UNIFYING ESTIMATION OF VARYING-COEFFICIENT MODELS

by

Weitong Yin

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Applied Mathematics

Charlotte

2019

Approved by:

____

Dr. Jiancheng Jiang

____

Dr. Qingning Zhou

____

Dr. Weihua Zhou

____

Dr. Weidong Tian

ABSTRACT

WEITONG YIN. Unifying Estimation of Varying-coefficient Models. (Under the direction of DR. JIANCHENG JIANG)

Varying-coefficient models are widely used to analyze the relationship between a response and a group of covariates. Existing research shows different convergence rates for the estimators of coefficient for the stationary part and the nonstationary part. It brings difficulties in statistical inference for the coefficient functions since an appropriate sampling distribution has to be carefully chosen. In this dissertation we propose a unifying two-step estimation procedure for varying-coefficients models, which facilitates the unifying inference for coefficients. In step one, a local smoother (LS) is adopted to give estimates of coefficients for the stationary part. In step two, we propose a weighted local score equation(WLSE) method for estimating the nonstationary part coefficients. The proposed two-step procedure will provide a unifying estimation procedure for the varying-coefficients models. The asymptotic joint distribution of the proposed estimators is established, which provides a Wald type of confidence regions for the coefficient functions. However, this confidence region does not work well when the conditional variance of the error changes. To solve this problem, we propose an empirical likelihood inference tool for the coefficient functions. Simulations demonstrate good finite sample performance of our estimators and coverage probability of proposed empirical likelihood confidence regions. A real example illustrates the value of our methodology.

KEY WORDS: Two-Step Estimation, Local Smoother(LS), Weighted Local Score Equation(WLSE), Asymptotic Normality, Empirical Likelihood

DEDICATION

Firstly, I would like to give thanks to my advisor Dr. Jiancheng Jiang. His intelligence and humor make every single discussion between us enlightening and enjoyable. He navigated me through all the major difficulties in my dissertation research so I could accomplish it. The way Dr. Jiang behaved made as unforgettable an impression on me as did his ideas.

Beyond his academic support, Dr. Jiang showered me with sincerity, encouragement and trust. He simply extended a one-way street of helpfulness. It was ordained at birth that he would be excellent, he elected to be kind. I'm being taught not only how to study wisely, but also being taught how to live wisely. All in all, I am very thankful and grateful for all the things he has done for me.

Moreover, I sincerely thank other committee members , Dr.Weidong Tian, Dr.Weihua Zhou and Dr.Qingning Zhou, for their time and insightful questions. Also, I'd like to acknowledge the financial support of math department for my PhD study.

Meanwhile, I would also express my deepest gratitude to all my friends and family members. Their everlasting love, support and unwavering faith in me give me the confidence and strength during my PhD study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1 Motivation

Varying coefficient regression models provide a very useful tool for analyzing the relationship between a response and predictors. They have similar structure and interpretability to the traditional linear regression model. Because of the infinite dimensionality of the corresponding parameter spaces, they are more flexible (Park and Mammen, 2015). For these advantages, they have been embraced by many applied researchers in statistics, economics and finance. In varying coefficient models, the coefficients are set to be functions of some other predictors. Varying coefficient models inherit the interpretability of the classical linear models while being nonparametric. They originated from real application. See Hastie and Tibshirani(1993), Fan and Zhang (2008) for details. Cai and Li (2009) studied the varying-coefficient models for time series data, which allows for stationary and nonstationary covariates. This model is specified as

$$y_t = x_{t-1}^\top \beta(z_t) + \theta(z_t) + v_t, \quad x_t = (x_{t,1}, \dots, x_{t,k})^\top,$$

which is equivalent to

$$y_t = W_{t-1}^\top \gamma(z_t) + v_t, \tag{1.1}$$

where $W_{t-1} = (x_{t-1}^\top, 1)^\top$ and $\gamma(z_t) = (\beta(z_t)^\top, \theta(z_t))^\top$. The covariates $x_{t,i}$'s are I(0), I(1) and NI(1) variables, z_t is stationary and v_t is a white noise independent of x_t and z_t at all leads and lags. Model (1.1) includes the well-known varying coefficient models widely studied in the literature. See, for example, Jiang and Mack (2001), Fan and Yao (2003), Cai, Li and Park (2009), and Xiao (2009a). Since model (1.1) allows the coefficient vector $\gamma(z_t)$ to depend on z_t, the "Curse of Dimensionality" will be avoided for low dimensional smoothing in z_t and the modeling bias in linear models will be reduced significantly.

