{s_2})-\overline{g}_n\right)
\end{aligned}$$

$$=\frac{(1+o(1))}{n^3}\sum_{\Theta_3} D_3(s_1,s_2,t)\Big(g(\tau_{s_1})-\overline{g}_n\Big)\Big(g(\tau_{s_2})-\overline{g}_n\Big)$$
$$+\frac{(1+o(1))}{n^2}\sum_{\Theta_2}\Big(g(\tau_{s_1})-\overline{g}_n\Big)\Big(g(\tau_{s_2})-\overline{g}_n\Big) \triangleq (1+o(1))(M_{2,2,1}+M_{2,2,2}). \qquad \text{(C.16)}$$

Since $g(\cdot)$ is a bounded function, there exists some positive value $C_g > 0$, such that $\max_{\tau\in[0,1]}|g(\tau)-\overline{g}| < C_g$ uniformly. Then we have

$$\begin{aligned}\left|M_{2,2,1}\right| &\le \frac{1}{n^3}\sum_{\Theta_3}\left|D_3(s_1,s_2,t)\Big(g(\tau_{s_1})-\overline{g}_n\Big)\Big(g(\tau_{s_2})-\overline{g}_n\Big)\right| \\ &\le \frac{C_g}{n^3}\sum_{\Theta_3}\left|\iint \frac{f_{s_1,s_2,t}(z,z,z)-f(z)^3}{f(z)^2}dz\right| = O(n^{-1}).\end{aligned} \qquad \text{(C.17)}$$

where the last step holds by Assumption 2.3.4. For $M_{2,2,2}$, it is straightforward that

$$\frac{1}{n^2}\sum_{s\neq t}\Big(g(\tau_{s_1})-\overline{g}_n\Big)\Big(g(\tau_{s_2})-\overline{g}_n\Big) = O(n^{-1}). \qquad \text{(C.18)}$$

Therefore, $M_{2,2,2} = O(n^{-1})$. To summarize, we have proved that $\mathrm{E}\Big[M_{2,i}(n)\Big] = o(1)$ for $i = 1,2$. They directly imply that $M_2(i,j) = o_p(1)$. Hence we complete the proof. ■

Equations (B.3) and (B.4) in Lemma B.1.1 Using the Cauchy-Schwarz inequality,

$$|M_{12}(i,j)| \le |M_1(i,i)|^{1/2}|M_2(j,j)|^{1/2} = O_p(1)o_p(1) = o_p(1). \qquad \text{(C.19)}$$

The same result holds for $M_{21}(i,j)$ for $i,j = 1,\ldots,k$. ■

Equation (B.5) in Lemma B.1.1 For $S_2(i,j)$, using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}|S_2(i,j)| &= \left|\frac{1}{n}\sum_{t=1}^{n}\left(v_{it}-\sum_{s=1}^{n}w_{ns}(t)v_{is}\right)\left(v_{jt}-\sum_{s=1}^{n}w_{ns}(t)v_{js}\right)\right| \\ &\le \left|\frac{1}{n}\sum_{t=1}^{n}\left(v_{it}-\sum_{s=1}^{n}w_{ns}(t)v_{is}\right)^2\right|^{1/2}\left|\frac{1}{n}\sum_{t=1}^{n}\left(v_{jt}-\sum_{s=1}^{n}w_{ns}(t)v_{js}\right)^2\right|^{1/2},\end{aligned} \qquad \text{(C.20)}$$

for $i,j = 1,\ldots,k$, where

$$\frac{1}{n}\sum_{t=1}^{n}\left(v_{it}-\sum_{s=1}^{n}w_{ns}(t)v_{is}\right)^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{s=1}^{n}w_{ns}(t)(v_{is}-v_{it})\right)^2$$

$$=\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}w_{ns}(t)^2(v_{is}-v_{it})^2+\frac{1}{n}\sum_{t=1}^{n}\sum_{s_1=1}^{n}\sum_{\substack{s_2=1,\\ s_2\neq s_1}}^{n}w_{ns_1}(t)w_{ns_2}(t)(v_{is_1}-v_{it})(v_{is_2}-v_{it}). \quad (C.21)$$

As in the previous proofs, we assume $k=1$. As $\widehat{f}(v)=f(v)+o_p(1)$, it suffices to prove that (C.21) is $o_p(1)$. Since (C.21) is always positive, we only need to show that the expectations of the following two expressions are $o(1)$.

$$S_{2,1}=\frac{1}{n^3}\sum_{t=1}^{n}\sum_{\substack{s=1\\ s\neq t}}^{n}\xi_{s,t}^2(v_s-v_t)^2, \quad (C.22)$$

$$S_{2,2}=\frac{1}{n^3}\sum_{t=1}^{n}\sum_{\substack{s_1=1\\ s_1\neq t}}^{n}\sum_{\substack{s_2=1,\\ s_2\neq t, s_2\neq s_1}}^{n}\xi_{s_1,t}\xi_{s_2,t}(v_{s_1}-v_t)(v_{s_2}-v_t), \quad (C.23)$$

where $\xi_{s,t}=\frac{K\left(\frac{v_s-v_t}{h}\right)}{hf(v_t)}$. Here, we only examine equation (C.23) and (C.22) is similar and simpler. Define $L(u)=uK(u)$, we have

$$\begin{aligned}
|\mathrm{E}[S_{2,2}]| =& \left|\mathrm{E}\left[\frac{1}{n^3}\sum_{\Theta_3}\frac{L\left(\frac{v_{s_1}-v_t}{h}\right)L\left(\frac{v_{s_2}-v_t}{h}\right)}{f(v_t)^2}\right]\right| \\
=& \left|\frac{1}{n^3}\sum_{\Theta_3}\iiint\frac{L\left(\frac{v_{s_1}-v_t}{h}\right)L\left(\frac{v_{s_2}-v_t}{h}\right)}{f(v_t)^2}f(v_{s_1},v_{s_2},v_t,)dv_{s_1}dv_{s_2}dv_t\right| \\
\leq& \frac{1}{n^3}\sum_{\Theta_3}\iiint\left|\frac{L(w_1)L(w_2)}{f(z)^2}f_{s_1,s_2,t}(z+w_1h,z+w_2h,z)\right|h^2dw_1dw_2dz \\
=& \frac{(1+o(1))h^2}{n^3}\sum_{\Theta_3}\iiint\left|\frac{L(w_1)L(w_2)}{f(z)^2}f_{s_1,s_2,t}(z,z,z)\right|dw_1dw_2dz \\
=& \frac{(1+o(1))h^2}{n^3}\sum_{\Theta_3}\left(\int|L(w_1)|dw_1\right)^2\int\frac{f_{s_1,s_2,t}(z,z,z)}{f(z)^2}dz=O(h^2),
\end{aligned} \quad (C.24)$$

where $\int|L(w_1)|dw_1<\infty$ and

$$\max_{s_1,s_2,t}\int\frac{f_{s_1,s_2,t}(z,z,z)}{f(z)^2}dz<\infty. \quad (C.25)$$

Under Assumption (2.3.6), $h\to 0$ as $n\to\infty$, therefore, $\mathrm{E}[S_{2,2}]\to 0$ as $n\to\infty$. This implies that $S_{2,2}=o_p(1)$. Similarly, $S_{2,1}=o_p(1)$ and therefore $S_2(i,j)=o_p(1)$ for any

$i, j = 1, 2, ..., k$. Thus we complete the proof. $\blacksquare$

Proof of Lemma B.1.2: Note that

$$\begin{aligned} I_{1n} =& \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \zeta(v_t) \widetilde{v}_t^2 = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \zeta(v_t) \left(\frac{\frac{1}{nh} \sum_{s=1}^{n} K\left(\frac{v_s - v_t}{h}\right)(v_s - v_t)}{\widehat{f}(v_t)} \right)^2 \\ =& \frac{(1 + o_p(1))}{\sqrt{n}} \sum_{t=1}^{n} \zeta(v_t) \left(\frac{\frac{1}{nh} \sum_{s=1}^{n} K\left(\frac{v_s - v_t}{h}\right)(v_s - v_t)}{f(v_t)} \right)^2 \triangleq (1 + o_p(1)) \widetilde{I}_{1n}. \end{aligned} \tag{C.26}$$

Hence, it is equivalent to show $\widetilde{I}_{1n} = o_p(1)$. We consider the square of $\widetilde{I}_{1n}$ as follows.

$$\begin{aligned} \widetilde{I}_{1n}^2 =& \frac{1}{n^5 h^4} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \zeta(v_{t_1}) \zeta(v_{t_2}) \left(\frac{\sum_{s=1}^{n} K\left(\frac{v_s - v_{t_1}}{h}\right)\left(v_s - v_{t_1}\right)}{f(v_{t_1})} \right)^2 \\ & \times \left(\frac{\sum_{l=1}^{n} K\left(\frac{v_l - v_{t_2}}{h}\right)\left(v_l - v_{t_2}\right)}{f(v_{t_2})} \right)^2 \\ =& \frac{1}{n^5} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \frac{\zeta(v_{t_1}) \zeta(v_{t_2})}{f(v_{t_1})^2 f(v_{t_2})^2} \left(\sum_{s=1}^{n} L\left(\frac{v_s - v_{t_1}}{h}\right) \right)^2 \left(\sum_{l=1}^{n} L\left(\frac{v_l - v_{t_2}}{h}\right) \right)^2 \\ =& \frac{1}{n^5} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \sum_{s_1=1}^{n} \sum_{s_2=1}^{n} \sum_{l_1=1}^{n} \sum_{l_2=1}^{n} \frac{\zeta(v_{t_1}) \zeta(v_{t_2})}{f(v_{t_1})^2 f(v_{t_2})^2} L\left(\frac{v_{s_1} - v_{t_1}}{h}\right) L\left(\frac{v_{s_2} - v_{t_1}}{h}\right) L\left(\frac{v_{l_1} - v_{t_2}}{h}\right) \\ & \times L\left(\frac{v_{l_2} - v_{t_2}}{h}\right). \end{aligned} \tag{C.27}$$

Therefore, we have to prove that the expectation of

$$\begin{aligned} & \frac{1}{n^5} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \sum_{s_1=1}^{n} \sum_{s_2=1}^{n} \sum_{l_1=1}^{n} \sum_{l_2=1}^{n} \frac{\zeta(v_{t_1}) \zeta(v_{t_2})}{f(v_{t_1})^2 f(v_{t_2})^2} L\left(\frac{v_{s_1} - v_{t_1}}{h}\right) L\left(\frac{v_{s_2} - v_{t_1}}{h}\right) \\ & \times L\left(\frac{v_{l_1} - v_{t_2}}{h}\right) L\left(\frac{v_{l_2} - v_{t_2}}{h}\right) \end{aligned} \tag{C.28}$$

is $o(1)$. We consider some typical cases of the index vector $(t_1, t_2, s_1, s_2, l_1, l_2)$, while under remaining conditions it can be proved similarly.

(1) Let Θ_6 denote $t_1 \neq t_2 \neq s_1 \neq s_2 \neq l_1 \neq l_2$,

$$\begin{aligned} & \frac{1}{n^5} \sum_{\Theta_6} \mathrm{E}\left[\frac{\zeta(v_{t_1}) \zeta(v_{t_2})}{f(v_{t_1})^2 f(v_{t_2})^2} L\left(\frac{v_{s_1} - v_{t_1}}{h}\right) L\left(\frac{v_{s_2} - v_{t_1}}{h}\right) L\left(\frac{v_{l_1} - v_{t_2}}{h}\right) L\left(\frac{v_{l_2} - v_{t_2}}{h}\right) \right] \\ =& \frac{1}{n^5} \sum_{\Theta_6} \int \cdots \int \frac{\zeta(v_{t_1}) \zeta(v_{t_2})}{f(v_{t_1})^2 f(v_{t_2})^2} L\left(\frac{v_{s_1} - v_{t_1}}{h}\right) L\left(\frac{v_{s_2} - v_{t_1}}{h}\right) L\left(\frac{v_{l_1} - v_{t_2}}{h}\right) \\ & L\left(\frac{v_{l_2} - v_{t_2}}{h}\right) f_{t_1, t_2, s_1, s_2, l_1, l_2}(v_{t_1}, v_{t_2}, v_{s_1}, v_{s_2}, v_{l_1}, v_{l_2}) dv_{t_1} dv_{t_2} dv_{s_1} dv_{s_2} dv_{l_1} dv_{l_2} \end{aligned}$$

$$=\frac{h^4}{n^5}\sum_{\Theta_6}\int\cdots\int\frac{\zeta(z_1)\zeta(z_2)}{f(z_1)^2f(z_2)^2}L(w_1)L(w_2)L(w_3)L(w_4)$$

$$f_{t_1,t_2,s_1,s_2,l_1,l_2}(z_1,z_2,z_1+w_1h,z_1+w_2h,z_2+w_3h,z_2+w_4h)dw_1dw_2dw_3dw_4dz_1dz_2$$

$$=\frac{(1+o(1))h^4}{n^5}\sum_{\Theta_6}\iiint\iiint\frac{\zeta(z_1)\zeta(z_2)}{f(z_1)^2f(z_2)^2}L(w_1)L(w_2)L(w_3)L(w_4)$$

$$f^{(4)(3,4,5,6)}_{t_1,t_2,s_1,s_2,l_1,l_2}(z_1,z_2,z_1,z_1,z_2,z_2)w_1w_2w_3w_4h^4dw_1dw_2dw_3dw_4dz_1dz_2$$

$$=\frac{(1+o(1))h^8}{n^5}\sum_{\Theta_6}\int w_1L(w_1)dw_1\int w_2L(w_2)dw_2\int w_3L(w_3)dw_3\int w_4L(w_4)dw_4$$

$$\iint\frac{\zeta(z_1)\zeta(z_2)f^{(4)(3,4,5,6)}_{t_1,t_2,s_1,s_2,l_1,l_2}(z_1,z_2,z_1,z_1,z_2,z_2)}{f(z_1)^2f(z_2)^2}dz_1dz_2$$

$$=\frac{(1+o(1))nh^8\kappa_{21}^4}{n^6}\sum_{\Theta_6}\iint\frac{\zeta(z_1)\zeta(z_2)f^{(4)(3,4,5,6)}_{t_1,t_2,s_1,s_2,l_1,l_2}(z_1,z_2,z_1,z_1,z_2,z_2)}{f(z_1)^2f(z_2)^2}dz_1dz_2 \tag{C.29}$$

where $\kappa_{21}=\int wL(w)dw=\int w^2K(w)dw<\infty$. Meanwhile, we require that

$$\max_{t_1,t_2,s_2,s_2,l_1,l_2}\left|\iint\frac{\zeta(z_1)\zeta(z_2)f^{(4)(3,4,5,6)}_{t_1,t_2,s_1,s_2,l_1,l_2}(z_1,z_2,z_1,z_1,z_2,z_2)}{f(z_1)^2f(z_2)^2}dz_1dz_2\right|<\infty. \tag{C.30}$$

Therefore, when $t_1\neq t_2\neq s_1\neq s_2\neq l_1\neq l_2$, as $n\to\infty$, $nh^8\to 0$, and

$$\frac{1}{n^5}\sum_{\Theta_6}\mathrm{E}\left[\frac{\zeta(v_{t_1})\zeta(v_{t_2})}{f(v_{t_1})^2f(v_{t_2})^2}L\left(\frac{v_{s_1}-v_{t_1}}{h}\right)L\left(\frac{v_{s_2}-v_{t_1}}{h}\right)L\left(\frac{v_{l_1}-v_{t_2}}{h}\right)L\left(\frac{v_{l_2}-v_{t_2}}{h}\right)\right]$$

$$=O(nh^8)=o(1). \tag{C.31}$$

(2) Let Θ_5 denote $t_1=t_2=t, t\neq s_1\neq s_2\neq l_1\neq l_2$. Therefore,

$$\frac{1}{n^5}\sum_{\Theta_5}\mathrm{E}\left[\frac{\zeta(v_t)^2}{f(v_t)^4}L\left(\frac{v_{s_1}-v_t}{h}\right)L\left(\frac{v_{s_2}-v_t}{h}\right)L\left(\frac{v_{l_1}-v_t}{h}\right)L\left(\frac{v_{l_2}-v_t}{h}\right)\right]$$

$$=\frac{1}{n^5}\sum_{\Theta_5}\int\cdots\int\frac{\zeta(v_t)^2}{f(v_t)^4}L\left(\frac{v_{s_1}-v_t}{h}\right)L\left(\frac{v_{s_2}-v_t}{h}\right)L\left(\frac{v_{l_1}-v_t}{h}\right)L\left(\frac{v_{l_2}-v_t}{h}\right)$$

$$f(v_t,v_{s_1},v_{s_2},v_{l_1},v_{l_2})dv_tdv_{s_1}dv_{s_2}dv_{l_1}dv_{l_2}$$

$$=\frac{1}{n^5}\sum_{\Theta_5}\int\cdots\int\frac{\zeta(z)^2}{f(z)^4}L(w_1)L(w_2)L(w_3)L(w_4)$$

$$f_{t,s_1,s_2,l_1,l_2}(z,z+w_1h,z+w_2h,z+w_3h,z+w_4h)h^4dzdw_1dw_2dw_3dw_4$$

$$\leq \frac{1}{n^5}\sum_{\Theta_5}\int\cdots\int\left|\frac{\zeta(z)^2}{f(z)^4}L(w_1)L(w_2)L(w_3)L(w_4)\right|$$

$$f_{t,s_1,s_2,l_1,l_2}(z,z+w_1h,z+w_2h,z+w_3h,z+w_4h)h^4dzdw_1dw_2dw_3dw_4$$

$$=\frac{(1+o(1))h^4}{n^5}\sum_{\Theta_5}\int\cdots\int\left|\frac{\zeta(z)^2}{f(z)^4}L(w_1)L(w_2)L(w_3)L(w_4)\right|$$

$$f_{t,s_1,s_2,l_1,l_2}(z,z,z,z,z)dzdw_1dw_2dw_3dw_4$$

$$=\frac{(1+o(1))h^4}{n^5}\sum_{\Theta_5}\left(\int|L(w)|dw\right)^4\int\left|\frac{\zeta(z)^2}{f(z)^4}\right|f_{t,s_1,s_2,l_1,l_2}(z,z,z,z,z)dz=O(h^4), \tag{C.32}$$

where $\int|L(w)|dw<\infty$, and

$$\max_{t,s_1,s_2,l_1,l_2}\int\frac{\zeta(z)^2}{f(z)^4}f_{t,s_1,s_2,l_1,l_2}(z,z,z,z,z)dz<\infty, \tag{C.33}$$

(3) Let Θ_3 denote $t_1=t_2=t, s_1=s_2=s, l_1=l_2=l, t\neq s\neq l$.

$$\frac{1}{n^5}\sum_{\Theta_3}\mathrm{E}\left[\frac{\zeta(v_t)^2}{f(v_t)^4}L^2\left(\frac{v_s-v_t}{h}\right)L^2\left(\frac{v_l-v_t}{h}\right)\right]$$

$$=\frac{1}{n^5}\sum_{\Theta_3}\iiint\frac{\zeta(v_t)^2}{f(v_t)^4}L^2\left(\frac{v_s-v_t}{h}\right)L^2\left(\frac{v_l-v_t}{h}\right)f_{t,s,l}(v_t,v_s,v_l)dv_tdv_sdv_l$$

$$=\frac{h^2}{n^5}\sum_{\Theta_3}\iiint\frac{\zeta(z)^2}{f(z)^4}L^2(w_1)L^2(w_2)f_{t,s,l}(z,z+w_1h,z+w_2h)dw_1dw_2dz$$

$$=\frac{(1+o(1))h^2}{n^5}\sum_{\Theta_3}\iiint\frac{\zeta(z)^2}{f(z)^4}L^2(w_1)L^2(w_2)f_{t,s,l}(z,z,z)dw_1dw_2dz$$

$$=\frac{(1+o(1))h^2}{n^5}\sum_{\Theta_3}\int L^2(w_1)dw_1\int L^2(w_2)dw_2\int\frac{\zeta(z)^2f_{t,s,l}(z,z,z)}{f(z)^4}dz$$

$$=\frac{(1+o(1))h^2\kappa_{22}^2}{n^2}\frac{1}{n^3}\sum_{\Theta_3}\int\frac{\zeta(z)^2f_{t,s,l}(z,z,z)}{f(z)^4}dz, \tag{C.34}$$

where $\kappa_{22}=\int L^2(w)dw=\int w^2K^2(w)dw<\infty$, and

$$\max_{t,s,l}\int\frac{\zeta(z)^2}{f(z)^4}f_{t,s,l}(z,z,z)dz<\infty. \tag{C.35}$$

Therefore, in this condition when $t_1=t_2=t, s_1=s_2=s, l_1=l_2=l, t\neq s\neq l$,

$$\frac{1}{n^5}\sum_{t\neq s\neq l}\mathrm{E}\left[\frac{\zeta(v_t)^2}{f(v_t)^4}L^2\left(\frac{v_s-v_t}{h}\right)L^2\left(\frac{v_l-v_t}{h}\right)\right]=O\left(\frac{h^2}{n^2}\right)=o(1). \tag{C.36}$$

To summarize,

$$\mathrm{E}[\widetilde{I}_{1n}^2] = o(1), \tag{C.37}$$

which implies that $\widetilde{I}_{1n} = o_p(1)$ and therefore $I_{1n} = o_p(1)$.

We consider the second moment of I_{2n}, which can be written as

$$\begin{aligned}
I_{2n} =& \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \widetilde{g}(\tau_t)\zeta(v_t)\widetilde{v}_t \\
=& \frac{1}{n^2\sqrt{n}h^2} \sum_{t=1}^{n} \zeta(v_t) \left(\frac{\sum_{s=1}^{n} K\left(\frac{v_s - v_t}{h}\right)\left(g(\tau_s) - g(\tau_t)\right)}{\widehat{f}(v_t)} \right) \left(\frac{\sum_{s=1}^{n} K\left(\frac{v_s - v_t}{h}\right)(v_s - v_t)}{\widehat{f}(v_t)} \right) \\
=& \frac{1 + o_p(1)}{n^2\sqrt{n}h^2} \sum_{t=1}^{n} \zeta(v_t) \left(\frac{\sum_{s=1}^{n} K\left(\frac{v_s - v_t}{h}\right)\left(g(\tau_s) - g(\tau_t)\right)}{f(v_t)} \right) \left(\frac{\sum_{s=1}^{n} K\left(\frac{v_s - v_t}{h}\right)(v_s - v_t)}{f(v_t)} \right) \\
=& (1 + o_p(1))\widetilde{I}_{2n}.
\end{aligned} \tag{C.38}$$

Hence, it is equivalent to prove that $\widetilde{I}_{2n}$ is $o_p(1)$. We then show that $\mathrm{E}[\widetilde{I}_{2n}^2] \to 0$ as $n \to \infty$.

$$\begin{aligned}
\widetilde{I}_{2n}^2 = & \frac{1}{n^5h^4} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \zeta(v_{t_1})\zeta(v_{t_2}) \left(\frac{\sum_{s=1}^{n} K\left(\frac{v_s - v_{t_1}}{h}\right)\left(g(\tau_s) - g(\tau_{t_1})\right)}{f(v_{t_1})} \right) \\
& \left(\frac{\sum_{s=1}^{n} K\left(\frac{v_s - v_{t_1}}{h}\right)\left(v_s - v_{t_1}\right)}{f(v_{t_1})} \right) \left(\frac{\sum_{l=1}^{n} K\left(\frac{v_l - v_{t_2}}{h}\right)\left(g(\tau_l) - g(\tau_{t_2})\right)}{f(v_{t_2})} \right) \\
& \left(\frac{\sum_{l=1}^{n} K\left(\frac{v_l - v_{t_2}}{h}\right)\left(v_l - v_{t_2}\right)}{f(v_{t_2})} \right) \\
= & \frac{1}{n^5h^2} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \frac{\zeta(v_{t_1})\zeta(v_{t_2})}{f(v_{t_1})^2 f(v_{t_2})^2} \left(\sum_{s_1=1}^{n} K\left(\frac{v_{s_1} - v_{t_1}}{h}\right)\left(g(\tau_{s_1}) - g(\tau_{t_1})\right) \right) \\
& \left(\sum_{s_2=1}^{n} L\left(\frac{v_{s_2} - v_{t_1}}{h}\right) \right) \left(\sum_{l_1=1}^{n} K\left(\frac{v_{l_1} - v_{t_2}}{h}\right)\left(g(\tau_s) - g(\tau_{t_2})\right) \right) \left(\sum_{l_2=1}^{n} L\left(\frac{v_{l_2} - v_{t_2}}{h}\right) \right) \\
= & \frac{1}{n^5h^2} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \sum_{s_1=1}^{n} \sum_{s_2=1}^{n} \sum_{l_1=1}^{n} \sum_{l_2=1}^{n} \frac{\zeta(v_{t_1})\zeta(v_{t_2})}{f(v_{t_1})^2 f(v_{t_2})^2} \left(g(\tau_{s_1}) - g(\tau_{t_1})\right)\left(g(\tau_{l_1}) - g(\tau_{t_2})\right) \\
& \times K\left(\frac{v_{s_1} - v_{t_1}}{h}\right) L\left(\frac{v_{s_2} - v_{t_1}}{h}\right) K\left(\frac{v_{l_1} - v_{t_2}}{h}\right) L\left(\frac{v_{l_2} - v_{t_2}}{h}\right).
\end{aligned} \tag{C.39}$$

Therefore,[1] we need to show that the expectation of

$$\frac{1}{n^5h^2}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\frac{\zeta(v_{t_1})\zeta(v_{t_2})}{f(v_{t_1})^2f(v_{t_2})^2}\Big(g(\tau_{s_1})-g(\tau_{t_1})\Big)\Big(g(\tau_{l_1})-g(\tau_{t_2})\Big)$$
$$K\Big(\frac{v_{s_1}-v_{t_1}}{h}\Big)L\Big(\frac{v_{s_2}-v_{t_1}}{h}\Big)K\Big(\frac{v_{l_1}-v_{t_2}}{h}\Big)L\Big(\frac{v_{l_2}-v_{t_2}}{h}\Big) \tag{C.40}$$

is $o(1)$. We only consider some typical conditions of the index vector $(t_1,t_2,s_1,s_2,l_1,l_2)$, and the proofs can be conducted similarly for other conditions.

(1) Let $\mathcal{G}(s_1,t_1,l_1,t_2)=\Big(g(\tau_{s_1})-g(\tau_{t_1})\Big)\Big(g(\tau_{l_1})-g(\tau_{t_2})\Big)$, when all the subscripts are not equal to each other,

$$\begin{aligned}
&\mathrm{E}\Bigg[\frac{1}{n^5h^2}\sum_{\Theta_6}\frac{\zeta(v_{t_1})\zeta(v_{t_2})}{f(v_{t_1})^2f(v_{t_2})^2}\mathcal{G}(s_1,t_1,l_1,t_2)K\Big(\frac{v_{s_1}-v_{t_1}}{h}\Big)L\Big(\frac{v_{s_2}-v_{t_1}}{h}\Big)K\Big(\frac{v_{l_1}-v_{t_2}}{h}\Big)\\
&L\Big(\frac{v_{l_2}-v_{t_2}}{h}\Big)\Bigg]\\
=&\frac{1}{n^5h^2}\sum_{\Theta_6}\mathcal{G}(s_1,t_1,l_1,t_2)\mathrm{E}\Bigg[\frac{\zeta(v_{t_1})\zeta(v_{t_2})}{f(v_{t_1})^2f(v_{t_2})^2}K\Big(\frac{v_{s_1}-v_{t_1}}{h}\Big)L\Big(\frac{v_{s_2}-v_{t_1}}{h}\Big)K\Big(\frac{v_{l_1}-v_{t_2}}{h}\Big)\\
&L\Big(\frac{v_{l_2}-v_{t_2}}{h}\Big)\Bigg]\\
=&\frac{1}{n^5h^2}\sum_{\Theta_6}\mathcal{G}(s_1,t_1,l_1,t_2)\int\dots\int\frac{\zeta(v_{t_1})\zeta(v_{t_2})}{f(v_{t_1})^2f(v_{t_2})^2}K\Big(\frac{v_{s_1}-v_{t_1}}{h}\Big)L\Big(\frac{v_{s_2}-v_{t_1}}{h}\Big)K\Big(\frac{v_{l_1}-v_{t_2}}{h}\Big)\\
&\qquad L\Big(\frac{v_{l_2}-v_{t_2}}{h}\Big)f_{t_1,t_2,s_1,s_2,l_1,l_2}(v_{t_1},v_{t_2},v_{s_1},v_{s_2},v_{l_1},v_{l_2})dv_{t_1}dv_{t_2}dv_{s_1}dv_{s_2}dv_{l_1}dv_{l_2}\\
=&\frac{1}{n^5h^2}\sum_{\Theta_6}\mathcal{G}(s_1,t_1,l_1,t_2)\int\dots\int\frac{\zeta(z_1)\zeta(z_2)}{f(z_1)^2f(z_2)^2}K(w_1)L(w_2)K(w_3)L(w_4)\\
&\qquad f_{t_1,t_2,s_1,s_2,l_1,l_2}(z_1,z_2,z_1+w_1h,z_1+w_2h,z_2+w_3h,z_2+w_4h)h^4dw_1dw_2dw_3dw_4dz_1dz_2\\
=&\frac{h^2}{n^5}\sum_{\Theta_6}\mathcal{G}(s_1,t_1,l_1,t_2)\int\dots\int\frac{\zeta(z_1)\zeta(z_2)}{f(z_1)^2f(z_2)^2}K(w_1)L(w_2)K(w_3)L(w_4)\Big(f_{t_1,t_2,s_1,s_2,l_1,l_2}(z_1,z_2,\\
&\qquad z_1+w_1h,z_1+w_2h,z_2+w_3h,z_2+w_4h)-f(z_1)f(z_2)f(z_1+w_1h)f(z_1+w_2h)\\
&\qquad f(z_2+w_3h)f(z_2+w_4h)+f(z_1)f(z_2)f(z_1+w_1h)f(z_1+w_2h)f(z_2+w_3h)f(z_2+w_4h)\Big)\\
&dw_1dw_2dw_3dw_4dz_1dz_2\\
=&D_3+G_3,
\end{aligned} \tag{C.41}$$

where the second term is

[1]The term in the summation becomes 0 when one of the following conditions happens, $s_1=t_1$, $s_2=t_1$, $s_1=t_2$, or $s_2=t_2$.

$$
\begin{aligned}
G_3 =& \frac{h^2}{n^5}\sum_{\Theta_6}\mathcal{G}(s_1,t_1,l_1,t_2)\int\dots\int\frac{\zeta(z_1)\zeta(z_2)}{f(z_1)^2f(z_2)^2}K(w_1)L(w_2)K(w_3)L(w_4)f(z_1)f(z_2)\\
&\qquad f(z_1+w_1h)f(z_1+w_2h)f(z_2+w_3h)f(z_2+w_4h)dw_1dw_2dw_3dw_4dz_1dz_2\\
=& \frac{h^2}{n^5}\sum_{\Theta_6}\mathcal{G}(s_1,t_1,l_1,t_2)\int\dots\int\frac{\zeta(z_1)\zeta(z_2)}{f(z_1)^2f(z_2)^2}K(w_1)L(w_2)K(w_3)L(w_4)f(z_1)f(z_2)f(z_1)\\
&\qquad f'(z_1)w_2hf(z_2)f'(z_2)w_4hdw_1dw_2dw_3dw_4dz_1dz_2\\
=& \frac{h^4}{n^5}\sum_{\Theta_6}\mathcal{G}(s_1,t_1,l_1,t_2)\left(\int\zeta(z_1)f'(z_1)dz_1\right)^2\left(\int w_2L(w_2)dw_2\right)^2\\
=& \frac{Ch^4}{n^3}\sum_{s_1\neq t_1\neq l_1\neq t_2}\mathcal{G}(s_1,t_1,l_1,t_2)=O(h^4),
\end{aligned}
\tag{C.42}
$$

in which we require $\int wL(w)dw<\infty$, $\int\zeta(z_1)f'(z_1)dz_1<\infty$ and it is straightforward that

$$
\frac{1}{n^3}\sum_{s_1\neq t_1\neq l_1\neq t_2}\mathcal{G}(s_1,t_1,l_1,t_2)=O(1). \tag{C.43}
$$

While D_3 is also $o(1)$ because

$$
\begin{aligned}
|D_3| =& \frac{(1+o(1))h^2}{n^5}\sum_{\Theta_6}|\mathcal{G}(s_1,t_1,l_1,t_2)|\int\dots\int\left|\frac{\zeta(z_1)\zeta(z_2)}{f(z_1)^2f(z_2)^2}\right|\left|K(w_1)L(w_2)K(w_3)L(w_4)\right|\\
&\qquad \left|f_{t_1,t_2,s_1,s_2,l_1,l_2}(z_1,z_2,z_1,z_1,z_2,z_2)-f(z_1)^3f(z_2)^3\right|dw_1dw_2dw_3dw_4dz_1dz_2\\
=& \frac{(1+o(1))h^2}{n^5}\sum_{\Theta_6}|\mathcal{G}(s_1,t_1,l_1,t_2)|\left(\int|K(w_1)|dw_1\right)^2\left(\int|L(w_2)|dw_2\right)^2\\
&\qquad \iint\left|\frac{\zeta(z_1)\zeta(z_2)}{f(z_1)^2f(z_2)^2}\right|\left|f_{t_1,t_2,s_1,s_2,l_1,l_2}(z_1,z_2,z_1,z_1,z_2,z_2)-f(z_1)^3f(z_2)^3\right|dz_1dz_2\\
=& O(h^2),
\end{aligned}
\tag{C.44}
$$

where we used the assumptions that $\int|K(w)|dw<\infty$, $\int|L(w)|dw<\infty$, $\max_{s_1,t_1,l_1,t_2}|\mathcal{G}(s_1,t_1,l_1,t_2)|<\infty$ and

$$
\sum_{\Theta_6}\iint\frac{\zeta(z_1)\zeta(z_2)}{f(z_1)^2f(z_2)^2}\left|f_{t_1,t_2,s_1,s_2,l_1,l_2}(z_1,z_2,z_1,z_1,z_2,z_2)-f(z_1)^3f(z_2)^3\right|dz_1dz_2=O(n^5). \tag{C.45}
$$

Hence,

$$
\frac{1}{n^5h^2}\sum_{\Theta_6}\mathcal{G}(s_1,t_1,l_1,t_2)\mathrm{E}\left[\frac{\zeta(v_{t_1})\zeta(v_{t_2})}{f(v_{t_1})^2f(v_{t_2})^2}K\left(\frac{v_{s_1}-v_{t_1}}{h}\right)L\left(\frac{v_{s_2}-v_{t_1}}{h}\right)K\left(\frac{v_{l_1}-v_{t_2}}{h}\right)\right.
$$

$$L\left(\frac{v_{l_2}-v_{t_2}}{h}\right)\Bigg] = o(1). \tag{C.46}$$

(2) Let $\mathcal{G}(s,t,l) = \mathcal{G}(s,t,l,t)$, and let Θ_3 denote $t \neq s \neq l$, and assume $t_1 = t_2 = t, s_1 = s_2 = s$, $l_1 = l_2 = l$,.

$$\begin{aligned}
&\frac{1}{n^5h^2}\sum_{\Theta_3}\mathcal{G}(s,t,l)\mathrm{E}\left[\frac{\zeta(v_t)^2}{f(v_t)^4}K\left(\frac{v_s-v_t}{h}\right)L\left(\frac{v_s-v_t}{h}\right)K\left(\frac{v_l-v_t}{h}\right)L\left(\frac{v_l-v_t}{h}\right)\right]\\
=&\frac{1}{n^5h^2}\sum_{\Theta_3}\mathcal{G}(s,t,l)\iiint\frac{\zeta(v_t)^2}{f(v_t)^4}K\left(\frac{v_s-v_t}{h}\right)L\left(\frac{v_s-v_t}{h}\right)K\left(\frac{v_l-v_t}{h}\right)L\left(\frac{v_l-v_t}{h}\right)\\
&\qquad\qquad \times f(v_t,v_s,v_l)dv_tdv_sdv_l\\
=&\frac{1}{n^5h^2}\sum_{\Theta_3}\mathcal{G}(s,t,l)\iiint\frac{\zeta(z)^2}{f(z)^4}K(w_1)L(w_1)K(w_2)L(w_2)f_{t,s,l}(z,z+w_1h,z+w_2h)\\
&\qquad\qquad \times h^2dw_1dw_2dz\\
=&\frac{1+o(1)}{n^5}\sum_{\Theta_3}\mathcal{G}(s,t,l)\iiint\frac{\zeta(z)^2}{f(z)^4}w_1K^2(w_1)w_2K^2(w_2)f_{t,s,l}(z,z,z)dw_1dw_2dz\\
\leq&\frac{1+o(1)}{n^5}\sum_{\Theta_3}|\mathcal{G}(s,t,l)|\iiint\left|\frac{\zeta(z)^2}{f(z)^4}w_1K^2(w_1)w_2K^2(w_2)f_{t,s,l}(z,z,z)\right|dw_1dw_1dz\\
=&\frac{1+o(1)}{n^5}\sum_{\Theta_3}|\mathcal{G}(s,t,l)|\left(\int|w_1|K^2(w_1)dw_1\right)^2\int\left|\frac{\zeta(z)^2}{f(z)^4}f_{t,s,l}(z,z,z)\right|dz\\
=&O\left(\frac{1}{n^2}\right),
\end{aligned} \tag{C.47}$$

where for any s,t,l, $|\mathcal{G}(s,t,l)| < \infty$, $\int|w|K^2(w)dw < \infty$, and

$$\max_{t,s,l}\int\frac{\zeta(z)^2}{f(z)^4}f_{t,s,l}(z,z,z)dz < \infty. \tag{C.48}$$

Therefore, when $t_1 = t_2 = t, s_1 = s_2 = s, l_1 = l_2 = l, s \neq t \neq l$,

$$\begin{aligned}
&\frac{1}{n^5h^2}\sum_{\Theta_2}\mathcal{G}(s,t,l,t)\mathrm{E}\left[\frac{\zeta(v_t)^2}{f(v_t)^4}K\left(\frac{v_s-v_t}{h}\right)L\left(\frac{v_s-v_t}{h}\right)K\left(\frac{v_l-v_t}{h}\right)L\left(\frac{v_l-v_t}{h}\right)\right]\\
&\quad =o(1).
\end{aligned} \tag{C.49}$$

To summarize,

$$\mathrm{E}[\widetilde{I}_{2n}^2] = o(1), \tag{C.50}$$

which implies that $I_{2n} = o_p(1)$. Thus we complete the proof. ■

Proof of Lemma B.1.4: Without loss of generality[2], we assume $k = 1$. Therefore, the objective is to prove that

$$\mathfrak{D}_1(n) = \frac{1}{n}\sum_{t=1}^{n}(\widehat{x_t} - \widetilde{x_t})^2 = o_p(1). \tag{C.51}$$

Substitute $\widehat{x_t}$ and $\widetilde{x_t}$ in $\mathfrak{D}_1(n)$,

$$\begin{aligned}
\mathfrak{D}_1(n) &= \frac{1}{n}\sum_{t=1}^{n}(\widehat{x_t} - \widetilde{x_t})^2 \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(x_t - \sum_{s=1}^{n}\overline{w}_{ns}(t)x_s - x_t + \sum_{s=1}^{n}w_{ns}(t)x_s\right)^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\sum_{s=1}^{n}\Big(w_{ns}(t) - \overline{w}_{ns}(t)\Big)x_s\right)^2 \\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\Big(w_{ns}(t) - \overline{w}_{ns}(t)\Big)^2 x_s^2 + \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{\substack{r=1\\ r\neq s}}^{n}\Big(w_{ns}(t) - \overline{w}_{ns}(t)\Big)\Big(w_{nr}(t) - \overline{w}_{nr}(t)\Big)x_s x_r \\
&= \mathfrak{D}_{11}(n) + \mathfrak{D}_{12}(n).
\end{aligned} \tag{C.52}$$

Let $\Pi_{s,t} = w_{ns}(t) - \overline{w}_{ns}(t)$. We define several useful notations as follows.

$$\begin{aligned}
\Xi_1(r,t) =& \frac{K\left(\frac{v_r - v_t}{h}\right)}{f(v_t)}(g(\tau_t) - g(\tau_r)), \quad \Xi_2(r,t) = \frac{K\left(\frac{v_r - v_t}{h}\right)}{f(v_t)}(v_t - v_r), \\
\Xi_3(l,t) =& K_2\left(\frac{\tau_l - \tau_t}{b}\right)(g(\tau_l) - g(\tau_t)), \Xi_4(l,t) = K_2\left(\frac{\tau_l - \tau_t}{b}\right)v_t, \\
\Xi_5(p,t,l) =& \Xi_3(l,p) + \Xi_4(l,p) - \Xi_3(l,t) - \Xi_4(l,t).
\end{aligned}$$

Note that

$$\begin{aligned}
\Pi_{s,t} =& \frac{K\left(\frac{v_s - v_t}{h}\right)}{nh\widehat{f}(v_t)} - \frac{K\left(\frac{\widehat{v_s} - \widehat{v_t}}{h}\right)}{nh\widehat{f}(\widehat{v_t})} = \frac{K\left(\frac{v_s - v_t}{h}\right)(\widehat{f}(v_t) - \widehat{f}(\widehat{v_t}))}{nh\widehat{f}(v_t)\widehat{f}(\widehat{v_t})} \\
&+ \frac{\left(K\left(\frac{v_t - v_s}{h}\right) - K\left(\frac{\widehat{v_t} - \widehat{v_s}}{h}\right)\right)}{nh\widehat{f}(\widehat{v_t})} \\
=& \frac{K\left(\frac{v_s - v_t}{h}\right)(\widehat{f}(v_t) - \widehat{f}(\widehat{v_t}))}{nh(f(v_t)^2 + o_p(1))} + \frac{\left(K\left(\frac{v_t - v_s}{h}\right) - K\left(\frac{\widehat{v_t} - \widehat{v_s}}{h}\right)\right)}{nh(f(v_t) + o_p(1))} \\
=& (1 + o_p(1))\left(\frac{K\left(\frac{v_s - v_t}{h}\right)(\widehat{f}(v_t) - \widehat{f}(\widehat{v_t}))}{nhf(v_t)^2} + \frac{\left(K\left(\frac{v_t - v_s}{h}\right) - K\left(\frac{\widehat{v_t} - \widehat{v_s}}{h}\right)\right)}{nhf(v_t)}\right).
\end{aligned} \tag{C.53}$$

[2] The proofs for the multivariate case can be carried out at the cost of more tedious derivations.

By Taylor expansion, and denote $K_d(p,t) = K'\left(\frac{v_p - v_t}{h}\right)$,

$$
\begin{aligned}
K\left(\frac{v_p - v_t}{h}\right) - K\left(\frac{\widehat{v}_p - \widehat{v}_t}{h}\right) &= (1 + o_p(1))K'\left(\frac{v_p - v_t}{h}\right)\left(\left(\frac{v_p - v_t}{h}\right) - \left(\frac{\widehat{v}_p - \widehat{v}_t}{h}\right)\right) \\
=& (1 + o_p(1))K_d(p,t)\left(\frac{v_p - \widehat{v}_p}{h}\right) - (1 + o_p(1))K_d(p,t)\left(\frac{v_t - \widehat{v}_t}{h}\right) \\
=& \frac{(1 + o_p(1))}{h}K_d(p,t)\left(\widehat{g}(\tau_p) - g(\tau_p)\right) - \frac{(1 + o_p(1))}{h}K_d(p,t)\left(\widehat{g}(\tau_t) - g(\tau_t)\right) \\
=& \frac{(1 + o_p(1))}{h}K_d(p,t)\left(\sum_{l=1}^{n} w^*_{nl}(p)x_p - g(\tau_p)\right) \\
&- \frac{(1 + o_p(1))}{h}K_d(p,t)\left(\sum_{l=1}^{n} w^*_{nl}(t)x_t - g(\tau_t)\right) \\
=& \frac{(1 + o_p(1))^2}{nhb}K_d(p,t)\left(\sum_{l=1}^{n} K_2\left(\frac{\tau_l - \tau_p}{b}\right)(g(\tau_l) - g(\tau_p)) + \sum_{l=1}^{n} K_2\left(\frac{\tau_l - \tau_p}{b}\right)v_l\right) \\
&- \frac{(1 + o_p(1))^2}{nhb}K_d(p,t)\left(\sum_{l=1}^{n} K_2\left(\frac{\tau_l - \tau_t}{b}\right)(g(\tau_l) - g(\tau_t)) + \sum_{l=1}^{n} K_2\left(\frac{\tau_l - \tau_t}{b}\right)v_t\right) \\
\triangleq& \frac{(1 + o_p(1))^2}{nhb}K_d(p,t)\sum_{l=1}^{n}\left(\Xi_3(l,p) + \Xi_4(l,p) - \Xi_3(l,t) - \Xi_4(l,t)\right) \\
\triangleq& \frac{(1 + o_p(1))^2}{nhb}K_d(p,t)\sum_{l=1}^{n}\Xi_5(p,t,l),
\end{aligned}
\tag{C.54}
$$

where $\Xi_5(p,t,l) = \Xi_3(l,p) + \Xi_4(l,p) - \Xi_3(l,t) - \Xi_4(l,t)$. Meanwhile,

$$
\begin{aligned}
\widehat{f}(v_t) - \widehat{f}(\widehat{v}_t) =& \frac{1}{nh}\sum_{p=1}^{n}\left(K\left(\frac{v_p - v_t}{h}\right) - K\left(\frac{\widehat{v}_p - \widehat{v}_t}{h}\right)\right) \\
=& \frac{(1 + o_p(1))}{n^2h^2b}\sum_{p=1}^{n}K_d(p,t)\sum_{l=1}^{n}\Xi_5(p,t,l).
\end{aligned}
\tag{C.55}
$$

Therefore,

$$
\begin{aligned}
\Pi_{s,t} =& (1 + o_p(1))\left(\frac{K\left(\frac{v_s - v_t}{h}\right)}{n^3h^3bf(v_t)^2}\sum_{p=1}^{n}K_d(p,t)\sum_{l=1}^{n}\Xi_5(p,t,l)\right) \\
&+ (1 + o_p(1))\left(\frac{K_d(s,t)\sum_{l=1}^{n}\Xi_5(s,t,l)}{n^2h^2bf(v_t)}\right) \triangleq (1 + o_p(1))(\Pi_{s,t,1} + \Pi_{s,t,2}).
\end{aligned}
\tag{C.56}
$$

We then have

$$
\mathfrak{D}_{11}(n) = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\left(w_{ns}(t) - \overline{w}_{ns}(t)\right)^2 x_s^2 = \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\left((1 + o_p(1))(\Pi_{s,t,1} + \Pi_{s,t,2})\right)^2 x_s^2
$$

$$=\frac{(1+o_p(1))}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}(\Pi_{s,t,1}+\Pi_{s,t,2})^2x_s^2 \triangleq (1+o_p(1))\widetilde{\mathfrak{D}}_{11}. \tag{C.57}$$

Also note that

$$\begin{aligned}\widetilde{\mathfrak{D}}_{11} =&\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}(\Pi_{s,t,1}+\Pi_{s,t,2})^2x_s^2\\ \leq&\frac{2}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\Pi_{s,t,1}^2x_s^2+\frac{2}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\Pi_{s,t,2}^2x_s^2 \triangleq 2\widetilde{\mathfrak{D}}_{11,1}+2\widetilde{\mathfrak{D}}_{11,2}\end{aligned} \tag{C.58}$$

Hence we need to show that $\widetilde{\mathfrak{D}}_{11,1}=o_p(1)$ and $\widetilde{\mathfrak{D}}_{11,2}=o_p(1)$. As the all the terms in $\widetilde{\mathfrak{D}}_{11,i}$ are positive, we only need to show that $\mathrm{E}[\widetilde{\mathfrak{D}}_{11,i}]\to 0$ as $n\to\infty$ for $i=1,2$. We omit the proofs here as they belong to special and simpler cases of our subsequent proofs for $\mathfrak{D}_{12}(n)$. Then we move on to prove the term $\mathfrak{D}_{12}(n)$.

$$\begin{aligned}\mathfrak{D}_{12}(n) =&\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{\substack{r=1\\ r\neq s}}^{n}\Big(w_{ns}(t)-\overline{w}_{ns}(t)\Big)\Big(w_{nr}(t)-\overline{w}_{nr}(t)\Big)x_sx_r\\ =&\frac{(1+o_p(1))}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{\substack{r=1\\ r\neq s}}^{n}\Big(\Pi_{s,t,1}+\Pi_{s,t,2}\Big)\Big(\Pi_{r,t,1}+\Pi_{r,t,2}\Big)x_sx_r\\ =&(1+o_p(1))\left(\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{\substack{r=1\\ r\neq s}}^{n}\Pi_{s,t,1}\Pi_{r,t,1}x_sx_r+\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{\substack{r=1\\ r\neq s}}^{n}\Pi_{s,t,1}\Pi_{r,t,2}x_sx_r\right.\\ &\left.+\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{\substack{r=1\\ r\neq s}}^{n}\Pi_{s,t,2}\Pi_{r,t,1}x_sx_r+\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{\substack{r=1\\ r\neq s}}^{n}\Pi_{s,t,2}\Pi_{r,t,2}x_sx_r\right)\\ \triangleq&(1+o_p(1))\Big(\widetilde{\mathfrak{D}}_{12,1}(n)+\widetilde{\mathfrak{D}}_{12,2}(n)+\widetilde{\mathfrak{D}}_{12,3}(n)+\widetilde{\mathfrak{D}}_{12,4}(n)\Big).\end{aligned} \tag{C.59}$$

Therefore, we need to prove that $\widetilde{\mathfrak{D}}_{12,i}(n)=o_p(1)$ for $i=1,2,3,4$. We take $\widetilde{\mathfrak{D}}_{12,1}(n)$ as an example, and other terms can be proved similarly.