In the recent years, there have been growing interests and activities in the study and application of varying coefficient models. Examples include Fan and Zhang (1999), Cai, Fan and Li (2000), Park and Hahn (1999), Robinson (1989, 1991), Cai (2007), Chen and Hong (2007), Fan and Zhang (2008), Huang, Wu and Zhou (2002), Fan, Yao and Cai (2003), Fan, Zhang (2000), Hastie and Tibshirani (1993) and Cai and Li (2009). In particular, model (1.1) with stationary x_{t-1} and z_t has been considered by Chen and Tsay (1993), Hastie and Tibshirani (1993) and Cai, Fan and Li (2000). Model (1.1) with stationary v_t, stationary x_{t-1} and $z_t = t$ has been tackled by Robinson (1989, 1991), Cai (2007) and Chen and Hong (2007). Park and Hahn (1999) and Chang and Martinez- Chombo (2003) studied model (1.1) with stationary v_t, nonstationary x_{t-1} and $z_t = t$, and Cai and Wang (2008) considered model (1.1) with nearly integrated x_{t-1}.

Cai and Li (2009) studied model (1.1) and shows different convergence rates for the estimation of coefficients for the stationary part and the nonstationary part. The difference in the limiting distributions of $\gamma(z)$ across different persistency level makes inference difficult since an appropriate sampling distribution has to be carefully chosen. In addition, when x_{t-1} is I(1) or NI(1), the limiting distribution usually depends on nuisance local parameters that cannot be consistently estimated, which makes the statistical inference even harder. For this reason, it motivates us to propose a unifying inference procedures that are robust to different levels of persistency.

To solve the difficulty above and to work with multiple predictors simultaneously, in this dissertation, we propose a two-step unifying inference tool for the proposed model and derive the asymptotic distributions of the estimators in both "stationary" and "nonstationary" cases. This approach leads to a unifying limiting distribution of the proposed estimators with the predictors being stationary, I(0), I(1), NI(1) or slightly explosive. The asymptotic joint distribution of the proposed estimators provides a Wald type of confidence regions for the coefficient functions. However, this confidence region does not work well when the conditional variance of the error changes. To solve this problem, we propose an empirical likelihood inference tool for the coefficient functions. Simulations demonstrate good finite sample performance of our estimators and accurate coverage probability of proposed empirical likelihood confidence region.

1.2 Outline of the Dissertation

The rest of this dissertation is organized as follows. In Chapter 2, the unifying two-step estimation procedure is proposed for the varying-coefficients models where an algorithm for grouping stationary and nonstationary variables is introduced. In Chapter 3, we investigate the asymptotic properties of proposed estimators. In Chapter 4, the empirical likelihood region is proposed for the parameter functions. In Chapter 5, we conduct simulations to evaluate the accuracy of our estimation. In Chapter 6, a real example is used to illustrate the performance of our proposed estimation procedure. Concluding remarks are presented in Chapter 7. Proofs of the main results are given in the Appendix.

CHAPTER 2: A UNIFYING TWO-STEP ESTIMATION PROCEDURE

Before proposing our estimation procedure, we classify the covariates into the "stationary" and nonstationary" groups. Specifically, let

$$\mathcal{I} = \{i : max_{1\leq t\leq n} n^{-1/2} log(n)|W_{t-1,i}| < C^*, i = 1, \ldots, k+1\},$$

and $\mathcal{I}^c = \{1, 2, \ldots, k+1\} \setminus \mathcal{I}$. The variables with indexes in $\mathcal{I}$ are classified into the stationary group, and those with indexes in $\mathcal{I}^c$ in the nonstationary group. If $W_{t-1,i}$ is stationary, it belongs to $\mathcal{I}$ with probability going to one, since

$$P(max_{1\leq t-1\leq n} n^{-1/2} log(n)|W_{t-1,i}| < C^*) \longrightarrow 1,$$

as $n \to \infty$. If $W_{t-1,i}$ is $I(1)$ or $NI(1)$, then with probability going to one, it belongs to $\mathcal{I}^c$.

Let $W_{t-1,\mathcal{I}}$ and $W_{t-1,\mathcal{I}^c}$ be the subvectors of W_{t-1} with indices in $\mathcal{I}$ and $\mathcal{I}^c$, respectively. Then model (1.1) can be rewritten as

$$y_t = W_{t-1,\mathcal{I}}^T \gamma_{\mathcal{I}}(z_t) + W_{t-1,\mathcal{I}^c}^T \gamma_{\mathcal{I}^c}(z_t) + v_t, \tag{2.1}$$

where $\gamma_{\mathcal{I}}$ and $\gamma_{\mathcal{I}^c}$ are coefficients of $W_{t-1,\mathcal{I}}$ and $W_{t-1,\mathcal{I}^c}$ respectively.