$$\begin{aligned}\widetilde{\mathfrak{D}}_{12,1}(n) =&\frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{\substack{r=1\\ r\neq s}}^{n}\Pi_{s,t,1}\Pi_{r,t,1}x_sx_r\\ =&\frac{1}{n^7h^6b^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{\substack{r=1\\ r\neq s}}^{n}\frac{K\left(\frac{v_s-v_t}{h}\right)K\left(\frac{v_r-v_t}{h}\right)}{f(v_t)^4}\end{aligned}$$

$$\times \sum_{p_1=1}^{n} K_d(p_1,t) \sum_{l_1=1}^{n} \Xi_5(p_1,t,l_1) \sum_{p_2=1}^{n} K_d(p_2,t) \sum_{l_2=1}^{n} \Xi_5(p_2,t,l_2) x_s x_r$$

$$= \frac{1}{n^7 h^6 b^2} \sum_{t=1}^{n} \sum_{s=1}^{n} \sum_{\substack{r=1 \\ r\neq s}}^{n} \frac{K\left(\frac{v_s - v_t}{h}\right) K\left(\frac{v_r - v_t}{h}\right)}{f(v_t)^4}$$

$$\times \sum_{p_1=1}^{n} K_d(p_1,t) \sum_{l_1=1}^{n} \Big(\Xi_3(l_1,p_1) + \Xi_4(l_1,p_1) - \Xi_3(l_1,t) - \Xi_4(l_1,t)\Big)$$

$$\times \sum_{p_2=1}^{n} K_d(p_2,t) \sum_{l_2=1}^{n} \Big(\Xi_3(l_2,p_2) + \Xi_4(l_2,p_2) - \Xi_3(l_2,t) - \Xi_4(l_2,t)\Big) x_s x_r. \quad \text{(C.60)}$$

We pick one typical term to demonstrate the proving methods, and other terms can be conducted similarly. Note that $x_t = g(\tau_t) + v_t$,

$$\frac{1}{n^7 h^6 b^2} \sum_{t=1}^{n} \sum_{s=1}^{n} \sum_{\substack{r=1 \\ r\neq s}}^{n} \sum_{l_1=1}^{n} \sum_{p_1=1}^{n} \sum_{l_2=1}^{n} \sum_{p_2=1}^{n} \frac{K\left(\frac{v_s - v_t}{h}\right) K\left(\frac{v_r - v_t}{h}\right)}{f(v_t)^4} K_d(p_1,t) \Xi_3(l_1,p_1) K_d(p_2,t)$$

$$\Xi_3(l_2,p_2) g(\tau_s) g(\tau_r) \sim \frac{1}{n^7 h^6 b^2} \sum_{\Theta_7} \frac{K\left(\frac{v_s - v_t}{h}\right) K\left(\frac{v_r - v_t}{h}\right)}{f(v_t)^4} K_d(p_1,t) K_2\left(\frac{\tau_{l_1} - \tau_{p_1}}{b}\right)$$

$$(g(\tau_{l_1}) - g(\tau_{p_1})) K_d(p_2,t) K_2\left(\frac{\tau_{l_2} - \tau_{p_2}}{b}\right) (g(\tau_{l_2}) - g(\tau_{p_2})) g(\tau_s) g(\tau_r), \quad \text{(C.61)}$$

where $\sim$ means that they have the same order. Also note that

$$\mathrm{E}\left[\left|\frac{K\left(\frac{v_s - v_t}{h}\right) K\left(\frac{v_r - v_t}{h}\right)}{h^4 f(v_t)^4} K_d(p_1,t) K_d(p_2,t)\right|\right] \leq C, \quad \text{(C.62)}$$

for some $C > 0$ and

$$\frac{1}{n^2 b^2} \sum_{l=1}^{n} \sum_{p=1}^{n} \left| K_2\left(\frac{\tau_l - \tau_p}{b}\right) (g(\tau_l) - g(\tau_p)) \right| = (1 + o(1)) C. \quad \text{(C.63)}$$

Therefore,

$$\mathrm{E}\left[\left|\frac{1}{n^7 h^6 b^2} \sum_{\Theta_7} \frac{K\left(\frac{v_s - v_t}{h}\right) K\left(\frac{v_r - v_t}{h}\right)}{f(v_t)^4} K_d(p_1,t) K_2\left(\frac{\tau_{l_1} - \tau_{p_1}}{b}\right) (g(\tau_{l_1}) - g(\tau_{p_1})) K_d(p_2,t)\right.\right.$$

$$\left.\left. K_2\left(\frac{\tau_{l_2} - \tau_{p_2}}{b}\right) (g(\tau_{l_2}) - g(\tau_{p_2})) g(\tau_s) g(\tau_r)\right|\right]$$

$$\leq \frac{1}{n^7 h^6 b^2} \sum_{\Theta_7} \mathrm{E}\left[\left|\frac{K\left(\frac{v_s - v_t}{h}\right) K\left(\frac{v_r - v_t}{h}\right)}{f(v_t)^4} K_d(p_1,t) K_d(p_2,t)\right|\right] \left| K_2\left(\frac{\tau_{l_1} - \tau_{p_1}}{b}\right) (g(\tau_{l_1}) - g(\tau_{p_1}))\right.$$

$$K_2\left(\frac{\tau_{l_2}-\tau_{p_2}}{b}\right)(g(\tau_{l_2})-g(\tau_{p_2}))g(\tau_s)g(\tau_r)\Bigg|$$
$$\leq\frac{C}{n^7h^2b^2}\sum_{\Theta_7}\left|K_2\left(\frac{\tau_{l_1}-\tau_{p_1}}{b}\right)(g(\tau_{l_1})-g(\tau_{p_1}))K_2\left(\frac{\tau_{l_2}-\tau_{p_2}}{b}\right)(g(\tau_{l_2})-g(\tau_{p_2}))g(\tau_s)g(\tau_r)\right|$$
$$=(1+o(1))Cb^2/h^2=O(b^2/h^2). \tag{C.64}$$

Similarly, we can show that other terms are all $o_p(1)$. Therefore, $\widetilde{\mathfrak{D}}_{12,1}=o_p(1)$, hence

$$\mathfrak{D}_{12,1}=o_p(1). \tag{C.65}$$

It is similar to show that

$$\mathfrak{D}_{12,i}=o_p(1). \tag{C.66}$$

for $i=2,3,4$. Hence, we are able to show that $\mathfrak{D}_{11}=o_p(1),\mathfrak{D}_{12}=o_p(1)$. To summarize,

$$\mathfrak{D}_1=o_p(1). \tag{C.67}$$

Thus,we complete the proof. ■

Proof of Lemma B.1.3: As $\widetilde{x}_t=\widetilde{g}(\tau_t)+\widetilde{v}_t$, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^n\widetilde{x}_t\overline{u}_t=\frac{1}{\sqrt{n}}\sum_{t=1}^n\widetilde{g}(\tau_t)\overline{u}_t+\frac{1}{\sqrt{n}}\sum_{t=1}^n\widetilde{v}_t\overline{u}_t\triangleq G_1(n)+G_2(n), \tag{C.68}$$

hence, we need to show that both $G_1(n)$ and $G_2(n)$ are $o_p(1)$. Note that by previous definition,

$$\begin{aligned}G_1(n)=&\frac{1}{\sqrt{n}}\sum_{t=1}^n\widetilde{g}(\tau_t)\overline{u}_t=\frac{1}{\sqrt{n}}\sum_{t=1}^n\left(g(\tau_t)-\sum_{s=1}^n w_{ns}(t)g(\tau_s)\right)\left(\sum_{l=1}^n w_{nl}(t)u_l\right)\\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^n\left(\frac{1}{nh}\sum_{s=1}^n\frac{K\left(\frac{v_s-v_t}{h}\right)(g(\tau_t)-g(\tau_s))}{\widehat{f}(v_t)}\right)\left(\frac{1}{nh}\sum_{l=1}^n\frac{K\left(\frac{v_l-v_t}{h}\right)u_l}{\widehat{f}(v_t)}\right)\\
=&\frac{1+o_p(1)}{n^2\sqrt{n}h^2}\sum_{t=1}^n\sum_{s=1}^n\sum_{l=1}^n\frac{K\left(\frac{v_s-v_t}{h}\right)(g(\tau_t)-g(\tau_s))K\left(\frac{v_l-v_t}{h}\right)u_l}{f(v_t)^2}\triangleq(1+o_p(1))\widetilde{G}_1(n).\end{aligned} \tag{C.69}$$

Therefore, it suffices to show that $\widetilde{G}_1(n)$ is $o_p(1)$. Note that

$$\widetilde{G}_1(n)^2=\left(\frac{1}{n^2\sqrt{n}h^2}\sum_{t=1}^n\sum_{s=1}^n\sum_{l=1}^n\frac{K\left(\frac{v_s-v_t}{h}\right)(g(\tau_t)-g(\tau_s))K\left(\frac{v_l-v_t}{h}\right)u_l}{f(v_t)^2}\right)^2$$

$$=\frac{1}{n^5h^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\frac{K\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)K\left(\frac{v_{l_1}-v_{t_1}}{h}\right)K\left(\frac{v_{l_2}-v_{t_2}}{h}\right)}{f(v_{t_1})^2f(v_{t_2})^2}$$
$$\times u_{l_1}u_{l_2}(g(\tau_{t_1})-g(\tau_{s_1}))(g(\tau_{t_2})-g(\tau_{s_2})). \tag{C.70}$$

By Assumption 2.3.3, for the α-mixing error term u_t, we have

$$\mathrm{E}\left[\frac{1}{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}u_{l_1}u_{l_2}\right]=\frac{1}{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\mathrm{E}\left[u_{l_1}u_{l_2}\right]=O(1). \tag{C.71}$$

Meanwhile,

$$\begin{aligned}
&\mathrm{E}\left[\frac{K\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)K\left(\frac{v_{l_1}-v_{t_1}}{h}\right)K\left(\frac{v_{l_2}-v_{t_2}}{h}\right)}{h^4f(v_{t_1})^2f(v_{t_2})^2}\right]\\
=&\int\cdots\int\frac{K\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)K\left(\frac{v_{l_1}-v_{t_1}}{h}\right)K\left(\frac{v_{l_2}-v_{t_2}}{h}\right)}{h^4f(v_{t_1})^2f(v_{t_2})^2}f(v_{s_1},v_{s_2},v_{l_1},v_{l_2},v_{t_1},v_{t_2})\\
&\qquad dv_{s_1}dv_{s_2}dv_{l_1}dv_{l_2}dv_{t_1}dv_{t_2}\\
=&\iiiint\frac{K(w_1)K(w_2)K(w_3)K(w_4)}{f(z_1)^2f(z_2)^2}f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1+w_1h,z_2+w_2h,z_1+w_3h\\
&\qquad ,z_2+w_4h,z_1,z_2)dw_1dw_2dw_3dw_4dz_1dz_2\\
=&(1+o(1))\left(\int K(w_1)dw_1\right)^4\iint\frac{f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1,z_2,z_1,z_2,z_1,z_2)}{f(z_1)^2f(z_2)^2}dz_1dz_2\\
=&(1+o(1))\iint\frac{f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1,z_2,z_1,z_2,z_1,z_2)}{f(z_1)^2f(z_2)^2}dz_1dz_2,
\end{aligned} \tag{C.72}$$

where

$$\begin{aligned}
&\iint\frac{f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1,z_2,z_1,z_2,z_1,z_2)}{f(z_1)^2f(z_2)^2}dz_1dz_2\\
=&\iint\frac{f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1,z_2,z_1,z_2,z_1,z_2)-f(z_1)^3f(z_2)^3}{f(z_1)^2f(z_2)^2}dz_1dz_2+\iint f(z_1)f(z_2)dz_1dz_2\\
\triangleq&B_4(s_1,s_2,t_1,t_2)+1.
\end{aligned} \tag{C.73}$$

Therefore,[3]

$$\mathrm{E}\left[\widetilde{G}_1(n)^2\right]=\frac{1+o(1)}{n^5}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\Big(B_4(s_1,s_2,t_1,t_2)+1\Big)\mathrm{E}\left[u_{l_1}u_{l_2}\right]$$

[3]Here, we only need to consider the condition when all the indexes are not equal to each other since the quantity would be a higher order when we have equal indexes.

$$
\begin{aligned}
&\quad\times(g(\tau_{t_1})-g(\tau_{s_1}))(g(\tau_{t_2})-g(\tau_{s_2}))\\
=&\frac{1+o(1)}{n^5}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}B_4(s_1,s_2,t_1,t_2)\mathrm{E}\left[u_{l_1}u_{l_2}\right]\\
&\quad\times(g(\tau_{t_1})-g(\tau_{s_1}))(g(\tau_{t_2})-g(\tau_{s_2}))\\
&+\frac{1+o(1)}{n^5}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\mathrm{E}\left[u_{l_1}u_{l_2}\right](g(\tau_{t_1})-g(\tau_{s_1}))(g(\tau_{t_2})-g(\tau_{s_2}))\\
=&(1+o(1))\left(\widetilde{EG}_{11}(n)+\widetilde{EG}_{12}(n)\right).
\end{aligned}
\tag{C.74}
$$

We show that

$$
\begin{aligned}
&\left|\widetilde{EG}_{11}(n)\right|\\
=&\left|\frac{1}{n^5}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}B_4(s_1,s_2,t_1,t_2)\mathrm{E}\left[u_{l_1}u_{l_2}\right](g(\tau_{t_1})-g(\tau_{s_1}))(g(\tau_{t_2})-g(\tau_{s_2}))\right|\\
\leq&\frac{1}{n^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\left|B_4(s_1,s_2,t_1,t_2)\right|\\
&\frac{1}{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\left|\mathrm{E}\left[u_{l_1}u_{l_2}\right]\right|\left|(g(\tau_{t_1})-g(\tau_{s_1}))(g(\tau_{t_2})-g(\tau_{s_2}))\right|\\
=&o(1)O(1)=o(1),
\end{aligned}
\tag{C.75}
$$

and by Assumption 2.3.4,

$$
\begin{aligned}
&\frac{1}{n^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\left|B_4(s_1,s_2,t_1,t_2)\right|\\
=&\frac{1}{n^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\left|\iint\frac{f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1,z_2,z_1,z_2,z_1,z_2)-f(z_1)^3f(z_2)^3}{f(z_1)^2f(z_2)^2}dz_1dz_2\right|\\
\leq&\frac{1}{n^4c_f^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\\
&\max_{l_1,l_2}\iint\left|f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1,z_2,z_1,z_2,z_1,z_2)-f(z_1)^3f(z_2)^3\right|dz_1dz_2=o(1).
\end{aligned}
\tag{C.76}
$$

Also note that

$$
\widetilde{EG}_{12}(n)=\frac{1}{n^5}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\mathrm{E}\left[u_{l_1}u_{l_2}\right](g(\tau_{t_1})-g(\tau_{s_1}))(g(\tau_{t_2})-g(\tau_{s_2}))
$$

$$
\begin{aligned}
=&\left(\frac{1}{n^4}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}(g(\tau_{t_1})-g(\tau_{s_1}))(g(\tau_{t_2})-g(\tau_{s_2}))\right)\left(\frac{1}{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\mathrm{E}\left[u_{l_1}u_{l_2}\right]\right)\\
=&O\left(\frac{1}{n}\right)O(1)=O\left(\frac{1}{n}\right),
\end{aligned}
\tag{C.77}
$$

where $\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\mathrm{E}\left[u_{l_1}u_{l_2}\right]=O(n)$ because u_t is a statioanry α-mixing process. Therefore, we have $\mathrm{E}[\widetilde{G}_1(n)^2]=o(1)$, which implies that $G_1(n)=o_p(1)$. While for $G_2(n)$,

$$
\begin{aligned}
G_2(n)=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{v}_t\overline{u}_t=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(v_t-\sum_{s=1}^{n}w_{ns}(t)v_s\right)\left(\sum_{l=1}^{n}w_{nl}(t)u_l\right)\\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{1}{nh}\sum_{s=1}^{n}\frac{K\left(\frac{v_s-v_t}{h}\right)(v_t-v_s)}{\widehat{f}(v_t)}\right)\left(\frac{1}{nh}\sum_{l=1}^{n}\frac{K\left(\frac{v_l-v_t}{h}\right)u_l}{\widehat{f}(v_t)}\right)\\
=&\frac{1+o_p(1)}{n^2\sqrt{n}h^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{l=1}^{n}\frac{K\left(\frac{v_s-v_t}{h}\right)(v_t-v_s)K\left(\frac{v_l-v_t}{h}\right)u_l}{f(v_t)^2}\triangleq(1+o_p(1))\widetilde{G}_2(n).
\end{aligned}
\tag{C.78}
$$

Thus it is sufficient to show $\widetilde{G}_2(n)=o_p(1)$. Let $L(u)=uK(u)$, and note that

$$
\begin{aligned}
\widetilde{G}_2(n)^2=&\left(\frac{1}{n^2\sqrt{n}h}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{l=1}^{n}\frac{L\left(\frac{v_s-v_t}{h}\right)K\left(\frac{v_l-v_t}{h}\right)u_l}{f(v_t)^2}\right)^2\\
=&\frac{1}{n^5h^2}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\frac{L\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)L\left(\frac{v_{l_1}-v_{t_1}}{h}\right)K\left(\frac{v_{l_2}-v_{t_2}}{h}\right)u_{l_1}u_{l_2}}{f(v_{t_1})^2f(v_{t_2})^2}.
\end{aligned}
\tag{C.79}
$$

Similar to the previous proof, we have

$$
\begin{aligned}
&\mathrm{E}\left[\widetilde{G}_2(n)^2\right]\\
=&\mathrm{E}\left[\frac{1}{n^5h^2}\sum_{t_1,t_2,s_1,s_2,l_1,l_2=1}^{n}\frac{L\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)L\left(\frac{v_{l_1}-v_{t_1}}{h}\right)K\left(\frac{v_{l_2}-v_{t_2}}{h}\right)u_{l_1}u_{l_2}}{f(v_{t_1})^2f(v_{t_2})^2}\right]\\
=&\frac{1}{n^5h^2}\sum_{t_1,t_2,s_1,s_2,l_1,l_2=1}^{n}\mathrm{E}\left[\frac{L\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)L\left(\frac{v_{l_1}-v_{t_1}}{h}\right)K\left(\frac{v_{l_2}-v_{t_2}}{h}\right)}{f(v_{t_1})^2f(v_{t_2})^2}\right]\mathrm{E}\left[u_{l_1}u_{l_2}\right]\\
=&\frac{O(h^4)}{n^5h^2}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\sum_{l_1=1}^{n}\sum_{l_2=1}^{n}\mathrm{E}\left[u_{l_1}u_{l_2}\right]=O(h^2)=o(1),
\end{aligned}
\tag{C.80}
$$

where we used the results that

$$
\left|\mathrm{E}\left[\frac{L\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)L\left(\frac{v_{l_1}-v_{t_1}}{h}\right)K\left(\frac{v_{l_2}-v_{t_2}}{h}\right)}{f(v_{t_1})^2f(v_{t_2})^2}\right]\right|
$$

$$
\begin{aligned}
=&\left|\int\cdots\int \frac{L\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)L\left(\frac{v_{l_1}-v_{t_1}}{h}\right)K\left(\frac{v_{l_2}-v_{t_2}}{h}\right)}{f(v_{t_1})^2 f(v_{t_2})^2}\right.\\
&\left. f(v_{s_1},v_{s_2},v_{l_1},v_{l_2},v_{t_1},v_{t_2})dv_{s_1}dv_{s_2}dv_{l_1}dv_{l_2}dv_{t_1}dv_{t_2}\right|\\
\leq&\int\cdots\int\left|\frac{L\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)L\left(\frac{v_{l_1}-v_{t_1}}{h}\right)K\left(\frac{v_{l_2}-v_{t_2}}{h}\right)}{f(v_{t_1})^2 f(v_{t_2})^2} f(v_{s_1},v_{s_2},v_{l_1},v_{l_2},v_{t_1},v_{t_2})\right|\\
& dv_{s_1}dv_{s_2}dv_{l_1}dv_{l_2}dv_{t_1}dv_{t_2}\\
=&h^4\int\cdots\int\left|\frac{L(w_1)K(w_2)L(w_3)K(w_4)}{f(z_1)^2 f(z_2)^2}\right.\\
&\left. f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1+w_1h,z_2+w_2h,z_1+w_3h,z_2+w_4h,z_1,z_2)\right| dw_1dw_2dw_3dw_4dz_1dz_2\\
=&(1+o(1))h^4\left(\int|L(w)|dw\right)^2\left(\int|K(w)|dw\right)^2\\
&\times\iint\left|\frac{f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1,z_2,z_1,z_2,z_1,z_2)}{f(z_1)^2 f(z_2)^2}\right|dz_1dz_2\\
\leq&(1+o(1))h^4\left(\int|L(w)|dw\right)^2\left(\int|K(w)|dw\right)^2\\
&\times\iint\frac{f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1,z_2,z_1,z_2,z_1,z_2)}{f(z_1)^2 f(z_2)^2}dz_1dz_2\\
=&O(h^4),
\end{aligned}
\tag{C.81}
$$

where as for $s_1 \neq s_2 \neq l_1 \neq l_2 \neq t_1 \neq t_2$ uniformly, we have

$$
\max_{s_1,s_2,l_1,l_2,t_1,t_2}\iint\frac{f_{s_1,s_2,l_1,l_2,t_1,t_2}(z_1,z_2,z_1,z_2,z_1,z_2)}{f(z_1)^2 f(z_2)^2}dz_1dz_2<\infty. \tag{C.82}
$$

Therefore, $\mathrm{E}\left[\widetilde{G}_2(n)^2\right]=o(1)$ implies that $G_2(n)=o_p(1)$. To summarize,

$$
\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t\overline{u}_t = G_1(n)+G_2(n)=o_p(1). \tag{C.83}
$$

Thus we complete the proof. ■

Proof of Lemma B.1.5 :

We first decompose the equation as

$$
\frac{1}{\sqrt{n}}\left(\widehat{X}'\widehat{e}-\widetilde{X}'\widetilde{e}\right)=\frac{1}{\sqrt{n}}\left(\widehat{X}'\widehat{e}-\widetilde{X}'\widehat{e}+\widetilde{X}'\widehat{e}-\widetilde{X}'\widetilde{e}\right)
$$

$$
\begin{aligned}
&=\frac{1}{\sqrt{n}}\left(\widehat{X}'\widehat{e}-\widetilde{X}'\widehat{e}\right)+\frac{1}{\sqrt{n}}\left(\widetilde{X}'\widehat{e}-\widetilde{X}'\widetilde{e}\right)\\
&=\frac{1}{\sqrt{n}}\left(\widehat{X}-\widetilde{X}\right)'\widehat{e}+\frac{1}{\sqrt{n}}\widetilde{X}'\left(\widehat{e}-\widetilde{e}\right)\\
&=\frac{1}{\sqrt{n}}\left(\widehat{X}-\widetilde{X}\right)'\left(\widehat{e}-\widetilde{e}\right)+\frac{1}{\sqrt{n}}\left(\widehat{X}-\widetilde{X}\right)'\widetilde{e}+\frac{1}{\sqrt{n}}\widetilde{X}'\left(\widehat{e}-\widetilde{e}\right)\\
&=P_1(n)+P_2(n)+P_3(n),
\end{aligned} \tag{C.84}
$$

where $P_1(n)=\left(\widehat{X}-\widetilde{X}\right)'(\widehat{e}-\widetilde{e})/\sqrt{n}$, $P_2(n)=\left(\widehat{X}-\widetilde{X}\right)'\widetilde{e}/\sqrt{n}$, $P_3(n)=\widetilde{X}'(\widehat{e}-\widetilde{e})/\sqrt{n}$. Therefore, it suffices to show that $P_i(n)=o_p(1)$ as $n\to\infty$ for $i=1,2,3$. Note that $\widehat{e}=\widehat{\lambda}(V)+\widehat{U}$, $\widetilde{e}=\widetilde{\lambda}(V)+\widetilde{U}$, and write $\lambda_t=\lambda(v_t)$. We have

$$
\begin{aligned}
P_1(n)=&\frac{1}{\sqrt{n}}\left(\widehat{X}-\widetilde{X}\right)'(\widehat{e}-\widetilde{e})=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\widehat{x}_t-\widetilde{x}_t)(\widehat{e}_t-\widetilde{e}_t)\\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\widehat{x}_t-\widetilde{x}_t)(\widehat{\lambda}_t-\widetilde{\lambda}_t)+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\widehat{x}_t-\widetilde{x}_t)(\widehat{u}_t-\widetilde{u}_t)=P_{11}(n)+P_{12}(n),
\end{aligned} \tag{C.85}
$$

$$
\begin{aligned}
P_2(n)=&\frac{1}{\sqrt{n}}\left(\widehat{X}-\widetilde{X}\right)'\widetilde{e}=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\widehat{x}_t-\widetilde{x}_t)\widetilde{e}_t\\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\widehat{x}_t-\widetilde{x}_t)\widetilde{\lambda}_t+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\widehat{x}_t-\widetilde{x}_t)\widetilde{u}_t=P_{21}(n)+P_{22}(n).
\end{aligned} \tag{C.86}
$$

$$
\begin{aligned}
P_3(n)=&\frac{1}{\sqrt{n}}\widetilde{X}'(\widehat{e}-\widetilde{e})=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t(\widehat{e}_t-\widetilde{e}_t)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t(\widehat{\lambda}_t-\widetilde{\lambda}_t)+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t(\widehat{u}_t-\widetilde{u}_t)\\
=&P_{31}(n)+P_{32}(n).
\end{aligned} \tag{C.87}
$$

Hence, our objective is to show that $P_{ij}(n)=o_p(1)$ for $i=1,2,3, j=1,2$. Note that

$$
P_{32}(n)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{x}_t(\widehat{u}_t-\widetilde{u}_t)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{s=1}^{n}\widetilde{x}_t\left(w_{ns}(t)-\overline{w}_{ns}(t)\right)u_s=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\sum_{s=1}^{n}\widetilde{x}_t\Pi_{s,t}u_s, \tag{C.88}
$$

where $\Pi_{s,t}=\left(w_{ns}(t)-\overline{w}_{ns}(t)\right)$. We define several useful notations as follows.

$$
\Xi_1(r,t)=\frac{K\left(\frac{v_r-v_t}{h}\right)}{f(v_t)}(g(\tau_t)-g(\tau_r)),\quad \Xi_2(r,t)=\frac{K\left(\frac{v_r-v_t}{h}\right)}{f(v_t)}(v_t-v_r),
$$

$$\begin{aligned}\Xi_3(l,t) =& K_2\left(\frac{\tau_l-\tau_t}{b}\right)(g(\tau_l)-g(\tau_t)), \quad \Xi_4(l,t)=K_2\left(\frac{\tau_l-\tau_t}{b}\right)v_t,\\ \Xi_5(p,t,l) =& \Xi_3(l,p)+\Xi_4(l,p)-\Xi_3(l,t)-\Xi_4(l,t).\end{aligned}$$

As $\widetilde{x_t}=\widetilde{g}(\tau_t)+\widetilde{v_t}$, we have

$$\begin{aligned}\widetilde{g}(\tau_t) =& g(\tau_t)-\sum_{r=1}^{n}w_{nr}(t)g(\tau_r)=\frac{1}{nh}\sum_{r=1}^{n}\frac{K\left(\frac{v_r-v_t}{h}\right)}{\widehat{f}(v_t)}(g(\tau_t)-g(\tau_r))\\ =& \frac{1+o_p(1)}{nh}\sum_{r=1}^{n}\frac{K\left(\frac{v_r-v_t}{h}\right)}{f(v_t)}(g(\tau_t)-g(\tau_r))\triangleq\frac{1+o_p(1)}{nh}\sum_{r=1}^{n}\Xi_1(r,t),\end{aligned} \tag{C.89}$$

Similarly,

$$\widetilde{v_t} = \frac{1+o_p(1)}{nh}\sum_{r=1}^{n}\frac{K\left(\frac{v_r-v_t}{h}\right)}{f(v_t)}(v_t-v_r)\triangleq\frac{1+o_p(1)}{nh}\sum_{r=1}^{n}\Xi_2(r,t). \tag{C.90}$$

Also note that

$$\begin{aligned}\Pi_{s,t} =& \frac{K\left(\frac{v_s-v_t}{h}\right)}{nh\widehat{f}(v_t)}-\frac{K\left(\frac{\widehat{v}_s-\widehat{v}_t}{h}\right)}{nh\widehat{f}(\widehat{v}_t)}=\frac{K\left(\frac{v_s-v_t}{h}\right)(\widehat{f}(v_t)-\widehat{f}(\widehat{v}_t))}{nh\widehat{f}(v_t)\widehat{f}(\widehat{v}_t)}+\frac{\left(K\left(\frac{v_t-v_s}{h}\right)-K\left(\frac{\widehat{v}_t-\widehat{v}_s}{h}\right)\right)}{nh\widehat{f}(\widehat{v}_t)}\\ =& \frac{K\left(\frac{v_s-v_t}{h}\right)(\widehat{f}(v_t)-\widehat{f}(\widehat{v}_t))}{nh(f(v_t)^2+o_p(1))}+\frac{\left(K\left(\frac{v_t-v_s}{h}\right)-K\left(\frac{\widehat{v}_t-\widehat{v}_s}{h}\right)\right)}{nh(f(v_t)+o_p(1))}\\ =& (1+o_p(1))\left(\frac{K\left(\frac{v_s-v_t}{h}\right)(\widehat{f}(v_t)-\widehat{f}(\widehat{v}_t))}{nhf(v_t)^2}+\frac{\left(K\left(\frac{v_t-v_s}{h}\right)-K\left(\frac{\widehat{v}_t-\widehat{v}_s}{h}\right)\right)}{nhf(v_t)}\right).\end{aligned} \tag{C.91}$$

By Taylor expansion, and denote $K_d(p,t)=K'\left(\frac{v_p-v_t}{h}\right)$,

$$\begin{aligned}& K\left(\frac{v_p-v_t}{h}\right)-K\left(\frac{\widehat{v}_p-\widehat{v}_t}{h}\right)=(1+o_p(1))K'\left(\frac{v_p-v_t}{h}\right)\left(\left(\frac{v_p-v_t}{h}\right)-\left(\frac{\widehat{v}_p-\widehat{v}_t}{h}\right)\right)\\ =& (1+o_p(1))K_d(p,t)\left(\frac{v_p-\widehat{v}_p}{h}\right)-(1+o_p(1))K_d(p,t)\left(\frac{v_t-\widehat{v}_t}{h}\right)\\ =& (1+o_p(1))h^{-1}K_d(p,t)\left(\widehat{g}(\tau_p)-g(\tau_p)\right)-(1+o_p(1))h^{-1}K_d(p,t)\left(\widehat{g}(\tau_t)-g(\tau_t)\right)\\ =& (1+o_p(1))h^{-1}K_d(p,t)\left(\sum_{l=1}^{n}w_{nl}^{*}(p)x_p-g(\tau_p)\right)\end{aligned}$$

$$
\begin{aligned}
&-(1+o_p(1))h^{-1}K_d(p,t)\left(\sum_{l=1}^{n} w_{nl}^{*}(t)x_t - g(\tau_t)\right) \\
=&\frac{(1+o_p(1))^2}{nhb}K_d(p,t)\left(\sum_{l=1}^{n} K_2\left(\frac{\tau_l-\tau_p}{b}\right)(g(\tau_l)-g(\tau_p)) + \sum_{l=1}^{n} K_2\left(\frac{\tau_l-\tau_p}{b}\right)v_l\right) \\
&-\frac{(1+o_p(1))^2}{nhb}K_d(p,t)\left(\sum_{l=1}^{n} K_2\left(\frac{\tau_l-\tau_t}{b}\right)(g(\tau_l)-g(\tau_t)) + \sum_{l=1}^{n} K_2\left(\frac{\tau_l-\tau_t}{b}\right)v_t\right) \\
\triangleq&\frac{(1+o_p(1))^2}{nhb}K_d(p,t)\sum_{l=1}^{n}\Big(\Xi_3(l,p)+\Xi_4(l,p)-\Xi_3(l,t)-\Xi_4(l,t)\Big) \\
\triangleq&\frac{(1+o_p(1))^2}{nhb}K_d(p,t)\sum_{l=1}^{n}\Xi_5(p,t,l),
\end{aligned}
\tag{C.92}
$$

where $\Xi_5(p,t,l) = \Xi_3(l,p)+\Xi_4(l,p)-\Xi_3(l,t)-\Xi_4(l,t)$, Meanwhile,

$$
\begin{aligned}
\widehat{f}(v_t) - \widehat{f}(\widehat{v_t}) =& \frac{1}{nh}\sum_{p=1}^{n}\left(K\left(\frac{v_p-v_t}{h}\right) - K\left(\frac{\widehat{v_p}-\widehat{v_t}}{h}\right)\right) \\
=&\frac{(1+o_p(1))}{n^2h^2b}\sum_{p=1}^{n}K_d(p,t)\sum_{l=1}^{n}\Xi_5(p,t,l).
\end{aligned}
\tag{C.93}
$$

Therefore,

$$
\begin{aligned}
\Pi_{s,t} =& (1+o_p(1))\left(\frac{K\left(\frac{v_s-v_t}{h}\right)}{n^3h^3bf(v_t)^2}\sum_{p=1}^{n}K_d(p,t)\sum_{l=1}^{n}\Xi_5(p,t,l)\right) \\
&+(1+o_p(1))\left(\frac{K_d(s,t)\sum_{l=1}^{n}\Xi_5(s,t,l)}{n^2h^2bf(v_t)}\right).
\end{aligned}
\tag{C.94}
$$

To summarize,

$$
\begin{aligned}
P_{32}(n) =& \frac{(1+o_p(1))}{\sqrt{n}}\sum_{t=1}^{n}\sum_{s=1}^{n}\left(\frac{1}{nh}\sum_{r=1}^{n}\Xi_1(r,t)+\Xi_2(r,t)\right) \\
&\left(\frac{K\left(\frac{v_s-v_t}{h}\right)}{n^3h^3bf(v_t)^2}\sum_{p=1}^{n}K_d(p,t)\sum_{l=1}^{n}\Xi_5(p,t,l) + \frac{K_d(s,t)\sum_{l=1}^{n}\Xi_5(s,t,l)}{n^2h^2bf(v_t)}\right)u_s \\
=&\frac{(1+o_p(1))}{\sqrt{n}n^4h^4b}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{r=1}^{n}\sum_{p=1}^{n}\sum_{l=1}^{n}\frac{K\left(\frac{v_s-v_t}{h}\right)K_d(p,t)}{f(v_t)^2}\Big(\Xi_1(r,t)+\Xi_2(r,t)\Big)\Xi_5(p,t,l)u_s \\
&+\frac{(1+o_p(1))}{\sqrt{n}n^3h^3b}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{r=1}^{n}\sum_{l=1}^{n}\frac{K_d(s,t)}{f(v_t)}\Big(\Xi_1(r,t)+\Xi_2(r,t)\Big)\Xi_5(s,t,l)u_s.
\end{aligned}
\tag{C.95}
$$

Replacing $\Xi_5(p,t,l) = \Xi_3(l,p) + \Xi_4(l,p) - \Xi_3(l,t) - \Xi_4(l,t)$, it is equivalent to prove that the following quantities are $o_p(1)$.

$$\mathfrak{W}_1 = \frac{1}{\sqrt{n}n^4h^4b}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{r=1}^{n}\sum_{p=1}^{n}\sum_{l=1}^{n}\frac{K\left(\frac{v_s-v_t}{h}\right)K_d(p,t)}{f(v_t)^2}\Big(\Xi_1(r,t)+\Xi_2(r,t)\Big)$$
$$\Big(\Xi_3(l,p)+\Xi_4(l,p)-\Xi_3(l,t)-\Xi_4(l,t)\Big)u_s. \tag{C.96}$$

$$\mathfrak{W}_2 = \frac{1}{\sqrt{n}n^3h^3b}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{r=1}^{n}\sum_{l=1}^{n}\frac{K_d(s,t)}{f(v_t)}\Big(\Xi_1(r,t)+\Xi_2(r,t)\Big)$$
$$\Big(\Xi_3(l,p)+\Xi_4(l,p)-\Xi_3(l,t)-\Xi_4(l,t)\Big)u_s. \tag{C.97}$$

We consider one typical term, for example,

$$\mathfrak{w}_1 = \frac{1}{\sqrt{n}n^4h^4b}\sum_{t=1}^{n}\sum_{s=1}^{n}\sum_{r=1}^{n}\sum_{p=1}^{n}\sum_{l=1}^{n}\frac{K\left(\frac{v_s-v_t}{h}\right)K_d(p,t)}{f(v_t)^2}\Xi_1(r,t)\Xi_3(l,p)u_s. \tag{C.98}$$

It suffices to show that $\mathfrak{w}_1$ is $o_p(1)$ if $\mathrm{E}[\mathfrak{w}_1^2] \to 0$ as $n \to \infty$. We only consider the condition that all subscripts are not equal (denoted as Θ_{10}), and it is straightforward that the conclusion still holds for the rest of the conditions.

$$\begin{aligned}
\Big|\mathrm{E}[\mathfrak{w}_1^2]\Big| = &\Big|\mathrm{E}\Big[\frac{1}{n^9h^8b^2}\sum_{\Theta_{10}}\frac{K\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K'\left(\frac{v_{p_1}-v_{t_1}}{h}\right)K\left(\frac{v_{r_1}-v_{t_1}}{h}\right)}{f(v_{t_1})^3}\\
&\frac{K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)K'\left(\frac{v_{p_2}-v_{t_2}}{h}\right)K\left(\frac{v_{r_2}-v_{t_2}}{h}\right)}{f(v_{t_2})^3}(g(\tau_{t_1})-g(\tau_{r_1}))(g(\tau_{t_2})-g(\tau_{r_2}))\\
&K_2\left(\frac{\tau_{l_1}-\tau_{p_1}}{b}\right)(g(\tau_{l_1})-g(\tau_{p_1}))K_2\left(\frac{\tau_{l_2}-\tau_{p_2}}{b}\right)(g(\tau_{l_2})-g(\tau_{p_2}))u_{s_1}u_{s_2}\Big]\Big|\\
\leq&\frac{1}{n^9h^8b^2}\sum_{\Theta_{10}}\Big|(g(\tau_{t_1})-g(\tau_{r_1}))(g(\tau_{t_2})-g(\tau_{r_2}))\Big|\Big|\mathrm{E}\Big[u_{s_1}u_{s_2}\Big]\\
&\mathrm{E}\Big[\frac{K\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K'\left(\frac{v_{p_1}-v_{t_1}}{h}\right)K\left(\frac{v_{r_1}-v_{t_1}}{h}\right)}{f(v_{t_1})^3}\frac{K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)K'\left(\frac{v_{p_2}-v_{t_2}}{h}\right)K\left(\frac{v_{r_2}-v_{t_2}}{h}\right)}{f(v_{t_2})^3}\Big]\Big|\\
&\Big|K_2\left(\frac{\tau_{l_1}-\tau_{p_1}}{b}\right)(g(\tau_{l_1})-g(\tau_{p_1}))K_2\left(\frac{\tau_{l_2}-\tau_{p_2}}{b}\right)(g(\tau_{l_2})-g(\tau_{p_2}))\Big|\\
\leq&\frac{1}{n^9h^8b^2}\sum_{\Theta_{10}}\Big|(g(\tau_{t_1})-g(\tau_{r_1}))(g(\tau_{t_2})-g(\tau_{r_2}))\Big|\Big|\mathrm{E}\Big[u_{s_1}u_{s_2}\Big]\Big|
\end{aligned}$$

$$Ch^6\left|K_2\left(\frac{\tau_{l_1}-\tau_{p_1}}{b}\right)(g(\tau_{l_1})-g(\tau_{p_1}))K_2\left(\frac{\tau_{l_2}-\tau_{p_2}}{b}\right)(g(\tau_{l_2})-g(\tau_{p_2}))\right|$$
$$=\frac{C}{n^5h^2b^2}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\left|\mathrm{E}\left[u_{s_1}u_{s_2}\right]\right|\sum_{l_1=1}^{n}\sum_{p_1=1}^{n}\left|K_2\left(\frac{\tau_{l_1}-\tau_{p_1}}{b}\right)(g(\tau_{l_1})-g(\tau_{p_1}))\right|$$
$$\sum_{l_2=1}^{n}\sum_{p_2=1}^{n}\left|K_2\left(\frac{\tau_{l_2}-\tau_{p_2}}{b}\right)(g(\tau_{l_2})-g(\tau_{p_2}))\right|$$
$$=\frac{Cb^2}{nh^2}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\left|\mathrm{E}\left[u_{s_1}u_{s_2}\right]\right|$$
$$\leq\frac{Cb^2}{nh^2}\sum_{s_1=1}^{n}\sum_{s_2=1}^{n}\alpha^{\frac{\delta}{2+\delta}}(s_1-s_2)\mathrm{E}|u_{s_1}|^{2+\delta}\mathrm{E}|u_{s_2}|^{2+\delta}=O(b^2/h^2), \tag{C.99}$$

where we used the results as follows.

$$\left|\mathrm{E}\left[\frac{K\left(\frac{v_{s_1}-v_{t_1}}{h}\right)K'\left(\frac{v_{p_1}-v_{t_1}}{h}\right)K\left(\frac{v_{r_1}-v_{t_1}}{h}\right)K\left(\frac{v_{s_2}-v_{t_2}}{h}\right)K'\left(\frac{v_{p_2}-v_{t_2}}{h}\right)K\left(\frac{v_{r_2}-v_{t_2}}{h}\right)}{f(v_{t_1})^3f(v_{t_2})^3}\right]\right|$$
$$=\left|\int\cdots\int\frac{K(w_1)K'(w_2)K(w_3)K(w_4)K'(w_5)K(w_6)}{f(z_1)^3f(z_2)^3}f_{s_1,p_1,r_1,s_2,p_2,r_2,t_1,t_2}(z_1+w_1h,z_1\right.$$
$$\left.+w_2h,z_1+w_3h,z_2+w_4h,z_2+w_5h,z_2+w_6h,z_1,z_2)h^6dw_1dw_2dw_3dw_4dw_5dw_6dz_1dz_2\right|$$
$$\leq\int\cdots\int\left|\frac{K(w_1)K'(w_2)K(w_3)K(w_4)K'(w_5)K(w_6)}{f(z_1)^3f(z_2)^3}f_{s_1,p_1,r_1,s_2,p_2,r_2,t_1,t_2}(z_1+w_1h,\right.$$
$$\left.z_1+w_2h,z_1+w_3h,z_2+w_4h,z_2+w_5h,z_2+w_6h,z_1,z_2)h^6\right|dw_1dw_2dw_3dw_4dw_5dw_6dz_1dz_2$$
$$=(1+o(1))h^6\int\cdots\int\left|\frac{K(w_1)K'(w_2)K(w_3)K(w_4)K'(w_5)K(w_6)}{f(z_1)^3f(z_2)^3}\right|\left|f_{s_1,p_1,r_1,s_2,p_2,r_2,t_1,t_2}\right.$$
$$\left.(z_1,z_1,z_1,z_2,z_2,z_2,z_1,z_2)\right|dw_1dw_2dw_3dw_4dw_5dw_6dz_1dz_2$$
$$=(1+o(1))h^6\int|K(w_1)|dw_1\int|K'(w_2)|dw_2\int|K(w_3)|dw_3\int|K(w_4)|dw_4\int|K'(w_5)|$$
$$dw_5\int|K(w_6)|dw_6\iint\left|\frac{f_{s_1,p_1,r_1,s_2,p_2,r_2,t_1,t_2}(z_1,z_1,z_1,z_2,z_2,z_2,z_1,z_2)}{f(z_1)^3f(z_2)^3}\right|dz_1dz_2$$
$$\leq(1+o(1))h^6\left(\int|K(w_1)|dw_1\right)^4\left(\int|K'(w_2)|dw_2\right)^2$$
$$\iint\left|\frac{f_{s_1,p_1,r_1,s_2,p_2,r_2,t_1,t_2}(z_1,z_1,z_1,z_2,z_2,z_2,z_1,z_2)}{f(z_1)^3f(z_2)^3}\right|dz_1dz_2\leq(1+o(1))Ch^6, \tag{C.100}$$

for some $C>0$ and

$$\max_{s_1,p_1,r_1,s_2,p_2,r_2,t_1,t_2}\iint\frac{f_{s_1,p_1,r_1,s_2,p_2,r_2,t_1,t_2}(z_1,z_1,z_1,z_2,z_2,z_2,z_1,z_2)}{f(z_1)^3f(z_2)^3}dz_1dz_2<\infty. \tag{C.101}$$

In the same time, by the definition of Riemann Integral,

$$\begin{aligned}
&\frac{1}{n^2}\sum_{l=1}^{n}\sum_{p=1}^{n}\left|K_2\left(\frac{\tau_l-\tau_p}{b}\right)(g(\tau_l)-g(\tau_p))\right| = \frac{1}{n^2}\sum_{l=1}^{n}\sum_{p=1}^{n}\left|K_2\left(\frac{\tau_l-\tau_p}{b}\right)(\tau_l-\tau_p)g'(\tau_p)\right| \\
\to&\iint\left|K_2\left(\frac{y-x}{b}\right)(y-x)\right|\left|g'(x)\right|dydx = b\iint\left|K_2\left(\frac{y-x}{b}\right)\left(\frac{y-x}{b}\right)\right|\left|g'(x)\right|dydx \\
=&b^2\int\left|K_2(w)w\right|dw\int\left|g'(x)\right|dx = O(b^2). \qquad \text{(C.102)}
\end{aligned}$$

where $\int|K_2(w)w|dw<\infty$, and $\int|g'(x)|dx<\infty$. Thus we complete the proof. ■

C.2 Proofs of the Lemmas in Appendix B.2

Proof of Lemma B.2.1:

$$\begin{aligned}
&\frac{1}{n^{\frac{d_i}{2}}}\sum_{t=1}^{n}g_i(t)(\widetilde{\epsilon}_{t-1}-\widetilde{\epsilon}_t) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{g_i(t)}{n^{\frac{d_i-1}{2}}}(\widetilde{\epsilon}_{t-1}-\widetilde{\epsilon}_t) \\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{g}_i(\tau_t)(\widetilde{\epsilon}_{t-1}-\widetilde{\epsilon}_t) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\widetilde{g}_i(\tau_t)\widetilde{\epsilon}_{t-1}-\widetilde{g}_i(\tau_t)\widetilde{\epsilon}_t\right) \\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\widetilde{g}_i(\tau_t)\widetilde{\epsilon}_{t-1}-\widetilde{g}_i(\tau_{t-1})\widetilde{\epsilon}_{t-1}+\widetilde{g}_i(\tau_{t-1})\widetilde{\epsilon}_{t-1}-\widetilde{g}_i(\tau_t)\widetilde{\epsilon}_t\right) \\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\widetilde{g}_i(\tau_t)\widetilde{\epsilon}_{t-1}-\widetilde{g}_i(\tau_{t-1})\widetilde{\epsilon}_{t-1}\right)+\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\widetilde{g}_i(\tau_{t-1})\widetilde{\epsilon}_{t-1}-\widetilde{g}_i(\tau_t)\widetilde{\epsilon}_t\right) \\
=&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\widetilde{g}_i(\tau_t)-\widetilde{g}_i(\tau_{t-1})\right)\widetilde{\epsilon}_{t-1}+\frac{1}{\sqrt{n}}\left(\widetilde{g}_i(\tau_0)\widetilde{\epsilon}_0-\widetilde{g}_i(\tau_n)\widetilde{\epsilon}_n\right), \qquad \text{(C.103)}
\end{aligned}$$

where the second term is $o_p(1/\sqrt{n})$. For the first term, by Taylor expansion,

$$\zeta_3(n) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\widetilde{g}_i(\tau_t)-\widetilde{g}_i(\tau_{t-1})\right)\widetilde{\epsilon}_{t-1} = \frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\widetilde{g}_i'(\tau_{t-1})\widetilde{\epsilon}_{t-1}. \qquad \text{(C.104)}$$

It is easy to show that $\mathrm{E}[\zeta_3(n)] = 0$ and $\mathrm{E}[\zeta_3(n)^2] = O(n^{-2})$. Therefore, as $n\to\infty$, we have

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\widetilde{g}_i(\tau_t)(\widetilde{\epsilon}_{t-1}-\widetilde{\epsilon}_t) \longrightarrow_P 0.$$

We then complete the proof. ■

Proof of Lemma B.2.2: We need to show that as $n \to \infty$,

$$\frac{1}{n^{d_i/2}}\widetilde{\epsilon\eta}^f_{i,0} \longrightarrow_P 0, \tag{C.105}$$
$$\frac{1}{n^{d_i/2}}\widetilde{\epsilon\eta}^f_{i,n} \longrightarrow_P 0, \tag{C.106}$$

where $\widetilde{\epsilon\eta}^f_{i,t} = \widetilde{f}_{i,0}(L)(\epsilon_t\eta_{it} - \theta_i) = \sum_{s=0}^{\infty}\widetilde{f}_{i,0s}(\epsilon_{t-s}\eta_{i,t-s} - \theta_i)$ and $\widetilde{f}_{i,0s} = \sum_{k=s+1}^{\infty}\phi_k\psi_{k,i}$. It is obvious that

$$\begin{aligned}
\mathrm{E}\left[\frac{1}{n^{d_i/2}}\widetilde{\epsilon\eta}^f_{i,n}\right] &= \frac{1}{n^{d_i/2}}\mathrm{E}\left[\sum_{s=0}^{\infty}\widetilde{f}_{i,0s}(\epsilon_{t-s}\eta_{i,t-s} - \theta_i)\right] \\
&= \frac{1}{n^{d_i/2}}\sum_{s=0}^{\infty}\widetilde{f}_{i,0s}(\mathrm{E}\left[\epsilon_{t-s}\eta_{i,t-s}\right] - \theta_i) = 0.
\end{aligned} \tag{C.107}$$