Without loss of generality, we write model (1.1) as

$$y_t = W_{t-1,1}^T \gamma_1(z_t) + W_{t-1,2}^T \gamma_2(z_t) + v_t, \tag{2.2}$$

where $W_{t-1,1}$ consists of all stationary or $I(0)$ variables and $W_{t-1,2}$ are formed by all NI(1) and I(1) variables. With a proper choice of C^*, model (2.1) will be equivalent to model (2.2) with probability going to one, since $P(max_{1\leq t-1\leq n} n^{-1/2} log(n)|W_{t-1,i}| < C^*) \longrightarrow 1$ as $n \to \infty$.

2.1 Step one - Local Smoother (LS)

For model (2.1), Cai and Li (2009) proposed a two-step estimation procedure. I employ its step one as the step one in our unifying two-step estimation procedure.

Step one - Local Smoother (LS)

We estimate $\gamma(z)$ by the local smoother. By condition (A.1), we assume $\gamma(z)$ is twice continuously differentiable in z for all $z \in R$. For any given z, we can do a local linear approximation $\gamma(z)+\gamma^{(1)}(z)(z_t-z)$ to approximate $\gamma(z_t)$, where $\gamma^{(q)} = d^q\gamma(z)/dz^q$.

By minimizing the following loss function,

$$(\hat{\theta}_0, \hat{\theta}_1)^T = argmin_{\theta_0,\theta_1} \sum_{t=1}^{n} [y_t - W_{t-1}^T \theta_0 - (z_t - z) W_{t-1}^T \theta_1]^2 K_{h_1}(z_t - z), \qquad (2.3)$$

where $K_{h_1}(u) = h^{-1}K(u/h_1)$. By condition (A.2), $K(.)$ is a kernel function which is a symmetric and continous density function, supported by$[-1, 1]$, with h_1 being a bandwidth used to control the amount of data in smoothing. Here we choose the rule-of-thumb bandwidth, $h_1 = c_1 n^{-1/5}$, where c_1 is a constant. $\hat{\theta}_0 = \hat{\gamma}(z)$ estimates $\gamma(z)$, and $\hat{\theta}_1 = \hat{\gamma}^{(1)}(z)$ estimates $\gamma^{(1)}(z)$.

Then, $(\hat{\gamma}(z), \hat{\gamma}^{(1)}(z))^T$ can be expressed as

$$
\begin{aligned}
(\hat{\gamma}(z), \hat{\gamma}^{(1)}(z))^T = & \Big[\sum_{t=1}^{n} \begin{pmatrix} W_{t-1} \\ (z_t - z)W_{t-1} \end{pmatrix}^{\otimes 2} K_{h_1}(z_t - z)\Big]^{-1} \\
& \times \Big[\sum_{t=1}^{n} \begin{pmatrix} W_{t-1} \\ (z_t - z)W_{t-1} \end{pmatrix} y_t K_{h_1}(z_t - z)\Big],
\end{aligned}
\tag{2.4}
$$

where $A^{\otimes 2} = AA^T$ for a vector or matrix A, $\otimes$ is the Kronecker product.

As discussed in section 2.3 of Cai and Li(2009), the difference in the limiting distributions of $\hat{\gamma}(z)$ across different persistency level makes the inference for $\gamma(z)$ difficult. For this reason, an appropriate limiting distribution has to be chosen to use. This motivates us to propose a unifying two-step estimation procedure.

2.2 Step two - Weighted Local Score Equation(WLSE) Method

For model (2.1), we have step-one estimation of $\gamma(z)$ under local smoother. Let $\hat{\gamma}_{\mathcal{I}}$ and $\hat{\gamma}_{\mathcal{I}^c}$ be the subvectors of $\hat{\gamma}(z)$ with indices in $\mathcal{I}$ and $\mathcal{I}^c$ respectively. To establish a unifying inference framework, a step-two estimation procedure is proposed based on **Weighted Local Score Equation(WLSE)** method. That is, we replace $\gamma_{\mathcal{I}}(z_t)$ by $\hat{\gamma}_{\mathcal{I}}(z_t)$ in model (2.1)to obtain the partial residual y_t^*:

$$
y_t^* = y_t - W_{t-1,\mathcal{I}}^T \hat{\gamma}_{\mathcal{I}}(z_t) = W_{t-1,\mathcal{I}^c}^T \gamma_{\mathcal{I}^c}(z_t) + v_t^*, \tag{2.5}
$$

where z_t is stationary, v_t is a white noise independent of $W_{t-1,1}$, $W_{t-1,2}$ at all leads and lags, $v_t^* = v_t - W_{t-1,\mathcal{I}}^T[\hat{\gamma}_{\mathcal{I}}(z_t) - \gamma_{\mathcal{I}}(z_t)]$.