Meanwhile,

$$\begin{aligned}
\mathrm{E}\left[\left(\frac{1}{n^{d_i/2}}\widetilde{\epsilon\eta}^f_{i,n}\right)^2\right] &= \frac{1}{n^{d_i}}\sum_{s=0}^{\infty}\widetilde{f}^2_{i,0s}\mathrm{E}[(\epsilon_{t-s}\eta_{i,t-s} - \theta_i)^2] \\
&\quad + \frac{2}{n^{d_i}}\sum_{s_1=0}^{\infty}\sum_{s_2=s_1+1}^{\infty}\widetilde{f}_{i,0s_1}\widetilde{f}_{i,0s_2}\mathrm{E}[(\epsilon_{t-s_1}\eta_{i,t-s_1} - \theta_i)(\epsilon_{t-s_2}\eta_{i,t-s_2} - \theta_i)] \\
&= \frac{\widetilde{\delta}_{22} - \theta_i^2}{n^{d_i}}\sum_{s=0}^{\infty}\widetilde{f}^2_{i,0s} = O(n^{-d_i}),
\end{aligned} \tag{C.108}$$

given that $\sum_{s=0}^{\infty}\widetilde{f}^2_{i,0s} < \infty$. Therefore, as $n \to \infty$, (C.105) holds. Similarly, we can prove (C.106) given the same condition. Thus we complete the proof for Lemma B.2.2. ■

Proof of Lemma B.2.3 and B.2.4 In these two Lemmas, we define

$$B^f_{i,t} = \sum_{q=1}^{\infty}\widetilde{f}_{i,q}(L)\epsilon_t\eta_{i,t-q}, \tag{C.109}$$
$$B^m_{i,t} = \sum_{q=1}^{\infty}\widetilde{m}_{i,q}(L)\epsilon_{t-q}\eta_{it}, \tag{C.110}$$

where $\widetilde{f}_{i,q}(L) = \sum_{s=0}^{\infty}\widetilde{f}_{i,qs}L^s$, $\widetilde{f}_{i,qs} = \sum_{p=s+1}^{\infty}\phi_p\psi_{p+q,i}$ and $\widetilde{m}_{i,q}(L) = \sum_{s=0}^{\infty}\widetilde{m}_{i,qs}L^s$, $\widetilde{m}_{i,qs} = \sum_{p=s+1}^{\infty}\phi_{p+q}\psi_{p,i}$. Therefore, it is obvious that for any t,

$$\mathrm{E}\left[\frac{1}{n^{d_i/2}}B^f_{i,t}\right] = \mathrm{E}\left[\frac{1}{n^{d_i/2}}\sum_{q=1}^{\infty}\sum_{s=0}^{\infty}f_{i,qs}\epsilon_{t-s}\eta_{i,t-q-s}\right]$$

$$=\frac{1}{n^{d_i/2}}\sum_{q=1}^{\infty}\sum_{s=0}^{\infty}f_{i,ks}\mathrm{E}[\epsilon_{t-s}\eta_{i,t-q-s}]=0, \tag{C.111}$$

and

$$\begin{aligned}\mathrm{E}\left[\frac{1}{n^{d_i/2}}B^m_{i,t}\right]=&\mathrm{E}\left[\frac{1}{n^{d_i/2}}\sum_{q=1}^{\infty}\sum_{s=0}^{\infty}m_{i,qs}\epsilon_{t-q-s}\eta_{i,t-s}\right]\\=&\frac{1}{n^{d_i/2}}\sum_{q=1}^{\infty}\sum_{s=0}^{\infty}m_{i,qs}\mathrm{E}[\epsilon_{t-q-s}\eta_{i,t-s}]=0.\end{aligned} \tag{C.112}$$

While for the second moments,

$$\mathrm{E}\left[\left(\frac{1}{n^{d_i/2}}B^f_{i,t}\right)^2\right]=\frac{1}{n^{d_i}}\sum_{q=1}^{\infty}\sum_{s=0}^{\infty}f^2_{i,qs}\mathrm{E}[\epsilon^2_{t-s}\eta^2_{i,t-q-s}]=O(n^{-d_i}), \tag{C.113}$$

given that $\sum_{q=1}^{\infty}\sum_{s=0}^{\infty}\widetilde{f}^2_{i,qs}<\infty$. Similarly, we can show that

$$\mathrm{E}\left[\left(\frac{1}{n^{d_i/2}}B^m_{i,t}\right)^2\right]=\frac{1}{n^{d_i}}\sum_{q=1}^{\infty}\sum_{s=0}^{\infty}m^2_{i,qs}\mathrm{E}[\epsilon^2_{t-s-q}\eta^2_{i,t-s}]=O(n^{-d_i}), \tag{C.114}$$

given that $\sum_{q=1}^{\infty}\sum_{s=0}^{\infty}\widetilde{m}^2_{i,qs}<\infty$. Therefore,

$$\frac{1}{n^{d_i/2}}B^m_{i,t}=o_P(1), \tag{C.115}$$

$$\frac{1}{n^{d_i/2}}B^f_{i,t}=o_P(1), \tag{C.116}$$

for $t=0$ and $t=n$.

Thus we complete the proof for Lemma B.2.3 and B.2.4. ∎

Proof of Lemma B.2.5: Denote $\boldsymbol{Z}_3=(a'M_{nt})^2$, then

$$\boldsymbol{Z}_3=\sum_{i=1}^{k}\sum_{j=1}^{K}a_ia_jM^i_{nt}M^j_{nt}. \tag{C.117}$$

Note that

$$\begin{aligned}&M^i_{nt}M^j_{nt}\\=&n^{-d_i/2}\left(g_i(\tau_t)\Phi(1)\epsilon_t+f_{i,0}(1)\left(\epsilon_t\eta_{it}-\theta_i\right)+\sum_{q=1}^{\infty}f_{i,q}(1)\epsilon_t\eta_{i,t-q}+\sum_{q=1}^{\infty}m_{i,q}(1)\epsilon_{t-q}\eta_{i,t}\right)\end{aligned}$$

$$\cdot n^{-d_j/2}\left(g_j(\tau_t)\Phi(1)\epsilon_t + f_{j,0}(1)\left(\epsilon_t\eta_{jt} - \theta_j\right) + \sum_{q=1}^{\infty} f_{j,q}(1)\epsilon_t\eta_{j,t-q} + \sum_{q=1}^{\infty} m_{j,q}(1)\epsilon_{t-q}\eta_{j,t}\right)$$

$$\triangleq \left(M_{1,nt}^i + M_{2,nt}^i + M_{3,nt}^i + M_{4,nt}^i\right)\left(M_{1,nt}^j + M_{2,nt}^j + M_{3,nt}^j + M_{4,nt}^j\right), \tag{C.118}$$

where for $i = 1, 2, ..., k$,

$$M_{1,nt}^i = n^{-d_i/2} g_i(\tau_t)\Phi(1)\epsilon_t, \tag{C.119}$$

$$M_{2,nt}^i = n^{-d_i/2} f_{i,0}(1)\left(\epsilon_t\eta_{it} - \theta_i\right), \tag{C.120}$$

$$M_{3,nt}^i = n^{-d_i/2}\sum_{q=1}^{\infty} f_{i,q}(1)\epsilon_t\eta_{i,t-q}, \tag{C.121}$$

$$M_{4,nt}^i = n^{-d_i/2}\sum_{q=1}^{\infty} m_{i,q}(1)\epsilon_{t-q}\eta_{i,t}. \tag{C.122}$$

Then

$$\begin{aligned}\sum_{t=1}^{n} \mathrm{E}\left[(a' M_{nt})^2 \middle| F_{t-1}\right] &= \sum_{t=1}^{n}\sum_{i=1}^{k}\sum_{j=1}^{k}\sum_{r_1=1}^{4}\sum_{r_2=1}^{4} \mathrm{E}\left[a_i a_j M_{r_1,nt}^i M_{r_2,nt}^j \middle| F_{t-1}\right] \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \sum_{r_1=1}^{4}\sum_{r_2=1}^{4} \frac{1}{n^{d_{ij}}}\sum_{t=1}^{n} \mathrm{E}\left[M_{r_1,nt}^i M_{r_2,nt}^j \middle| F_{t-1}\right] \\ &= \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \sum_{r_1=1}^{4}\sum_{r_2=1}^{4} \boldsymbol{Z}_4(i,j,r_1,r_2), \end{aligned} \tag{C.123}$$

where $\boldsymbol{Z}_4(i,j,r_1,r_2) \triangleq n^{-d_{ij}}\sum_{t=1}^{n} \mathrm{E}\left[M_{r_1,nt}^i M_{r_2,nt}^j \middle| F_{t-1}\right]$. For given $i, j = 1, 2, ..., k$, we analyze the terms in (C.123) one by one with respect to $r_1, r_2 = 1, 2, 3, 4$.

(1) When $r_1 = 1, r_2 = 1$, as $n \to \infty$, we have

$$\begin{aligned}&\sum_{t=1}^{n} \mathrm{E}\left[M_{1,nt}^i M_{1,nt}^j \middle| F_{t-1}\right] \\ =&\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n} \mathrm{E}[g_i(\tau_t)g_j(\tau_t)\Phi(1)^2\epsilon_t^2 | F_{t-1}] \\ =&\frac{\sigma_1^2\Phi(1)^2}{n^{d_{ij}}}\sum_{t=1}^{n} g_i(\tau_t)g_j(\tau_t) \longrightarrow \sigma_1^2\Phi(1)^2\boldsymbol{Q}_{ij}.\end{aligned} \tag{C.124}$$

(2) When $r_1 = 2, r_2 = 2$, as $n \to \infty$, we have

$$\sum_{t=1}^{n} \mathrm{E}\left[M_{2,nt}^i M_{2,nt}^j \middle| F_{t-1}\right]$$

$$=f_{i,0}(1)f_{j,0}(1)\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n}\mathrm{E}[\left(\epsilon_t\eta_{it}-\theta_i\right)\left(\epsilon_t\eta_{jt}-\theta_j\right)]. \tag{C.125}$$

when $d_{ij}=1$,

$$\begin{aligned}\sum_{t=1}^{n}\mathrm{E}\left[M_{2,nt}^{i}M_{2,nt}^{j}\middle|F_{t-1}\right] &=f_{i,0}(1)f_{j,0}(1)(\mathrm{E}[\epsilon_t^2\eta_{it}\eta_{jt}]-\theta_i\theta_j)\\ &=f_{i,0}(1)f_{j,0}(1)(\delta_{2ij}-\theta_i\theta_j),\end{aligned} \tag{C.126}$$

where $\delta_{2ij}=\mathrm{E}[\epsilon_t^2\eta_{it}\eta_{jt}]$.

When $d_{ij}>1$,

$$\sum_{t=1}^{n}\mathrm{E}\left[M_{2,nt}^{i}M_{2,nt}^{j}\middle|F_{t-1}\right]=f_{i,0}(1)f_{j,0}(1)\frac{\mathrm{E}[\epsilon_t^2\eta_{it}\eta_{jt}]-\theta_i\theta_j}{n^{d_{ij}-1}}\longrightarrow 0. \tag{C.127}$$

(3) When $r_1=1,r_2=2$ or $r_1=2,r_2=1$, as $n\to\infty$, we have

$$\begin{aligned}&\sum_{t=1}^{n}\mathrm{E}\left[M_{1,nt}^{i}M_{2,nt}^{j}\middle|F_{t-1}\right]\\ =&\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n}\mathrm{E}[g_i(\tau_t)\Phi(1)\epsilon_t f_{j,0}(1)\left(\epsilon_t\eta_{jt}-\theta_j\right)|F_{t-1}]\\ =&\frac{\Phi(1)f_{j,0}(1)}{n^{d_{ij}}}\sum_{t=1}^{n}g_i(\tau_t)\mathrm{E}[\epsilon_t\left(\epsilon_t\eta_{jt}-\theta_j\right)]\\ =&\Phi(1)f_{j,0}(1)\delta_{2j}\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n}g_i(\tau_t)\\ =&\Phi(1)f_{j,0}(1)\delta_{2j}\frac{1}{n^{(d_j-1)/2}}\frac{1}{n}\sum_{t=1}^{n}\frac{g_i(\tau_t)}{n^{(d_i-1)/2}}\end{aligned} \tag{C.128}$$

where $\delta_{2j}=\mathrm{E}[\epsilon_t^2\eta_{jt}]$. When $d_j=1$,

$$\sum_{t=1}^{n}\mathrm{E}\left[M_{1,nt}^{i}M_{2,nt}^{j}\middle|F_{t-1}\right]\longrightarrow\Phi(1)f_{j,0}(1)\delta_{2j}\overline{g}_i, \tag{C.129}$$

in which $\overline{g}_i=\int_0^1 g_i^N(\tau)d\tau$.

When $d_j>1$,

$$\sum_{t=1}^{n}\sum_{i=1}^{K}\sum_{j=1}^{K}\mathrm{E}\left[a_ia_jM_{1,nt}^{i}M_{2,nt}^{j}\middle|F_{t-1}\right]\longrightarrow 0. \tag{C.130}$$

Similarly result holds for $r_1 = 2, r_2 = 1$.

(4) When $r_1 = r_2 = 3$, as $n \to \infty$, we have

$$
\begin{aligned}
&\sum_{t=1}^{n} \mathrm{E}\left[M_{3,nt}^{i} M_{3,nt}^{j} \middle| F_{t-1}\right] \\
=&\frac{1}{n^{d_{ij}}} \sum_{t=1}^{n} \mathrm{E}\left[\sum_{q_1=1}^{\infty} f_{i,q_1}(1)\epsilon_t \eta_{i,t-q_1} \sum_{q_2=1}^{\infty} f_{j,q_2}(1)\epsilon_t \eta_{j,t-q_2} \middle| F_{t-1}\right] \\
=&\frac{\sigma_1^2}{n^{d_{ij}}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} \sum_{q_2=1}^{\infty} f_{i,q_1}(1) f_{j,q_2}(1) \eta_{i,t-q_1} \eta_{j,t-q_2} \\
=&\frac{\sigma_1^2}{n^{d_{ij}}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) \eta_{i,t-q_1} \eta_{j,t-q_1} + \frac{\sigma_1^2}{n^{d_{ij}}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} \sum_{\substack{q_2=1 \\ q_2 \neq q_1}}^{\infty} f_{i,q_1}(1) f_{j,q_2}(1) \eta_{i,t-q_1} \eta_{j,t-q_2} \\
\triangleq& U_{11}(n) + U_{12}(n).
\end{aligned} \tag{C.131}
$$

Note that when $d_{ij} = 1$,

$$
\mathrm{E}[U_{11}(n)] = \frac{\sigma_1^2}{n^{d_{ij}}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) \mathrm{E}\left[\eta_{i,t-q_1} \eta_{j,t-q_1}\right] = \sigma_1^2 \sigma_{ij} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1), \tag{C.132}
$$

and

$$
\mathrm{E}[U_{12}(n)] = \frac{\sigma_1^2}{n} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} \sum_{\substack{q_2=1 \\ q_2 \neq q_1}}^{\infty} f_{i,q_1}(1) f_{j,q_2}(1) \mathrm{E}[\eta_{i,t-q_1} \eta_{j,t-q_2}] = 0. \tag{C.133}
$$

Meanwhile,

$$
\begin{aligned}
Var[U_{11}(n)] =& \frac{\sigma_1^4}{n^2} \mathrm{E}\left[\left(\sum_{t=1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) (\eta_{i,t-q_1} \eta_{j,t-q_1} - \sigma_{ij})\right)^2\right] \\
=& \frac{\sigma_1^4}{n^2} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \sum_{q_1=1}^{\infty} \sum_{q_2=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) f_{i,q_2}(1) f_{j,q_2}(1) \\
& \mathrm{E}\left[(\eta_{i,t_1-q_1} \eta_{j,t_1-q_1} - \sigma_{ij})(\eta_{i,t_2-q_2} \eta_{j,t_2-q_2} - \sigma_{ij})\right] \\
=& \frac{\sigma_1^4}{n^2} \sum_{t_1=1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1)^2 f_{j,q_1}(1)^2 \mathrm{E}\left[(\eta_{i,t_1-q_1} \eta_{j,t_1-q_1} - \sigma_{ij})^2\right] \\
& + \frac{2\sigma_1^4}{n^2} \sum_{t_1=1}^{n-1} \sum_{t_2=t_1+1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) f_{i,t_2-t_1+q_1}(1) f_{j,t_2-t_1+q_1}(1)
\end{aligned}
$$

$$
\begin{aligned}
&\mathrm{E}\left[(\eta_{i,t_1-q_1}\eta_{j,t_1-q_1}-\sigma_{ij})^2\right]\\
=&\frac{\sigma_1^4(\delta_{ij}-\sigma_{ij}^2)}{n^2}\sum_{t_1=1}^{n}\sum_{q_1=1}^{\infty}f_{i,q_1}(1)^2f_{j,q_1}(1)^2\\
&+\frac{2\sigma_1^4(\delta_{ij}-\sigma_{ij}^2)}{n^2}\sum_{t_1=1}^{n-1}\sum_{p=1}^{n-t_1}\sum_{q_1=1}^{\infty}f_{i,q_1}(1)f_{j,q_1}(1)f_{i,p+q_1}(1)f_{j,p+q_1}(1)\\
=&O(n^{-1}),
\end{aligned}
\tag{C.134}
$$

given that $\sum_{q_1=1}^{\infty}f_{i,q_1}(1)^2f_{j,q_1}(1)^2<\infty$ and $\sum_{p=1}^{\infty}\sum_{q_1=1}^{\infty}|f_{i,q_1}(1)f_{j,q_1}(1)f_{i,p+q_1}(1)f_{j,p+q_1}(1)|<\infty$. Also,

$$
\begin{aligned}
&Var[U_{12}(n)]\\
=&\frac{\sigma_1^4}{n^2}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty}\sum_{l_1=1}^{\infty}\sum_{\substack{l_2=1\\l_2\neq l_1}}^{\infty}f_{i,q_1}(1)f_{j,q_2}(1)f_{i,l_1}(1)f_{j,l_2}(1)\mathrm{E}[\eta_{i,t_1-q_1}\eta_{j,t_1-q_2}\eta_{i,t_2-l_1}\eta_{j,t_2-l_2}]\\
=&\frac{\sigma_1^4}{n^2}\sum_{t_1=1}^{n}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty}f_{i,q_1}(1)^2f_{j,q_2}(1)^2\mathrm{E}[\eta_{i,t_1-q_1}^2\eta_{j,t_1-q_2}^2]\\
&+\frac{2\sigma_1^4}{n^2}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty}f_{i,q_1}(1)f_{j,q_2}(1)f_{i,t_2-t_1+q_1}(1)f_{j,t_2-t_1+q_2}(1)\mathrm{E}[\eta_{i,t_1-q_1}^2\eta_{j,t_1-q_2}^2]\\
=&\frac{\sigma_1^4\sigma_{ii}\sigma_{jj}}{n^2}\sum_{t_1=1}^{n}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty}f_{i,q_1}(1)^2f_{j,q_2}(1)^2\\
&+\frac{2\sigma_1^4\sigma_{ii}\sigma_{jj}}{n^2}\sum_{t_1=1}^{n-1}\sum_{p=1}^{n-t_1}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty}f_{i,q_1}(1)f_{j,q_2}(1)f_{i,p+q_1}(1)f_{j,p+q_2}(1)\\
=&O(n^{-1}),
\end{aligned}
\tag{C.135}
$$

given that $\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty}f_{i,q_1}(1)^2f_{j,q_2}(1)^2<\infty$ and

$$\sum_{p=1}^{\infty}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty}f_{i,q_1}(1)f_{j,q_2}(1)f_{i,p+q_1}(1)f_{j,p+q_2}(1)<\infty.$$

Therefore, when $d_{ij}=1$, as $n\to\infty$,

$$U_{11}(n)\longrightarrow_P\sigma_1^2\sigma_{ij}\sum_{q_1=1}^{\infty}f_{i,q_1}(1)f_{j,q_1}(1),\tag{C.136}$$

$$U_{12}(n) \longrightarrow_P 0. \tag{C.137}$$

When $d_{ij} > 1$,

$$\begin{aligned}
\mathrm{E}[U_{11}(n)] =& \frac{\sigma_1^2}{n^{d_{ij}}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) \mathrm{E}\left[\eta_{i,t-q_1}\eta_{j,t-q_1}\right] \\
=& n^{d_{ij}-1}\sigma_1^2 \sigma_{ij} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) \longrightarrow 0,
\end{aligned} \tag{C.138}$$

and

$$\mathrm{E}[U_{12}(n)] = \frac{\sigma_1^2}{n^{d_{ij}}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} \sum_{\substack{q_2=1 \\ q_2 \neq q_1}}^{\infty} f_{i,q_1}(1) f_{j,q_2}(1) \mathrm{E}[\eta_{i,t-q_1}\eta_{j,t-q_2}] = 0. \tag{C.139}$$

Meanwhile,

$$\begin{aligned}
Var[U_{11}(n)] =& \frac{\sigma_1^4}{n^{2d_{ij}}} \mathrm{E}\left[\left(\sum_{t=1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) (\eta_{i,t-q_1}\eta_{j,t-q_1})\right)^2\right] \\
=& \frac{\sigma_1^4}{n^{2d_{ij}}} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \sum_{q_1=1}^{\infty} \sum_{q_2=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) f_{i,q_2}(1) f_{j,q_2}(1) \\
& \mathrm{E}\left[\eta_{i,t_1-q_1}\eta_{j,t_1-q_1}\eta_{i,t_2-q_2}\eta_{j,t_2-q_2}\right] \\
=& \frac{\sigma_1^4}{n^{2d_{ij}}} \sum_{t_1=1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1)^2 f_{j,q_1}(1)^2 \mathrm{E}\left[(\eta_{i,t_1-q_1}\eta_{j,t_1-q_1})^2\right] \\
& + \frac{2\sigma_1^4}{n^{2d_{ij}}} \sum_{t_1=1}^{n-1} \sum_{t_2=t_1+1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) f_{i,t_2-t_1+q_1}(1) f_{j,t_2-t_1+q_1}(1) \\
& \mathrm{E}\left[(\eta_{i,t_1-q_1}\eta_{j,t_1-q_1})^2\right] \\
=& \frac{\sigma_1^4 \delta_{ij}}{n^{2d_{ij}}} \sum_{t_1=1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1)^2 f_{j,q_1}(1)^2 \\
& + \frac{2\sigma_1^4 \delta_{ij}}{n^{2d_{ij}}} \sum_{t_1=1}^{n-1} \sum_{p=1}^{n-t_1} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) f_{i,p+q_1}(1) f_{j,p+q_1}(1) \\
=& O(n^{-1}),
\end{aligned} \tag{C.140}$$

given that $\sum_{q_1=1}^{\infty} f_{i,q_1}(1)^2 f_{j,q_1}(1)^2 < \infty$ and $\sum_{p=1}^{\infty}\sum_{q_1=1}^{\infty} |f_{i,q_1}(1) f_{j,q_1}(1) f_{i,p+q_1}(1) f_{j,p+q_1}(1)| < \infty$. Also,

$$Var[U_{12}(n)]$$

$$=\frac{\sigma_1^4}{n^{2d_{ij}}}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty}\sum_{l_1=1}^{\infty}\sum_{\substack{l_2=1\\l_2\neq l_1}}^{\infty} f_{i,q_1}(1)f_{j,q_2}(1)f_{i,l_1}(1)f_{j,l_2}(1)\mathrm{E}[\eta_{i,t_1-q_1}\eta_{j,t_1-q_2}\eta_{i,t_2-l_1}\eta_{j,t_2-l_2}]$$

$$=\frac{\sigma_1^4}{n^{2d_{ij}}}\sum_{t_1=1}^{n}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty} f_{i,q_1}(1)^2 f_{j,q_2}(1)^2\mathrm{E}[\eta_{i,t_1-q_1}^2\eta_{j,t_1-q_2}^2]$$

$$+\frac{2\sigma_1^4}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty} f_{i,q_1}(1)f_{j,q_2}(1)f_{i,t_2-t_1+q_1}(1)f_{j,t_2-t_1+q_2}(1)\mathrm{E}[\eta_{i,t_1-q_1}^2\eta_{j,t_1-q_2}^2]$$

$$=\frac{\sigma_1^4\sigma_{ii}\sigma_{jj}}{n^{2d_{ij}}}\sum_{t_1=1}^{n}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty} f_{i,q_1}(1)^2 f_{j,q_2}(1)^2$$

$$+\frac{2\sigma_1^4\sigma_{ii}\sigma_{jj}}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{p=1}^{n-t_1}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty} f_{i,q_1}(1)f_{j,q_2}(1)f_{i,p+q_1}(1)f_{j,p+q_2}(1)$$

$$=O(n^{-2d_{ij}+1}), \tag{C.141}$$

given that $\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty} f_{i,q_1}(1)^2 f_{j,q_2}(1)^2<\infty$ and

$$\sum_{p=1}^{\infty}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\q_2\neq q_1}}^{\infty} f_{i,q_1}(1)f_{j,q_2}(1)f_{i,p+q_1}(1)f_{j,p+q_2}(1)<\infty.$$

Therefore, when $d_{ij}>1$, as $n\to\infty$,

$$U_{11}(n)\longrightarrow_P 0, \tag{C.142}$$

$$U_{12}(n)\longrightarrow_P 0. \tag{C.143}$$

Hence, when $d_{ij}=1$,

$$\sum_{t=1}^{n}\mathrm{E}[M_{3,nt}^{i}M_{3,nt}^{j}|F_{t-1}]\longrightarrow_P \sigma_1^2\sigma_{ij}\sum_{q_1=1}^{\infty} f_{i,q_1}(1)f_{j,q_1}(1), \tag{C.144}$$

when $d_{ij}>1$,

$$\sum_{t=1}^{n}\mathrm{E}[M_{3,nt}^{i}M_{3,nt}^{j}|F_{t-1}]\longrightarrow_P 0. \tag{C.145}$$

(5) When $r_1=r_2=4$,

$$\sum_{t=1}^{n}\mathrm{E}[M_{4,nt}^{i}M_{4,nt}^{j}|F_{t-1}]=\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n}\mathrm{E}\left[\sum_{q_1=1}^{\infty} m_{i,q_1}(1)\epsilon_{t-q_1}\eta_{i,t}\sum_{q_2=1}^{\infty} m_{j,q_2}(1)\epsilon_{t-q_2}\eta_{j,t}|F_{t-1}\right]$$

$$
\begin{aligned}
&=\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n}\sum_{q_1=1}^{\infty}\sum_{q_2=1}^{\infty} m_{i,q_1}(1)m_{j,q_2}(1)\epsilon_{t-q_1}\epsilon_{t-q_2}\mathrm{E}\Big[\eta_{i,t}\eta_{j,t}\Big]\\
&=\frac{\sigma_{ij}}{n^{d_{ij}}}\sum_{t=1}^{n}\sum_{q_1=1}^{\infty}\sum_{q_2=1}^{\infty} m_{i,q_1}(1)m_{j,q_2}(1)\epsilon_{t-q_1}\epsilon_{t-q_2}\\
&=\frac{\sigma_{ij}}{n^{d_{ij}}}\sum_{t=1}^{n}\sum_{q_1=1}^{\infty} m_{i,q_1}(1)m_{j,q_1}(1)\epsilon_{t-q_1}^2+\frac{\sigma_{ij}}{n^{d_{ij}}}\sum_{t=1}^{n}\sum_{q_1=1}^{\infty}\sum_{\substack{q_2=1\\ q_2\neq q_1}}^{\infty} m_{i,q_1}(1)m_{j,q_2}(1)\epsilon_{t-q_1}\epsilon_{t-q_2}\\
&\triangleq U_{21}(n)+U_{22}(n). \qquad \text{(C.146)}
\end{aligned}
$$

It is easy to show that when $d_{ij}=1$,

$$\mathrm{E}[U_{21}(n)]=\sigma_{ij}\sigma_1^2\sum_{q_1=1}^{\infty} m_{i,q_1}(1)m_{j,q_1}(1), \qquad \text{(C.147)}$$

$$\mathrm{E}[U_{22}(n)]=0. \qquad \text{(C.148)}$$

Using the same method as in the previous case, we can show that

$$Var[U_{21}(n)]\longrightarrow 0, \qquad \text{(C.149)}$$

$$Var[U_{22}(n)]\longrightarrow 0. \qquad \text{(C.150)}$$

Therefore, when $d_{ij}=1$,

$$U_{21}(n)\longrightarrow_P \sigma_{ij}\sigma_1^2\sum_{q_1=1}^{\infty} m_{i,q_1}(1)m_{j,q_1}(1), \qquad \text{(C.151)}$$

$$U_{22}(n)\longrightarrow_P 0. \qquad \text{(C.152)}$$

When $d_{ij}>1$, we can show that

$$\mathrm{E}[U_{21}(n)]=0, \mathrm{E}[U_{22}(n)]=0, \qquad \text{(C.153)}$$

$$Var[U_{21}(n)]\longrightarrow 0, Var[U_{22}(n)]\longrightarrow 0. \qquad \text{(C.154)}$$

Therefore, when $d_{ij}>1$,

$$U_{21}(n)\longrightarrow_P 0, \qquad \text{(C.155)}$$

$$U_{22}(n)\longrightarrow_P 0. \qquad \text{(C.156)}$$

Hence, when $d_{ij} = 1$,

$$\sum_{t=1}^{n} \mathrm{E}[M^i_{4,nt} M^j_{4,nt} | F_{t-1}] \longrightarrow_P \sigma_{ij}\sigma_1^2 \sum_{q_1=1}^{\infty} m_{i,q_1}(1) m_{j,q_1}(1). \tag{C.157}$$

when $d_{ij} > 1$,

$$\sum_{t=1}^{n} \mathrm{E}[M^i_{4,nt} M^j_{4,nt} | F_{t-1}] \longrightarrow_P 0. \tag{C.158}$$

(6) When $r_1 = 1, r_2 = 3$ or $r_1 = 3, r_2 = 1$,

$$\begin{aligned}
\sum_{t=1}^{n} \mathrm{E}[M^i_{1,nt} M^j_{3,nt} | F_{t-1}] =& \frac{1}{n^{d_{ij}}} \sum_{t=1}^{n} \mathrm{E}\left[g_i(\tau_t)\Phi(1)\epsilon_t \sum_{q=1}^{\infty} f_{j,q}(1)\epsilon_t \eta_{j,t-q} | F_{t-1} \right] \\
=& \frac{\sigma_1^2 \Phi(1)}{n^{d_{ij}}} \sum_{t=1}^{n} g_i(\tau_t) \sum_{q=1}^{\infty} f_{j,q}(1)\eta_{j,t-q} \triangleq U_3(n).
\end{aligned} \tag{C.159}$$

Note that

$$\mathrm{E}[U_{3n}] = 0. \tag{C.160}$$

Meanwhile,

$$\begin{aligned}
\mathrm{E}[U_3(n)^2] =& \frac{\sigma_1^4 \Phi(1)^2}{n^{2d_{ij}}} \mathrm{E}\left[\left(\sum_{t=1}^{n} g_i(\tau_t) \sum_{q=1}^{\infty} f_{j,q}(1)\eta_{j,t-q} \right)^2 \right] \\
=& \frac{\sigma_1^4 \Phi(1)^2}{n^{2d_{ij}}} \sum_{t_1=1}^{n} \sum_{t_2=1}^{n} \sum_{q_1=1}^{\infty} \sum_{q_2=1}^{\infty} g_i(\tau_{t_1}) g_i(\tau_{t_2}) f_{j,q_1}(1) f_{j,q_2}(1) \mathrm{E}\left[\eta_{j,t_1-q_1} \eta_{j,t_2-q_2} \right] \\
=& \frac{\sigma_1^4 \Phi(1)^2}{n^{2d_{ij}}} \sum_{t_1=1}^{n} \sum_{q_1=1}^{\infty} g_i(\tau_{t_1})^2 f_{j,q_1}(1)^2 \mathrm{E}\left[\eta^2_{j,t_1-q_1} \right] \\
&+ \frac{2\sigma_1^4 \Phi(1)^2}{n^{2d_{ij}}} \sum_{t_1=1}^{n-1} \sum_{t_2=t_1+1}^{n} \sum_{q_1=1}^{\infty} g_i(\tau_{t_1}) g_i(\tau_{t_2}) f_{j,q_1}(1) f_{j,t_2-t_1+q_1}(1) \mathrm{E}\left[\eta^2_{j,t_1-q_1} \right] \\
=& \frac{\sigma_1^4 \Phi(1)^2 \sigma_{jj}}{n^{2d_{ij}}} \sum_{t_1=1}^{n} \sum_{q_1=1}^{\infty} g_i(\tau_{t_1})^2 f_{j,q_1}(1)^2 \\
&+ \frac{2\sigma_1^4 \Phi(1)^2 \sigma_{jj}}{n^{2d_{ij}}} \sum_{t_1=1}^{n-1} \sum_{t_2=t_1+1}^{n} \sum_{q_1=1}^{\infty} g_i(\tau_{t_1}) g_i(\tau_{t_2}) f_{j,q_1}(1) f_{j,t_2-t_1+q_1}(1),
\end{aligned} \tag{C.161}$$

where the first term is $O(n^{-d_j})$ given that $\sum_{q_1=1}^{\infty} f_{j,q_1}(1)^2 < \infty$.

For the second term

$$\frac{1}{n^{2d_{ij}-1}} \sum_{t_1=1}^{n-1} \sum_{t_2=t_1+1}^{n} \sum_{q_1=1}^{\infty} g_i(\tau_{t_1}) g_i(\tau_{t_2}) f_{j,q_1}(1) f_{j,t_2-t_1+q_1}(1)$$

$$=\frac{1}{n^{d_j-1}}\int_0^1\int_{\tau_1}^1\frac{g_i(\tau_1)g_i(\tau_2)}{n^{d_i-1}}\gamma_2(n(\tau_2-\tau_1),j)d\tau_1 d\tau_2\longrightarrow 0, \tag{C.162}$$

given that $\gamma_2(d,j)=\sum_{d=1}^{\infty}\sum_{q_1=1}^{\infty}f_{j,q_1}(1)f_{j,d+q_1}(1)$ and

$$\lim_{n\to\infty}\frac{1}{n^{d_j-1}}\int_0^1\int_{\tau_1}^1\frac{g_i(\tau_1)g_i(\tau_2)}{n^{d_i-1}}\gamma_2(n(\tau_2-\tau_1),j)d\tau_1 d\tau_2=0. \tag{C.163}$$

Therefore, as $n\to\infty$,

$$\sum_{t=1}^{n}\mathrm{E}[M_{1,nt}^{i}M_{3,nt}^{j}|F_{t-1}]\longrightarrow_P 0. \tag{C.164}$$

Similar result holds when $r_1=3$ and $r_2=1$.

(7) When $r_1=1, r_2=4$ or $r_1=4, r_2=1$,

$$\begin{aligned}\sum_{t=1}^{n}\mathrm{E}[M_{1,nt}^{i}M_{4,nt}^{j}|F_{t-1}]&=\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n}\mathrm{E}\left[g_i(\tau_t)\Phi(1)\epsilon_t\sum_{q=1}^{\infty}m_{j,q}(1)\epsilon_{t-q}\eta_{j,t}|F_{t-1}\right]\\&=\frac{\theta_j\Phi(1)}{n^{d_{ij}}}\sum_{t=1}^{n}g_i(\tau_t)\sum_{q=1}^{\infty}m_{j,q}(1)\epsilon_{t-q}\triangleq U_4(n).\end{aligned} \tag{C.165}$$

We can show that

$$\mathrm{E}[U_4(n)]=0. \tag{C.166}$$

Meanwhile,

$$\begin{aligned}\mathrm{E}[U_4(n)^2]=&\frac{\theta_i^2\Phi(1)^2}{n^{2d_{ij}}}\mathrm{E}\left[\left(\sum_{t=1}^{n}g_i(\tau_t)\sum_{q=1}^{\infty}m_{j,q}(1)\epsilon_{t-q}\right)^2\right]\\=&\frac{\theta_i^2\Phi(1)^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{q_1=1}^{\infty}\sum_{q_2=1}^{\infty}g_i(\tau_{t_1})g_i(\tau_{t_2})m_{j,q_1}(1)m_{j,q_2}(1)\mathrm{E}[\epsilon_{t_1-q_1}\epsilon_{t_2-q_2}]\\=&\frac{\theta_i^2\Phi(1)^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n}\sum_{q_1=1}^{\infty}g_i(\tau_{t_1})^2 m_{i,q_1}(1)^2\mathrm{E}[\epsilon_{t_1-q_1}^2]\\&+\frac{\theta_i^2\Phi(1)^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty}g_i(\tau_{t_1})g_i(\tau_{t_2})m_{j,q_1}(1)m_{j,t_2-t_1+q_1}(1)\mathrm{E}[\epsilon_{t_1-q_1}^2]\\=&\frac{\theta_i^2\Phi(1)^2\sigma_1^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n}\sum_{q_1=1}^{\infty}g_i(\tau_{t_1})^2 m_{j,q_1}(1)^2\\&+\frac{\theta_i^2\Phi(1)^2\sigma_1^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty}g_i(\tau_{t_1})g_i(\tau_{t_2})m_{j,q_1}(1)m_{j,t_2-t_1+q_1}(1),\end{aligned} \tag{C.167}$$

where the first term is $O(n^{-d_j})$ given that $\sum_{q_1=1}^{\infty} m_{j,q_1}(1)^2 < \infty$.

For the second term

$$\begin{aligned}
&\frac{1}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty} g_i(\tau_{t_1})g_i(\tau_{t_2})m_{j,q_1}(1)m_{j,t_2-t_1+q_1}(1)\\
=&\frac{1}{n^{d_j-1}}\int_0^1\int_{\tau_1}^1 \frac{g_i(\tau_1)g_i(\tau_2)}{n^{d_i-1}}\gamma_3(n(\tau_2-\tau_1),j)d\tau_1 d\tau_2 \longrightarrow 0,
\end{aligned} \tag{C.168}$$

where $\gamma_3(d,j) = \sum_{q_1=1}^{\infty} m_{j,q_1}(1)m_{j,d+q_1}(1)$ and

$$\lim_{n\to\infty}\frac{1}{n^{d_j-1}}\int_0^1\int_{\tau_1}^1 \frac{g_i(\tau_1)g_i(\tau_2)}{n^{d_i-1}}\gamma_3(n(\tau_2-\tau_1),j)d\tau_1 d\tau_2 = 0. \tag{C.169}$$

Therefore, as $n \to \infty$,

$$\sum_{t=1}^{n}\mathrm{E}[M_{1,nt}^{i}M_{4,nt}^{j}|F_{t-1}] \longrightarrow_P 0. \tag{C.170}$$

Similar result holds when $r_1 = 4$ and $r_2 = 1$.

(8) When $r_1 = 2, r_2 = 3$ or $r_1 = 3, r_2 = 2$,

$$\begin{aligned}
\sum_{t=1}^{n}\mathrm{E}[M_{2,nt}^{i}M_{3,nt}^{j}|F_{t-1}] =& \frac{1}{n^{d_{ij}}}\sum_{t=1}^{n}\mathrm{E}\left[f_{i,0}(1)\left(\epsilon_t\eta_{it}-\theta_i\right)\sum_{q=1}^{\infty} f_{j,q}(1)\epsilon_t\eta_{j,t-q}\middle|F_{t-1}\right]\\
=&\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n} f_{i,0}(1)\sum_{q=1}^{\infty} f_{j,q}(1)\eta_{j,t-q}\mathrm{E}\left[\left(\epsilon_t^2\eta_{it}\right)\right] = \frac{\delta_{2i}f_{i,0}(1)}{n^{d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty} f_{j,q}(1)\eta_{j,t-q} \triangleq U_5(n).
\end{aligned} \tag{C.171}$$

It is obvious that

$$\mathrm{E}[U_5(n)] = 0. \tag{C.172}$$

Meanwhile,

$$\begin{aligned}
\mathrm{E}[U_5(n)^2] =&\mathrm{E}\left[\left(\frac{\delta_{2i}f_{i,0}(1)}{n^{d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty} f_{j,q}(1)\eta_{j,t-q}\right)^2\right]\\
=&\frac{\delta_{2i}^2 f_{i,0}(1)^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n}\sum_{t_2=1}^{n}\sum_{q_1=1}^{\infty}\sum_{q_2=1}^{\infty} f_{j,q_1}(1)f_{j,q_2}(1)\mathrm{E}\left[\eta_{j,t_1-q_1}\eta_{j,t_2-q_2}\right]\\
=&\frac{\delta_{2i}^2 f_{i,0}(1)^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n}\sum_{q_1=1}^{\infty} f_{j,q_1}(1)^2\mathrm{E}\left[\eta_{j,t_1-q_1}^2\right]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\delta_{2i}^2 f_{i,0}(1)^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty} f_{j,q_1}(1) f_{j,t_2-t_1+q_1}(1)\mathrm{E}\left[\eta_{j,t_1-q_1}^2\right] \\
=&\frac{\delta_{2i}^2 f_{i,0}(1)^2\sigma_{jj}}{n^{2d_{ij}}}\sum_{t_1=1}^{n}\sum_{q_1=1}^{\infty} f_{j,q_1}(1)^2+\frac{\delta_{2i}^2 f_{i,0}(1)^2\sigma_{jj}}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty} f_{j,q_1}(1) f_{j,t_2-t_1+q_1}(1) \\
=&O(n^{-2d_{ij}+1}),
\end{aligned} \tag{C.173}
$$

given that $\sum_{q_1=1}^{\infty} f_{j,q_1}(1)^2<\infty$ and $\sum_{p=1}^{\infty}\sum_{q_1=1}^{\infty} f_{j,q_1}(1) f_{j,p+q_1}(1)<\infty$. Therefore,

$$
\sum_{t=1}^{n}\mathrm{E}[M_{2,nt}^i M_{3,nt}^j|F_{t-1}]\longrightarrow_P 0. \tag{C.174}
$$

Similar result holds when $r_1=3$ and $r_2=2$.

(9) When $r_1=2, r_2=4$ or $r_1=4, r_2=2$,

$$
\begin{aligned}
\sum_{t=1}^{n}\mathrm{E}[M_{2,nt}^i M_{4,nt}^j|F_{t-1}]=&\frac{1}{n^{d_{ij}}}\sum_{t=1}^{n}\mathrm{E}\left[f_{i,0}(1)\left(\epsilon_t\eta_{it}-\theta_i\right)\sum_{q=1}^{\infty} m_{j,q}(1)\epsilon_{t-q}\eta_{j,t}\middle|F_{t-1}\right] \\
=&\frac{f_{i,0}(1)}{n^{d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty} m_{j,q}(1)\epsilon_{t-q}\mathrm{E}[\epsilon_t\eta_{it}\eta_{j,t}] \\
=&\frac{f_{i,0}(1)\delta_{1ij}}{n^{d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty} m_{j,q}(1)\epsilon_{t-q}=U_6(n),
\end{aligned} \tag{C.175}
$$

where $\delta_{1ij}=\mathrm{E}[\epsilon_t\eta_{it}\eta_{j,t}]$. Similar as in the previous case, we can show that

$$
\mathrm{E}[U_6(n)]=0. \tag{C.176}
$$

Meanwhile,

$$
\begin{aligned}
\mathrm{E}\left[U_6(n)^2\right]=&\mathrm{E}\left[\left(\frac{f_{i,0}(1)\delta_{1ij}}{n^{d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty} m_{j,q}(1)\epsilon_{t-q}\right)^2\right] \\
=&\frac{f_{i,0}(1)^2\delta_{1ij}^2}{n^{2d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty} m_{j,q}(1)^2\mathrm{E}\left[\epsilon_{t-q}^2\right] \\
&+\frac{f_{i,0}(1)^2\delta_{1ij}^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q=1}^{\infty} m_{j,q}(1) m_{j,t_2-t_1+q}\mathrm{E}\left[\epsilon_{t_1-q}^2\right] \\
=&\frac{f_{i,0}(1)^2\delta_{1ij}^2\sigma_1^2}{n^{2d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty} m_{j,q}(1)^2+\frac{f_{i,0}(1)^2\delta_{1ij}^2\sigma_1^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{p=1}^{n-t_1}\sum_{q=1}^{\infty} m_{j,q}(1) m_{j,p+q}
\end{aligned}
$$

$$=O(n^{-2d_{ij}+1}), \tag{C.177}$$

given that $\sum_{q=1}^{\infty} m_{j,q}(1)^2 < \infty$ and $\sum_{p=1}^{\infty}\sum_{q=1}^{\infty} m_{j,q}(1)m_{j,p+q} < \infty$. Therefore, as $n \to \infty$,

$$\sum_{t=1}^{n} \mathrm{E}[M_{2,nt}^{i} M_{4,nt}^{j} | F_{t-1}] \longrightarrow_P 0. \tag{C.178}$$

Similar result holds when $r_1 = 4$ and $r_2 = 2$.

(10) When $r_1 = 3, r_2 = 4$ or $r_1 = 4, r_2 = 3$,

$$\begin{aligned}
\sum_{t=1}^{n} \mathrm{E}[M_{3,nt}^{i} M_{4,nt}^{j} | F_{t-1}] &= \frac{1}{n^{d_{ij}}} \sum_{t=1}^{n} \mathrm{E}\left[\sum_{q=1}^{\infty} f_{i,q}(1)\epsilon_t \eta_{i,t-q} \sum_{l=1}^{\infty} m_{j,l}(1)\epsilon_{t-l}\eta_{j,t} \middle| F_{t-1}\right] \\
&= \frac{1}{n^{d_{ij}}} \sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=1}^{\infty} f_{i,q}(1)\eta_{i,t-q} m_{j,l}(1)\epsilon_{t-l} \mathrm{E}\left[\epsilon_t \eta_{j,t}\right] \\
&= \frac{\theta_j}{n^{d_{ij}}} \sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=1}^{\infty} f_{i,q}(1) m_{j,l}(1)\eta_{i,t-q}\epsilon_{t-l} \\
&= \frac{\theta_j}{n^{d_{ij}}} \sum_{t=1}^{n}\sum_{q=1}^{\infty} f_{i,q}(1) m_{j,q}(1)\eta_{i,t-q}\epsilon_{t-q} + \frac{\theta_j}{n^{d_{ij}}} \sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1) m_{j,l}(1)\eta_{i,t-q}\epsilon_{t-l} \\
&\triangleq U_{71}(n) + U_{72}(n).
\end{aligned} \tag{C.179}$$

Then, when $d_{ij} = 1$,

$$\mathrm{E}[U_{71}(n)] = \mathrm{E}\left[\frac{\theta_j}{n} \sum_{t=1}^{n}\sum_{q=1}^{\infty} f_{i,q}(1) m_{j,q}(1)\eta_{i,t-q}\epsilon_{t-q}\right] = \theta_j \theta_i \sum_{q=1}^{\infty} f_{i,q}(1) m_{j,q}(1), \tag{C.180}$$

and

$$\mathrm{E}[U_{72}(n)] = \frac{\theta_j}{n} \sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1) m_{j,l}(1) \mathrm{E}\left[\eta_{i,t-q}\epsilon_{t-l}\right] = 0. \tag{C.181}$$

Meanwhile,

$$\begin{aligned}
\mathrm{E}\left[\left(U_{71}(n) - \mathrm{E}[U_{71}(n)]\right)^2\right] &= \mathrm{E}\left[\left(\frac{\theta_j}{n} \sum_{t=1}^{n}\sum_{q=1}^{\infty} f_{i,q}(1) m_{j,q}(1)(\eta_{i,t-q}\epsilon_{t-q} - \theta_i)\right)^2\right] \\
&= \frac{\theta_j^2}{n^2} \sum_{t=1}^{n}\sum_{q=1}^{\infty} f_{i,q}(1)^2 m_{j,q}(1)^2 \mathrm{E}\left[(\eta_{i,t-q}\epsilon_{t-q} - \theta_i)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&+\frac{\theta_j^2}{n^2}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty} f_{i,q_1}(1)m_{j,q_1}(1)f_{i,t_2-t_1+q_1}(1)m_{j,t_2-t_1+q_1}(1)\mathrm{E}\left[(\eta_{i,t_1-q_1}\epsilon_{t_1-q_1}-\theta_i)^2\right]\\
=&\frac{\theta_j^2(\delta_{22i}-\theta_i^2)}{n^2}\sum_{t=1}^{n}\sum_{q=1}^{\infty} f_{i,q}(1)^2 m_{j,q}(1)^2\\
&+\frac{\theta_j^2(\delta_{22i}-\theta_i^2)}{n^2}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty} f_{i,q_1}(1)m_{j,q_1}(1)f_{i,t_2-t_1+q_1}(1)m_{j,t_2-t_1+q_1}(1)\\
=&O(n^{-1}),
\end{aligned}
\tag{C.182}
$$

given that $\sum_{q=1}^{\infty} f_{i,q}(1)^2 m_{j,q}(1)^2 < \infty$ and

$$
\sum_{p=1}^{\infty}\sum_{q_1=1}^{\infty} f_{i,q_1}(1)m_{j,q_1}(1)f_{i,p+q_1}(1)m_{j,p+q_1}(1) < \infty.
$$

$$
\begin{aligned}
\mathrm{E}[U_{72}(n)^2] =&\mathrm{E}\left[\left(\frac{\theta_j}{n}\sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1)m_{j,l}(1)\eta_{i,t-q}\epsilon_{t-l}\right)^2\right]\\
=&\frac{\theta_j^2}{n^2}\sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1)^2 m_{j,l}(1)^2\mathrm{E}\left[\eta_{i,t-q}^2\epsilon_{t-l}^2\right]+\frac{\theta_j^2}{n^2}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty}\sum_{q_2=1}^{\infty}\\
&\sum_{l_1=q+1}^{\infty}\sum_{l_2=q_2+1}^{\infty} f_{i,q_1}(1)f_{i,q_2}(1)m_{j,l_1}(1)m_{j,l_2}(1)\mathrm{E}\left[\eta_{i,t_1-q_1}\eta_{i,t_2-q_2}\epsilon_{t_1-l_1}\epsilon_{t_2-l_2}\right]\\
=&\frac{\theta_j^2}{n^2}\sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1)^2 m_{j,l}(1)^2\mathrm{E}\left[\eta_{i,t-q}^2\epsilon_{t-l}^2\right]+\frac{\theta_j^2}{n^2}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty}\\
&\sum_{l_1=q+1}^{\infty} f_{i,q_1}(1)f_{i,t_2-t_1+q_1}(1)m_{j,l_1}(1)m_{j,t_2-t_1+l_1}(1)\mathrm{E}\left[\eta_{i,t_1-q_1}^2\epsilon_{t_1-l_1}^2\right]\\
=&\frac{\theta_j^2\delta_{22i}}{n^2}\sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1)^2 m_{j,l}(1)^2\\
&+\frac{\theta_j^2\delta_{22i}}{n^2}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty}\sum_{l_1=q+1}^{\infty} f_{i,q_1}(1)f_{i,t_2-t_1+q_1}(1)m_{j,l_1}(1)m_{j,t_2-t_1+l_1}(1)\\
=&O(n^{-1}),
\end{aligned}
\tag{C.183}
$$

given that $\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1)^2 m_{j,l}(1)^2 < \infty$ and

$$
\sum_{p=1}^{\infty}\sum_{q_1=1}^{\infty}\sum_{l_1=q+1}^{\infty} |f_{i,q_1}(1)f_{i,p+q_1}(1)m_{j,l_1}(1)m_{j,p+l_1}(1)| < \infty.
$$

Therefore, when $d_{ij} = 1$, as $n \to \infty$,

$$U_{71}(n) \longrightarrow_P \theta_j \theta_i \sum_{q=1}^{\infty} f_{i,q}(1) m_{j,q}(1), \tag{C.184}$$

$$U_{72}(n) \longrightarrow_P 0. \tag{C.185}$$

Hence,

$$\sum_{t=1}^{n} \mathrm{E}[M_{3,nt}^{i} M_{4,nt}^{j} | F_{t-1}] = U_{71}(n) + U_{72}(n) \longrightarrow_P \theta_j \theta_i \sum_{q=1}^{\infty} f_{i,q}(1) m_{j,q}(1). \tag{C.186}$$

When $d_{ij} > 1$,

$$\begin{aligned}\mathrm{E}[U_{71}(n)] =& \mathrm{E}\left[\frac{\theta_j}{n^{d_{ij}}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} f_{i,q}(1) m_{j,q}(1) \eta_{i,t-q} \epsilon_{t-q}\right] \\ =& \frac{\theta_j \theta_i}{n^{d_{ij}-1}} \sum_{q=1}^{\infty} f_{i,q}(1) m_{j,q}(1) \longrightarrow 0,\end{aligned} \tag{C.187}$$

and

$$\mathrm{E}[U_{72}(n)] = \frac{\theta_j}{n^{d_{ij}}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} \sum_{l=q+1}^{\infty} f_{i,q}(1) m_{j,l}(1) \mathrm{E}\left[\eta_{i,t-q} \epsilon_{t-l}\right] = 0. \tag{C.188}$$

Meanwhile,

$$\begin{aligned}&\mathrm{E}\left[U_{71}(n)^2\right] = \mathrm{E}\left[\left(\frac{\theta_j}{n^{d_{ij}}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} f_{i,q}(1) m_{j,q}(1) (\eta_{i,t-q} \epsilon_{t-q})\right)^2\right] \\ &= \frac{\theta_j^2}{n^{2d_{ij}}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} f_{i,q}(1)^2 m_{j,q}(1)^2 \mathrm{E}\left[(\eta_{i,t-q} \epsilon_{t-q})^2\right] \\ &\quad + \frac{\theta_j^2}{n^{2d_{ij}}} \sum_{t_1=1}^{n-1} \sum_{t_2=t_1+1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) m_{j,q_1}(1) f_{i,t_2-t_1+q_1}(1) m_{j,t_2-t_1+q_1}(1) \mathrm{E}\left[(\eta_{i,t_1-q_1} \epsilon_{t_1-q_1})^2\right] \\ &= \frac{\theta_j^2 \delta_{22i}}{n^{2d_{ij}}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} f_{i,q}(1)^2 m_{j,q}(1)^2 \\ &\quad + \frac{\theta_j^2 \delta_{22i}}{n^{2d_{ij}}} \sum_{t_1=1}^{n-1} \sum_{t_2=t_1+1}^{n} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) m_{j,q_1}(1) f_{i,t_2-t_1+q_1}(1) m_{j,t_2-t_1+q_1}(1) \\ &= O(n^{-2d_{ij}+1}),\end{aligned} \tag{C.189}$$

given that $\sum_{q=1}^{\infty} f_{i,q}(1)^2 m_{j,q}(1)^2 < \infty$ and

$$\sum_{p=1}^{\infty}\sum_{q_1=1}^{\infty} f_{i,q_1}(1)m_{j,q_1}(1)f_{i,p+q_1}(1)m_{j,p+q_1}(1) < \infty.$$

$$\begin{aligned}
\mathrm{E}[U_{72}(n)^2] =& \mathrm{E}\left[\left(\frac{\theta_j}{n^{d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1)m_{j,l}(1)\eta_{i,t-q}\epsilon_{t-l}\right)^2\right] \\
=& \frac{\theta_j^2}{n^{2d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1)^2 m_{j,l}(1)^2 \mathrm{E}\left[\eta_{i,t-q}^2\epsilon_{t-l}^2\right] + \frac{\theta_j^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty}\sum_{q_2=1}^{\infty} \\
& \sum_{l_1=q+1}^{\infty}\sum_{l_2=q_2+1}^{\infty} f_{i,q_1}(1)f_{i,q_2}(1)m_{j,l_1}(1)m_{j,l_2}(1)\mathrm{E}\left[\eta_{i,t_1-q_1}\eta_{i,t_2-q_2}\epsilon_{t_1-l_1}\epsilon_{t_2-l_2}\right] \\
=& \frac{\theta_j^2}{n^{2d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1)^2 m_{j,l}(1)^2 \mathrm{E}\left[\eta_{i,t-q}^2\epsilon_{t-l}^2\right] + \frac{\theta_j^2}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty} \\
& \sum_{l_1=q+1}^{\infty} f_{i,q_1}(1)f_{i,t_2-t_1+q_1}(1)m_{j,l_1}(1)m_{j,t_2-t_1+l_1}(1)\mathrm{E}\left[\eta_{i,t_1-q_1}^2\epsilon_{t_1-l_1}^2\right] \\
=& \frac{\theta_j^2\delta_{22i}}{n^{2d_{ij}}}\sum_{t=1}^{n}\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1)^2 m_{j,l}(1)^2 \\
& + \frac{\theta_j^2\delta_{22i}}{n^{2d_{ij}}}\sum_{t_1=1}^{n-1}\sum_{t_2=t_1+1}^{n}\sum_{q_1=1}^{\infty}\sum_{l_1=q+1}^{\infty} f_{i,q_1}(1)f_{i,t_2-t_1+q_1}(1)m_{j,l_1}(1)m_{j,t_2-t_1+l_1}(1) \\
=& O(n^{-2d_{ij}+1}), \qquad (C.190)
\end{aligned}$$

given that $\sum_{q=1}^{\infty}\sum_{l=q+1}^{\infty} f_{i,q}(1)^2 m_{j,l}(1)^2 < \infty$ and

$$\sum_{p=1}^{\infty}\sum_{q_1=1}^{\infty}\sum_{l_1=q+1}^{\infty} |f_{i,q_1}(1)f_{i,p+q_1}(1)m_{j,l_1}(1)m_{j,p+l_1}(1)| < \infty.$$

Therefore, when $d_{ij} = 1$, as $n \to \infty$,

$$U_{71}(n) \longrightarrow_P 0, \qquad (C.191)$$

$$U_{72}(n) \longrightarrow_P 0. \qquad (C.192)$$

Hence, when $d_{ij} = 1$,

$$\sum_{t=1}^{n}\mathrm{E}[M_{3,nt}^{i}M_{4,nt}^{j}|F_{t-1}] = U_{71}(n) + U_{72}(n) \longrightarrow_P \theta_j\theta_i\sum_{q=1}^{\infty} f_{i,q}(1)m_{j,q}(1). \qquad (C.193)$$

When $d_{ij} > 1$,

$$\sum_{t=1}^{n} \mathrm{E}[M^i_{3,nt} M^j_{4,nt} | F_{t-1}] = U_{71}(n) + U_{72}(n) \longrightarrow_P 0. \tag{C.194}$$

Similar result holds when $r_1 = 4$ and $r_2 = 3$.

To conclude,

$$\sum_{t=1}^{n} \mathrm{E}\left[(a' M_{nt})^2 | F_{t-1}\right] \longrightarrow_P a'\Omega a, \tag{C.195}$$

where Ω is a $K \times K$ variance-covariance matrix defined as follows.

For $1 \le i \le K_1$ and $1 \le j \le K_1$, i.e., $d_i = d_j = 1$,

$$\begin{aligned}
\Omega_{ij} =& \sigma_1^2 \Phi(1)^2 \boldsymbol{Q}_{ij} + f_{i,0}(1) f_{j,0}(1)(\delta_{2ij} - \theta_i\theta_j) + \Phi(1) f_{j,0}(1)\delta_{2j}\bar{g}_i + \Phi(1) f_{i,0}(1)\delta_{2i}\bar{g}_j \\
&+ \sigma_1^2 \sigma_{ij} \sum_{q_1=1}^{\infty} f_{i,q_1}(1) f_{j,q_1}(1) + \sigma_1^2 \sigma_{ij} \sum_{q_1=1}^{\infty} m_{i,q_1}(1) m_{j,q_1}(1) \\
&+ \theta_j \theta_i \sum_{q=1}^{\infty} f_{i,q}(1) m_{j,q}(1) + \theta_i \theta_j \sum_{q=1}^{\infty} f_{j,q}(1) m_{i,q}(1).
\end{aligned} \tag{C.196}$$

For $K_1 < i \le n$ and $1 \le j \le K_1$, i.e., $d_i > 1$, $d_j = 1$,

$$\Omega_{ij} = \sigma_1^2 \Phi(1)^2 \boldsymbol{Q}_{ij} + \Phi(1) f_{j,0}(1) \delta_{2j} \bar{g}_i. \tag{C.197}$$

Finally, when $K_1 < i \le K$ and $K_1 < j \le K$, i.e., $d_i > 1$, $d_j > 1$,

$$\Omega_{ij} = \sigma_1^2 \Phi(1)^2 \boldsymbol{Q}_{ij}. \tag{C.198}$$

We then examine the second condition of the CLT for martingale difference sequence that as $n \to \infty$,

$$\sum_{t=1}^{n} \mathrm{E}\left[(a' M_{nt})^4 \middle| F_{t-1}\right] \longrightarrow_P 0. \tag{C.199}$$

It is then equivalent to show that for any i,

$$\sum_{t=1}^{n} \mathrm{E}\left[a_i^4 (M^i_{nt})^4 \middle| F_{t-1}\right] \longrightarrow_P 0. \tag{C.200}$$

Then, it is equivlent to prove

$$\sum_{t=1}^{n} \mathrm{E}\left[(M^i_{p,nt})^4 \middle| F_{t-1}\right] \longrightarrow_P 0. \tag{C.201}$$

for $p = 1, 2, 3, 4$, respectively.

When $p = 1$,

$$\begin{aligned}\sum_{t=1}^{n} \mathrm{E}\left[(M_{1,nt}^{i})^4 | F_{t-1}\right] =& \frac{1}{n^{2d_i}} \sum_{t=1}^{n} g_i(\tau_t)^4 \Phi(1)^4 \mathrm{E}[\epsilon_t^4] \\ =& \frac{C\Phi(1)^4}{n^{2d_i}} \sum_{t=1}^{n} g_i(\tau_t)^4, \end{aligned} \tag{C.202}$$

which is purely deterministic. Note that

$$\frac{1}{n} \sum_{t=1}^{n} \left(n^{-\frac{d_i-1}{2}} g_i(\tau_t) \right)^4 \longrightarrow \int_0^1 g_i^N(\tau)^4 d\tau < \infty,$$

where $g_i^N(\tau) = n^{-\frac{d_i-1}{2}} g_i(\tau_t)$ is the rescaled trend function. Therefore, given that $\mathrm{E}[\epsilon_t^4] < \infty$, we have

$$\sum_{t=1}^{n} \mathrm{E}\left[(M_{1,nt}^{i})^4 | F_{t-1}\right] = \frac{1}{n} \left(C\Phi(1)^4 \int_0^1 g_i^N(\tau) d\tau \right) \longrightarrow 0, \tag{C.203}$$

as $n \to \infty$. When $p = 2$,

$$\begin{aligned}\sum_{t=1}^{n} \mathrm{E}\left[(M_{2,nt}^{i})^4 | F_{t-1}\right] =& \frac{1}{n^{2d_i}} \sum_{t=1}^{n} f_{i,0}(1)^4 \mathrm{E}\left[\left(\epsilon_t \eta_{it} - \sigma_{12} \right)^4 \right] \\ =& \frac{(\delta_{44i} - \sigma_{12}^4)}{n^{2d_i}} \sum_{t=1}^{n} f_{i,0}(1)^4 \\ =& \frac{(\delta_{44i} - \sigma_{12}^4) f_{i,0}(1)^4}{n^{2d_i - 1}} \end{aligned} \tag{C.204}$$

where $\delta_{44i} = \mathrm{E}[\epsilon_t^4 \eta_{it}^4] < \infty$. Since it is also purely deterministic and $2d_i - 1 > 0$, therefore, as $n \to \infty$

$$\sum_{t=1}^{n} \mathrm{E}\left[(M_{2,nt}^{i})^4 | F_{t-1}\right] \longrightarrow 0. \tag{C.205}$$

When $p = 3$,

$$\begin{aligned}\sum_{t=1}^{n} \mathrm{E}\left[(M_{3,nt}^{i})^4 | F_{t-1}\right] =& \frac{1}{n^{2d_i}} \sum_{t=1}^{n} \left(\sum_{q=1}^{\infty} f_{i,q}(1) \eta_{i,t-q} \right)^4 \mathrm{E}[\epsilon_t^4] \\ =& \frac{C}{n^{2d_i}} \sum_{t=1}^{n} \left(\sum_{q=1}^{\infty} f_{i,q}(1) \eta_{i,t-q} \right)^4 \triangleq C \cdot U_8(n). \end{aligned} \tag{C.206}$$

Note that

$$\begin{aligned}
\mathrm{E}[U_8(n)] =& \frac{1}{n^{2d_i}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} \sum_{q_2=1}^{\infty} \sum_{q_3=1}^{\infty} \sum_{q_4=1}^{\infty} f_{i,q_1}(1) f_{i,q_2}(1) f_{i,q_3}(1) f_{i,q_4}(1) \mathrm{E}[\eta_{i,t-q_1}\eta_{i,t-q_2}\eta_{i,t-q_3}\eta_{i,t-q_4}] \\
=& \frac{1}{n^{2d_i}} \sum_{t=1}^{n} \sum_{q=1}^{\infty} f_{i,q}(1)^4 \mathrm{E}\left[\eta_{i,t-q}^4\right] + \frac{1}{n^{2d_i}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} \sum_{q_2=q_1+1}^{\infty} f_{i,q_1}(1)^2 f_{i,q_2}(1)^2 \mathrm{E}\left[\eta_{i,t-q_1}^2 \eta_{i,t-q_2}^2\right] \\
=& \frac{C}{n^{2d_i-1}} \sum_{k=1}^{\infty} f_{i,q}(1)^4 + \frac{C}{n^{2d_i-1}} \sum_{q_1=1}^{\infty} \sum_{q_2=q_1+1}^{\infty} f_{q_1}(1)^2 f_{q_2}(1)^2 \\
=& O(n^{-2d_i+1}) = o(1), \qquad \text{(C.207)}
\end{aligned}$$

since $2d_i - 1 > 0$ and given that $\mathrm{E}\left[\eta_{i,t}^4\right] < \infty$, $\mathrm{E}\left[\eta_{i,t-q_1}^2 \eta_{i,t-q_2}^2\right] < \infty$ and $\sum_{q=1}^{\infty} f_{i,q}(1)^4 < \infty$, $\sum_{q_1=1}^{\infty} \sum_{q_2=q_1+1}^{\infty} f_{i,q_1}(1)^2 f_{i,q_2}(1)^2 < \infty$.

Since $U_8(n) \geq 0$, then as $n \to \infty$, $\mathrm{E}[U_8(n)] \longrightarrow 0$ implies

$$U_8(n) \longrightarrow_P 0. \qquad \text{(C.208)}$$

Using the same method as above, we can show that when $p = 4$,

$$\begin{aligned}
\sum_{t=1}^{n} \mathrm{E}\left[(M_{4,nt}^i)^4 | F_{t-1}\right] =& \frac{1}{n^{2d_i}} \sum_{t=1}^{n} \left(\sum_{q=1}^{\infty} m_{i,q}(1) \epsilon_{t-q} \right)^4 \mathrm{E}[\eta_{it}^4] \\
=& \frac{C}{n^{2d_i}} \sum_{t=1}^{n} \left(\sum_{q=1}^{\infty} m_{i,q}(1) \epsilon_{t-q} \right)^4 \triangleq C \cdot U_9(n). \qquad \text{(C.209)}
\end{aligned}$$

Note that

$$\begin{aligned}
\mathrm{E}[U_9(n)] =& \frac{1}{n^{2d_i}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} \sum_{q_2=1}^{\infty} \sum_{q_3=1}^{\infty} \sum_{q_4=1}^{\infty} m_{i,q_1}(1) m_{i,q_2}(1) m_{i,q_3}(1) m_{i,q_4}(1) \mathrm{E}[\epsilon_{t-q_1}\epsilon_{t-q_2}\epsilon_{t-q_3}\epsilon_{t-q_4}] \\
=& \frac{1}{n^{2d_i}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} m_{i,q_1}(1)^4 \mathrm{E}[\epsilon_{t-q_1}^4] + \frac{1}{n^{2d_i}} \sum_{t=1}^{n} \sum_{q_1=1}^{\infty} \sum_{q_2=q_1+1}^{\infty} m_{i,q_1}(1)^2 m_{i,q_2}(1)^2 \mathrm{E}[\epsilon_{t-q_1}^2 \epsilon_{t-q_2}^2] \\
=& \frac{C}{n^{2d_i-1}} \sum_{q_1=1}^{\infty} m_{i,q_1}(1)^4 + \frac{C}{n^{2d_i-1}} \sum_{q_1=1}^{\infty} \sum_{q_2=q_1+1}^{\infty} m_{i,q_1}(1)^2 m_{i,q_2}(1)^2 \\
=& O(n^{-2d_i+1}) = o(1), \qquad \text{(C.210)}
\end{aligned}$$

given that $\mathrm{E}[\epsilon_{t-q_1}^4] < \infty$ and $\mathrm{E}[\epsilon_{t-q_1}^2 \epsilon_{t-q_2}^2] < \infty$. Meanwhile, $\sum_{q=1}^{\infty} m_{i,q}(1)^4 < \infty$ and $\sum_{q_1=1}^{\infty} \sum_{q_2=q_1+1}^{\infty} m_{i,q_1}(1)^2 m_{i,q_2}(1)^2 < \infty$. Hence, as $U_9(n) \geq 0$, $\mathrm{E}[U_9(n)] \longrightarrow 0$ implies

$$U_9(n) \longrightarrow_P 0. \qquad \text{(C.211)}$$

as $n \to \infty$. Therefore, (C.200) holds. We then complete the proof for Lemma B.2.5. ■

Appendix D

Verification of Assumption 2.3.4

It is sufficient to show that the stationary sequence $\{v_t\}$ satisfy

$$\sum_{\substack{t_1,t_2,\dots,t_p=1\\ t_1\neq t_2\neq\dots\neq t_p}}^{n}\left|f_{t_1,t_2,\dots,t_p}(z_1,z_2,\dots,z_p)-\prod_{i=1}^{p}f(z_i)\right|=O(n^{p-1}),\tag{D.1}$$

for $p=2,3,\dots,6$. In addition to the mixing conditions, this assumption describes the asymptotic independence of the mixing sequence in terms of the joint density and the marginal density. Without loss of generality, suppose $p=2$, and the sequence $\{v_t\}$ follows AR(1) process as

$$v_t=\rho v_{t-1}+\epsilon_t,\tag{D.2}$$

where $0<\rho<1$, and $\epsilon_t \overset{i.i.d.}{\sim} N(0,1-\rho^2)$. Therefore, the marginal distribution of $\{v_t\}$ is the standard normal distribution. Meanwhile, the joint density of v_t and v_s is

$$f_{t,s}(x,y)=\frac{1}{2\pi\sqrt{1-\rho^{2j}}}\exp\left(-\frac{x^2+y^2-2\rho^j xy}{2(1-\rho^{2j})}\right).\tag{D.3}$$

Therefore, our objective is to show that

$$\sum_{t=1}^{n}\sum_{\substack{s=1\\ s\neq t}}^{n}\left|f_{t,s}(x,y)-f(x)f(y)\right|=O(n).\tag{D.4}$$

Let $j=|s-t|$, based on stationarity, the joint density only depend on j. Hence we can write $f_j(x,y)\triangleq f_{t,s}(x,y)$. Let $A=(x^2+y^2)/2>0$, $B=(x^2+y^2-2\rho^j xy)/2$, note that

$$|f_j(x,y)-f(x)f(y)|=\left|\frac{1}{2\pi\sqrt{1-\rho^{2j}}}\exp\left(-\frac{x^2+y^2-2\rho^j xy}{2(1-\rho^{2j})}\right)-\frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)\right|$$

$$\leq \left| \frac{1}{2\pi\sqrt{1-\rho^{2j}}} \exp\left(-\frac{x^2+y^2-2\rho^j xy}{2(1-\rho^{2j})}\right) - \frac{1}{2\pi\sqrt{1-\rho^{2j}}} \exp\left(-\frac{x^2+y^2}{2}\right) \right|$$
$$+ \left| \frac{1}{2\pi\sqrt{1-\rho^{2j}}} \exp\left(-\frac{x^2+y^2}{2}\right) - \frac{1}{2\pi} \exp\left(-\frac{x^2+y^2}{2}\right) \right| \triangleq F_1(j) + F_2(j), \tag{D.5}$$

where

$$F_1(j) = \left| \frac{1}{2\pi\sqrt{1-\rho^{2j}}} \exp\left(-\frac{B}{1-\rho^{2j}}\right) - \frac{1}{2\pi\sqrt{1-\rho^{2j}}} \exp(-A) \right|$$
$$= \left| \frac{1}{2\pi\sqrt{1-\rho^{2j}}} \right| \left| \exp\left(-\frac{B}{1-\rho^{2j}}\right) - \exp(-A) \right|$$
$$\leq \left| \frac{1}{2\pi} \right| \left| \exp\left(-\frac{B}{1-\rho^{2j}}\right) - \exp\left(-\frac{A}{1-\rho^{2j}}\right) + \exp\left(-\frac{A}{1-\rho^{2j}}\right) - \exp(-A) \right|$$
$$\leq \left| \frac{1}{2\pi} \right| \left| \exp\left(-\frac{B}{1-\rho^{2j}}\right) - \exp\left(-\frac{A}{1-\rho^{2j}}\right) \right| + \left| \frac{1}{2\pi} \right| \left| \exp\left(-\frac{A}{1-\rho^{2j}}\right) - \exp(-A) \right|$$
$$\triangleq F_{11}(j) + F_{12}(j). \tag{D.6}$$

Notice that

$$\sum_{j=1}^{\infty} |f_j(x,y) - f(x)f(y)| \leq \sum_{j=1}^{\infty} \left|F_2(j)\right| + \sum_{j=1}^{\infty} \left|F_{11}(j)\right| + \sum_{j=1}^{\infty} \left|F_{12}(j)\right| \tag{D.7}$$

Further,

$$\sum_{j=1}^{\infty} \left|F_{11}(j)\right| = \sum_{j=1}^{\infty} \left| \frac{1}{2\pi} \right| \left| \exp\left(-\frac{B}{1-\rho^{2j}}\right) - \exp\left(-\frac{A}{1-\rho^{2j}}\right) \right|$$
$$= \left| \frac{1}{2\pi} \right| \sum_{j=1}^{\infty} \left| \exp\left(-\frac{A}{1-\rho^{2j}}\right) \left(\exp\left(\frac{\rho^j xy}{(1-\rho^{2j})}\right) - 1 \right) \right|$$
$$\leq \left| \frac{1}{2\pi} \exp(-A) \right| \sum_{j=1}^{\infty} \left| \left(\exp\left(\frac{\rho^j xy}{(1-\rho^{2j})}\right) - 1 \right) \right|$$
$$= \left| \frac{\exp(-A)}{2\pi} \right| \sum_{j=1}^{\infty} \left| \sum_{k=1}^{\infty} \frac{(xy)^k}{k!} \left(\frac{\rho^j}{1-\rho^{2j}} \right)^k \right|$$
$$\leq \left| \frac{\exp(-A)}{2\pi} \right| \sum_{k=1}^{\infty} \frac{(|xy|)^k}{k!} \sum_{j=1}^{\infty} \left(\frac{\rho^j}{1-\rho^{2j}} \right)^k. \tag{D.8}$$

Also note that

$$\begin{aligned}
\sum_{j=1}^{\infty}\left|F_{12}(j)\right| &= \sum_{j=1}^{\infty}\left|\frac{1}{2\pi}\right|\left|\exp\left(-\frac{A}{1-\rho^{2j}}\right)-\exp(-A)\right| \\
&= \sum_{j=1}^{\infty}\left|\frac{\exp(-A)}{2\pi}\right|\left|\exp\left(-\frac{A}{(1-\rho^{2j})}+A\right)-1\right| \\
&= \frac{\exp(-A)}{2\pi}\sum_{j=1}^{\infty}\left|\exp\left(-\frac{\rho^{2j}A}{(1-\rho^{2j})}\right)-1\right| \\
&= \frac{\exp(-A)}{2\pi}\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}\frac{(-1)^k A^k}{k!}\left(\frac{\rho^{2j}}{1-\rho^{2j}}\right)^k\right| \\
&\le \frac{\exp(-A)}{2\pi}\sum_{k=1}^{\infty}\frac{A^k}{k!}\sum_{j=1}^{\infty}\left(\frac{\rho^{2j}}{1-\rho^{2j}}\right)^k. \qquad \text{(D.9)}
\end{aligned}$$

Meanwhile,

$$\begin{aligned}
&\sum_{j=1}^{\infty}\left|F_2(j)\right| = \sum_{j=1}^{\infty}\left|\frac{1}{2\pi\sqrt{1-\rho^{2j}}}\exp(-A)-\frac{1}{2\pi}\exp(-A)\right| \\
&= \sum_{j=1}^{\infty}\frac{\exp(-A)}{2\pi}\left|\frac{1}{\sqrt{1-\rho^{2j}}}-1\right| = \sum_{j=1}^{\infty}\frac{\exp(-A)}{2\pi}\left|\sum_{k=1}^{\infty}\frac{(-1)^k(2k-1)!!\rho^{2jk}}{2^k k!}\right| \\
&= \frac{\exp(-A)}{2\pi}\sum_{k=1}^{\infty}\left|\frac{(-1)^k(2k-1)!!}{2^k k!}\right|\sum_{j=1}^{\infty}\rho^{2jk} \le \frac{\exp(-A)}{2\pi}\sum_{k=1}^{\infty}\left|\frac{(-1)^k(2k-1)!!}{2^k k!}\right|\left|\frac{\rho^{2k}}{1-\rho^{2k}}\right| \\
&= \frac{\exp(-A)}{2\pi}\sum_{k=1}^{\infty}\left|\frac{(2k-1)}{2k}\times\frac{2k-3}{2(k-1)}\times\cdots\times\frac{5}{6}\times\frac{3}{4}\times\frac{1}{2}\right|\left|\frac{\rho^{2k}}{1-\rho^{2k}}\right| \\
&\le \frac{\exp(-A)}{2\pi}\sum_{k=1}^{\infty}\left|\frac{\rho^{2k}}{1-\rho^{2k}}\right|. \qquad \text{(D.10)}
\end{aligned}$$

Define[1] $k_1 = \left[\frac{\ln(1/2)}{2\ln\rho}\right]+1$, for $0<\rho<1$. Therefore, we have $1-\rho^{2j} > 1/2$ when $j > k_1$. Note that

$$\begin{aligned}
&\left|\sum_{j=1}^{\infty}F_{11}(j)\right| \le \left|\frac{\exp(-A)}{2\pi}\right|\sum_{k=1}^{\infty}\frac{(|xy|)^k}{k!}\sum_{j=1}^{\infty}\left(\frac{\rho^j}{1-\rho^{2j}}\right)^k \\
&\le \left|\frac{\exp(-A)}{2\pi}\right|\sum_{k=1}^{\infty}\frac{(|xy|)^k}{k!}\sum_{j=1}^{k_1}\left(\frac{\rho^j}{1-\rho^{2j}}\right)^k + \left|\frac{\exp(-A)}{2\pi}\right|\sum_{k=1}^{\infty}\frac{(|xy|)^k}{k!}\sum_{j=k_1+1}^{\infty}\left(\frac{\rho^j}{1-\rho^{2j}}\right)^k
\end{aligned}$$

[1] $[x]$ denotes the integer part of x.

$$
\begin{aligned}
&\leq \left|\frac{\exp(-A)}{2\pi}\right| \sum_{k=1}^{\infty} \frac{(|xy|)^k}{k!} C + \left|\frac{\exp(-A)}{2\pi}\right| \sum_{k=1}^{\infty} \frac{(|xy|)^k}{k!} \sum_{j=k_1+1}^{\infty} \left(\frac{\rho^j}{1/2}\right)^k \\
&= \left|\frac{\exp(-A)C}{2\pi}\right| \sum_{k=1}^{\infty} \frac{(|xy|)^k}{k!} + \left|\frac{\exp(-A)}{2\pi}\right| \sum_{k=1}^{\infty} \frac{(2|xy|)^k}{k!} \frac{\rho^{k_1+1}}{1-\rho^k} \\
&= \left|\frac{\exp(-A)C}{2\pi}\right| \Big(\exp(|xy|)-1\Big) + \left|\frac{\exp(-A)}{2\pi}\right| \rho^{k_1+1} \left(\sum_{k=1}^{k_1} \frac{(2|xy|)^k}{k!(1-\rho^k)} + \sum_{k=k_1+1}^{\infty} \frac{(2|xy|)^k}{k!(1-\rho^k)}\right) \\
&\leq \left|\frac{\exp(-A)C}{2\pi}\right| \Big(\exp(|xy|)-1\Big) + \left|\frac{\exp(-A)}{2\pi}\right| \rho^{k_1+1} C + \left|\frac{\exp(-A)}{2\pi}\right| \rho^{k_1+1} \sum_{k=k_1+1}^{\infty} \frac{(2|xy|)^k}{k!(1/2)} \\
&\leq \left|\frac{\exp(-A)C}{2\pi}\right| \Big(\exp(|xy|)-1\Big) + \left|\frac{\exp(-A)}{2\pi}\right| \rho^{k_1+1} C + \left|\frac{\exp(-A)}{\pi}\right| \rho^{k_1+1} \Big(\exp(2|xy|)-1\Big) < \infty.
\end{aligned}
\tag{D.11}
$$

Following the same method, it can be easily shown that $\sum_{j=1}^{\infty} \left|F_{12}(j)\right| < \infty$, and $\sum_{j=1}^{\infty} \left|F_2(j)\right| < \infty$. Therefore,

$$
\sum_{j=1}^{\infty} |f_j(x,y) - f(x)f(y)| < \infty, \tag{D.12}
$$

and it suffices to show that

$$
\begin{aligned}
\sum_{t=1}^{n} \sum_{\substack{s=1 \\ s\neq t}}^{n} |f_{t,s}(x,y) - f(x)f(y)| &= \sum_{t=2}^{n-1} \sum_{s=1}^{t-1} |f_j(x,y) - f(x)f(y)| + \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} |f_j(x,y) - f(x)f(y)| \\
= \sum_{t=1}^{n-1} \sum_{j=1}^{t-1} |f_j(x,y) - f(x)f(y)| &+ \sum_{t=1}^{n-1} \sum_{j=1}^{n-t} |f_j(x,y) - f(x)f(y)| = O(n).
\end{aligned}
\tag{D.13}
$$

This result can be generalized to the cases when there are more than two variables. i.e.,

$$
\sum_{\substack{t_1,t_2,\ldots,t_p=1 \\ t_1\neq t_2\neq\ldots\neq t_p}}^{n} \left| f_{t_1,t_2,\ldots,t_p}(z_1,z_2,\ldots,z_p) - \prod_{i=1}^{p} f(z_i) \right| = O(n^{p-1}). \tag{D.14}
$$

Appendix E

Proofs of Theorems in Appendix A

For simplicity, we only prove the Theorems when the error terms are i.i.d. innovations. The case of mixing innovations can be proved with more complicated mathematical techniques.

Proof of Theorem A.3.1: The trend function is estimated by local constant method, i.e.,

$$\widehat{g}(\tau) = \frac{1}{nh\widehat{f}(\tau)} \sum_{s=1}^{n} K\left(\frac{\tau_s - \tau}{h}\right) x_s. \tag{E.1}$$

We can write the above equation as

$$\begin{aligned} \widehat{f}(\tau)(\widehat{g}(\tau) - g(\tau)) =& \frac{1}{nh} \sum_{s=1}^{n} K\left(\frac{\tau_s - \tau}{h}\right)(g(\tau_s) - g(\tau)) + \frac{1}{nh} \sum_{s=1}^{n} K\left(\frac{\tau_s - \tau}{h}\right) v_s \\ \triangleq& GN_1(n) + GN_2(n). \end{aligned} \tag{E.2}$$

We show that $GN_1(n)$ forms the bias term in the estimation, while $GN_2(n)$ forms the variance term. By Taylor Expansion, $g(\tau_s) - g(\tau) = g'(\tau)(\tau_s - \tau) + g''(\tau)(\tau_s - \tau)^2/2 + o(\tau_s - \tau)^2$. Therefore, the leading term of $GN_1(n)$ is

$$\begin{aligned} \widetilde{GN}_1(n) =& \frac{1}{nh} \sum_{s=1}^{n} K\left(\frac{\tau_s - \tau}{h}\right)(g(\tau_s) - g(\tau)) \\ =& \frac{1}{nh} \sum_{s=1}^{n} K\left(\frac{\tau_s - \tau}{h}\right)(g'(\tau)(\tau_s - \tau) + g''(\tau)(\tau_s - \tau)^2/2 \\ =& \frac{1}{nh} \sum_{s=1}^{n} K\left(\frac{\tau_s - \tau}{h}\right) g'(\tau)(\tau_s - \tau) + \frac{1}{2nh} \sum_{s=1}^{n} K\left(\frac{\tau_s - \tau}{h}\right) g''(\tau)(\tau_s - \tau)^2 \end{aligned}$$

$$=\frac{g'(\tau)}{n}\sum_{s=1}^{n}K\left(\frac{\tau_s-\tau}{h}\right)\left(\frac{\tau_s-\tau}{h}\right)+\frac{hg''(\tau)}{2n}\sum_{s=1}^{n}K\left(\frac{\tau_s-\tau}{h}\right)\left(\frac{\tau_s-\tau}{h}\right)^2, \tag{E.3}$$

where the first term converges to $\int uK(u)du = 0$ and the second term converges to $h^2g''(\tau)\int u^2K(u)du/2$. Therefore, as $n\to\infty$,

$$\widetilde{GN}_1(n) \longrightarrow h^2g''(\tau)\int u^2K(u)du/2. \tag{E.4}$$

The variance of $GN_1(n)$ is 0 because this term is deterministic.

We then look at the second term $GN_2(n)$. It is obvious that the expectation is 0. The variance of $GN_2(n)$ is

$$\begin{aligned}Var(GN_2(n)) =& Var\left[\frac{1}{nh}\sum_{s=1}^{n}K\left(\frac{\tau_s-\tau}{h}\right)v_s\right] = \frac{1}{n^2h^2}\sum_{s=1}^{n}K^2\left(\frac{\tau_s-\tau}{h}\right)\Omega_v\\ =& \frac{\Omega_v}{n^2h^2}\sum_{s=1}^{n}K^2\left(\frac{\tau_s-\tau}{h}\right)\to\frac{\Omega_v}{nh}\int K^2(u)du.\end{aligned} \tag{E.5}$$

Since $g(\cdot)$ is defined at fixed designed points, we have $\widehat{f}(\tau) = 1 + o(1)$. To summarize, by Liapunov's CLT, we have as $n\to\infty$,

$$\sqrt{nh}\left(\widehat{g}(\tau)-g(\tau)-\frac{h^2}{2}g''(\tau)\mu_2\right)\longrightarrow_D \mathcal{N}(0,\Omega_v\kappa_2), \tag{E.6}$$

where $\mu_2 = \int u^2K(u)du$, and $\kappa_2 = \int K^2(u)du$.

Proofs of Theorem A.3.2: We first show that

$$\frac{1}{n}\sum_{t=1}^{n}\pi(v_t)x_t' \longrightarrow_P \Sigma_v. \tag{E.7}$$

Note that

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\pi(v_t)x_t' =& \frac{1}{n}\sum_{t=1}^{n}\pi(v_t)g(\tau_t)' + \frac{1}{n}\sum_{t=1}^{n}\pi(v_t)v_t'\\ =& \frac{1}{n}\sum_{t=1}^{n}\Big(\pi(v_t)-\mathrm{E}[\pi(v_t)]\Big)g(\tau_t)' + \frac{1}{n}\sum_{t=1}^{n}\mathrm{E}[\pi(v_t)]g(\tau_t)' + \frac{1}{n}\sum_{t=1}^{n}\pi(v_t)v_t'\\ \triangleq& PN_1(n)+PN_2(n)+PN_3(n).\end{aligned} \tag{E.8}$$

By Law of Large Numbers, we can show that

$$PN_3(n) \to_P \mathrm{E}[\pi(v_t)v_t'], \tag{E.9}$$

as $n \to \infty$. Also, we have

$$PN_2(n) \to \mathrm{E}[\pi(v_1)]\bigg(\int g(\tau)d\tau\bigg). \tag{E.10}$$

For the first term $PN_1(n)$, we have $\mathrm{E}[PN_1(n)] = 0$ and [1]

$$\begin{aligned}
\mathrm{E}\Big[PN_1(n)^2\Big] =& \mathrm{E}\left[\frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\Big(\pi(v_t)-\mathrm{E}[\pi(v_1)]\Big)\Big(\pi(v_s)-\mathrm{E}[\pi(v_1)]\Big)g(\tau_t)g(\tau_s)\right] \\
=& \mathrm{E}\left[\frac{1}{n^2}\sum_{t=1}^{n}\Big(\pi(v_t)-\mathrm{E}[\pi(v_1)]\Big)^2 g(\tau_t)^2\right] \\
& + \mathrm{E}\left[\frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1,s\neq t}^{n}\Big(\pi(v_t)-\mathrm{E}[\pi(v_1)]\Big)\Big(\pi(v_s)-\mathrm{E}[\pi(v_1)]\Big)g(\tau_t)g(\tau_s)\right] \\
=& \frac{C_v}{n^2}\sum_{t=1}^{n} g(\tau_t)^2 \longrightarrow \frac{C_v}{n}\int_0^1 g(\tau)^2 d\tau = O(n^{-1}),
\end{aligned} \tag{E.11}$$

where $C_v = \mathrm{E}[(\pi(v_1)-\mathrm{E}[\pi(v_1)])^2] < \infty$. To summarize, as $n \to \infty$,

$$\frac{1}{n}\sum_{t=1}^{n}\pi(v_t)x_t \longrightarrow_P \mathrm{E}[\pi(v_1)]\int_0^1 g(\tau)d\tau + \mathrm{E}[\pi(v_1)v_1] = \Sigma_v, \tag{E.12}$$

where Σ_v is defined in Assumption A.1.2.

Next, we consider $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\pi(v_t)e_t$. By Central Limit Theorem, it is easy to show that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\pi(v_t)e_t \longrightarrow_D \mathcal{N}(0,\, \Gamma). \tag{E.13}$$

where Γ is the variance-covariance matrix of $\xi_t = \pi(v_t)e_t$. Therefore, by Slutsky's Theorem, we have

$$\sqrt{n}(\widehat{\beta}-\beta) \longrightarrow_D \mathcal{N}(0, \Sigma_v^{-1}\Gamma\Sigma_v^{-1}). \tag{E.14}$$

Thus, we complete the proof of this Theorem. ■

[1] We only consider the univariate case just for simplicity.

Bibliography

Amihud, Y. and Hurvich, C. M. (2004), 'Predictive regressions: A reduced-bias estimation method', *Journal of Financial and Quantitative Analysis* **39**(4), 813–841.

Andrews, D. W. and McDermott, C. J. (1995), 'Nonlinear econometric models with deterministically trending variables', *The Review of Economic Studies* **62**(3), 343–360.

Bernanke, B. (1985), 'Adjustment costs, durables, and aggregate consumption', *Journal of Monetary Economics* **15**(1), 41–68.

Bierens, H. J. (1997), 'Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the us price level and interest rate', *Journal of Econometrics* **81**(1), 29–64.

Cai, Z. (2007), 'Trending time-varying coefficient time series models with serially correlated errors', *Journal of Econometrics* **136**(1), 163–188.

Cai, Z. and Wang, Y. (2014), 'Testing predictive regression models with nonstationary regressors', *Journal of Econometrics* **178**(1), 4–14.

Campbell, J. Y. (1987), 'Does saving anticipate declining labor income? An alternative test of the permanent income hypothesis', *Econometrica* **55**(6), 1249–1273.

Campbell, J. Y. and Mankiw, N. G. (1990), 'Permanent income, current income, and consumption', *Journal of Business and Economic Statistics* **8**(3), 265–279.

Campbell, J. Y. and Yogo, M. (2006), 'Efficient tests of stock return predictability', *Journal of Financial Economics* **81**(1), 27–60.

Canova, F. (1998), 'Detrending and business cycle facts', *Journal of Monetary Economics* **41**(3), 475–512.

Chen, J., Gao, J. and Li, D. (2012), 'Semiparametric trending panel data models with cross-sectional dependence', *Journal of Econometrics* **171**(1), 71–85.

Cochrane, J. H. (2012), A brief parable of over-differencing, Working paper, University of Chicago, http://faculty.chicagobooth.edu/john.cochrane/research/papers /overdifferencing.pdf.

DeJong, D. N., Nankervis, J. C., Savin, N. E. and Whiteman, C. H. (1992), 'The power problems of unit root test in time series with autoregressive errors', *Journal of Econometrics* **53**(1), 323–343.

DeJong, D. N. and Whiteman, C. H. (1991), 'Reconsidering 'trends and random walks in macroeconomic time series'', *Journal of Monetary Economics* **28**(2), 221–254.

Dickey, D. A. and Fuller, W. A. (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association* **74**(366a), 427–431.

Diebold, F. X. and Rudebusch, G. D. (1991), 'On the power of dickey-fuller tests against fractional alternatives', *Economics Letters* **35**(2), 155–160.

Dong, C. and Gao, J. (2014), 'Specification testing in structural nonparametric cointegration', Working paper, Monash University. Available at SSRN: https://ssrn.com/abstract=2379928.

Engle, R. F. and Granger, C. W. (1987), 'Co-integration and error correction: Representation, estimation, and testing', *Econometrica* **55**(2), 251–276.

Fan, J. and Gijbels, I. (1996), *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, Vol. 66, CRC Press.

Fan, J. and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York.

Fernández-Villaverde, J. and Krueger, D. (2007), 'Consumption over the life cycle: Facts from consumer expenditure survey data', *The Review of Economics and Statistics* **89**(3), 552–565.

Flavin, M. A. (1981), 'The adjustment of consumption to changing expectations about future income', *The Journal of Political Economy* (5), 974–1009.

Flavin, M. A. (1984), 'Excess sensitivity of consumption to current income: liquidity constraints or myopia?', Working paper, NBER working paper No. 1341.

Fuller, W. A. (1976), *Introduction to statistical time series*, John Wiley & Sons, New York.

Gao, J. (2007), *Nonlinear Time Series: Semiparametric and Nonparametric Methods*, Chapman & Hall, London.

Gao, J. and Hawthorne, K. (2006), 'Semiparametric estimation and testing of the trend of temperature series', *The Econometrics Journal* **9**(2), 332–355.

Gao, J. and Phillips, P. C. B. (2013), 'Functional coefficient nonstationary regression', Cowles fundation discussion paper 1911, Cowles Fundation for Research in Economics, Yale University, https://ideas.repec.org/p/cwl/cwldpp/1911.html#cites.

Granger, C. W. (1981), 'Some properties of time series data and their use in econometric model specification', *Journal of Econometrics* **16**(1), 121–130.

Hall, R. E. (1978), 'Stochastic implications of the life cycle-permanent income hypothesis: Theory and evidence', *The Journal of Political Economy* **86**(6), 971–987.

Han, H. and Ogaki, M. (1997), 'Consumption, income and cointegration', *International Review of Economics and Finance* **6**(2), 107–117.

Hansen, B. E. (1992), 'Tests for parameter instability in regressions with I(1) processes', *Journal of Business and Economic Statistics* **10**(3), 321–35.

Hansen, B. E. and Phillips, P. C. (1990), 'Estimation and inference in models of cointegration: A simulation study', *Advances in Econometrics* **8**, 225–248.

Härdle, W., Liang, H. and Gao, J. (2000), *Partially Linear Models*, Physica-Verlag, Heidelberg.

Härdle, W. and Vieu, P. (1992), 'Kernel regression smoothing of time series', *Journal of Time Series Analysis* **13**(3), 209–232.

Harvey, A. (1997), 'Trends, cycles and autoregressions', *The Economic Journal* **107**(440), 192–201.

Hodrick, R. J. and Prescott, E. C. (1997), 'Postwar U.S. business cycles: an empirical investigation', *Journal of Money, Credit, and Banking* **29**(1), 1–16.

King, R. G., Plosser, C. I., Stock, J. H. and Watson, M. W. (1991), 'Stochastic trends and economic fluctuations', *The American Economic Review* **81**(4), 819–840.

Krugman, P. R. (1995), *The Age of Diminished Expectations*, MIT Press, Cambridge.

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

Li, Q. and Racine, J. S. (2007), *Nonparametric Econometrics: Theory and Practice*, Princeton University Press, New Jersey.

Liang, Z. and Li, Q. (2012), 'Functional coefficient regression models with time trend', *Journal of Econometrics* **170**(1), 15–31.

Mankiw, N. G. and Shapiro, M. D. (1985), 'Trends, random walks, and tests of the permanent income hypothesis', *Journal of Monetary Economics* **16**(2), 165–174.

Mankiw, N. G. and Shapiro, M. D. (1986), 'Do we reject too often?: Small sample properties of tests of rational expectations models', *Economics Letters* **20**(2), 139–145.

Nelson, C. R. and Kang, H. (1981), 'Spurious periodicity in inappropriately detrended time series', *Econometrica* **49**(3), 741–751.

Nelson, C. R. and Kang, H. (1984), 'Pitfalls in the use of time as an explanatory variable in regression', *Journal of Business and Economic Statistics* **2**(1), 73–82.

Nelson, C. R. and Plosser, C. R. (1982), 'Trends and random walks in macroeconomic time series: Some evidence and implications', *Journal of Monetary Economics* **10**(2), 139–162.

Newey, W. K. and West, K. D. (1987), 'A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix', *Econometrica* **55**(3), 703–708.

Park, J. Y. and Hahn, S. B. (1999), 'Cointegrating regressions with time varying coefficients', *Econometric Theory* **15**(5), 664–703.

Perron, P. (1988), 'Trends and random walks in macroeconomic time series: Further evidence from a new approach', *Journal of Economic Dynamics and Control* **12**(2-3), 297–332.

Perron, P. (1989), 'The great crash, the oil price shock, and the unit root hypothesis', *Econometrica* **57**(7), 1361–1401.

Phillips, P. C. B. (2001), 'Trending time series and macroeconomic activity: Some present and future challenges', *Journal of Econometrics* **100**(1), 21–27.

Phillips, P. C. B. (2003), 'Laws and limits of econometrics', *The Economic Journal* **113**(486), C26–C52.

Phillips, P. C. B. (2005), 'Challenges of trending time series econometrics', *Mathematics and Computers in Simulation* **68**(5), 401–416.

Phillips, P. C. B. (2010), 'The mysteries of trend', *Macroeconomic Review* **9**(2), 82–89.

Phillips, P. C. B. and Hansen, B. E. (1990), 'Statistical inference in instrumental variables regression with I(1) processes', *The Review of Economic Studies* **57**(1), 99–125.

Phillips, P. C. B., Li, D. and Gao, J. (2017), 'Estimating smooth structural change in cointegration models', *Journal of Econometrics* **196**(1), 180–195.

Phillips, P. C. B. and Perron, P. (1988), 'Testing for a unit root in time series regression', *Biometrika* **75**(2), 335–346.

Phillips, P. C., Shi, S. and Yu, J. (2015), 'Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500', *International Economic Review* **56**(4), 1043–1078.

Phillips, P. C., Wu, Y. and Yu, J. (2011), 'Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?', *International Economic Review* **52**(1), 201–226.

Quintos, C. E. and Phillips, P. C. B. (1993), 'Parameter constancy in cointegrating regressions', *Empirical Economics* **18**(4), 675–706.

Rao, B. B. (2010), 'Deterministic and stochastic trends in the time series models: A guide for the applied economist', *Applied Economics* **42**(17), 2193–2202.

Robinson, P. M. (1988), 'Root-N-consistent semiparametric regression', *Econometrica* **56**(4), 931–954.

Robinson, P. M. (2012), 'Nonparametric trending regression with cross-sectional dependence', *Journal of Econometrics* **169**(1), 4–14.

Rudebusch, G. D. (1993), 'The uncertain unit root in real GNP', *The American Economic Review* **83**(1), 264–272.

Said, S. E. and Dickey, D. A. (1984), 'Testing for unit roots in autoregressive-moving average models of unknown order', *Biometrika* **71**(3), 599–607.

Stambaugh, R. F. (1999), 'Predictive regressions', *Journal of Financial Economics* **54**(3), 375–421.

Wang, Q. and Phillips, P. C. B. (2009), 'Structural nonparametric cointegrating regression', *Econometrica* **77**(6), 1901–1948.

Welch, I. and Goyal, A. (2008), 'A comprehensive look at the empirical performance of equity premium prediction', *Review of Financial Studies* **21**(4), 1455–1508.

White, H. and Granger, C. W. (2011), 'Consideration of trends in time series', *Journal of Time Series Econometrics* **3**(1), article 2.

Xiao, Z. (2009a), 'Functional-coefficient cointegration models', *Journal of Econometrics* **152**(2), 81–92.

Xiao, Z. (2009b), 'Quantile cointegrating regression', *Journal of Econometrics* **150**(2), 248–260.

Yatchew, A. (2003), *Semiparametric Regression for the Applied Econometrician*, Cambridge University Press, Cambridge.