In the Step 2, we regress the partial residual y_t^* on $W_{t-1,\mathcal{I}^c}$. Since $W_{t-1,\mathcal{I}^c}$ is a group of "nonstationary" variables, direct least square estimation will lead to a mixing-normal distribution with the convergence rate being n. To get a joint asymptotic distribution, we propose a Weighted Local Score Equation(WLSE) method to estimate $\gamma_{\mathcal{I}^c}(z)$.

Step-two - Weighted Local Score Equation(WLSE) Method:

$$\sum_{t=1}^{n} Q_t^*[y_t^* - W_{t-1,\mathcal{I}^c}^T\{\xi + \eta(z_t - z)\}]K_{h_2}(z_t - z) = 0, \tag{2.6}$$

where $Q_t^* = (1, z_t - z)^\top \otimes (\Omega_t^* W_{t-1,\mathcal{I}^c})$ with the $\Omega_t^* = diag\{w_{t,1}, \ldots, \omega_{t,d_2}\}$, $w_{t,i} = (1 + ||W_{t-1,\mathcal{I}^c}||^2)^{-1/2}$

The resulting estimation of $(\gamma_{\mathcal{I}^c}(z), \gamma_{\mathcal{I}^c}^{(1)}(z))$ admits the closed form:

$$(\hat{\gamma}_{\mathcal{I}^c}(z), \hat{\gamma}_{\mathcal{I}^c}^{(1)}(z))^T = A_n^{-1}B_n, \tag{2.7}$$

where $A_n = \sum_{t=1}^{n}\{(1, z_t - z)^\top(1, z_t - z)\}\otimes(\Omega_t^* W_{t-1,\mathcal{I}^c}^{\otimes 2})K_{h_2}(z_t - z)$ and $B_n = \sum_{t=1}^{n} Q_t^* y_t^* K_{h_2}(z_t - z)$. We choose rule-of-thumb bandwidth, $h_2 = 1.06Sn^{-1/3}$, where S is the sample standard deviation of z_t.

2.3 Algorithm for C^* Selection

We have introduced the following index $\mathcal{I}$ to regroup the variables.

$$\begin{cases} i \in \mathcal{I} \quad if \quad max_{1\leq t\leq n} n^{-1/2} log(n)|W_{t-1,i}| < C^* \\ i \notin \mathcal{I} \quad if \quad max_{1\leq t\leq n} n^{-1/2} log(n)|W_{t-1,i}| \geq C^* \end{cases}$$

where C^* is a positive tuning parameter chosen to maximize the efficiency of our proposed estimator, $\hat{\gamma}(z)$. Since the proposed algorithm involves the positive tuning parameter C^*, different values of C^* may lead to different index $\mathcal{I}$, which naturally causes different estimations. While adopting larger C^*, more covariates will be regrouped as "stationary", on the other side of the coin, more nonstionary covariates will be missed as well. Conversely, with smaller C^*, more stationary covariates will be missed. So there is an obvious trade-off between them. For a better performance of our proposed estimator, a proper C^* has to be carefully chosen.

In any model building process, it has always been a challenging task to choose a tuning parameter. Optimal tuning parameters are "difficult to calibrate in practice" (Lederer and Müller, 2015) and are "not practically feasible" (Fan and Tang, 2013). Since specific techniques have their opponents and proponents, the task becomes even more difficult. We use cross-validation to choose the tuning parameter, C^*. Cross-validation is a commonly used technique to evaluate predictive models by dividing the sample data into a validation set and a training set to check model performance (Efron and Tibshirani, 1995). Tibshirani calls cross-validation "A simple, intuitive way to estimate prediction error". By comparing prediction errors across different splittings, we will be able to evaluate the overall model performance on a given data set. Grid search builds a model for every possible combination of parameters specified, then evaluates each model. Here, we combine the K-fold cross-validation and grid search technique to find an "optimal" C^*.

Algorithm: K-fold Cross-validation and Grid Search for C^* Selection

Set $K = 10$, Split the data into K folds; $\{E_1, E_2, \ldots, E_K\} \leftarrow \mathbf{0}$

Set $m = 20$, $Grid = \{c_1, c_2, \ldots, c_m\}$; $\{CV_1, CV_2, \ldots, CV_m\} \leftarrow \mathbf{0}$

for $i = 1 : m$ **do**

  for $k = 1 : K$ **do**

    Fit the model with $C^* = c_i$ to the other $K - 1$ part, giving $\hat{\gamma}^{-k}(z_t)$;

    Compute its error in estimating the k_{th} part:

    $E_k \leftarrow \sum_{t \in k_{th} part} \{y_t - W_{t-1}^T \hat{\gamma}^{-k}(z_t)\}^2$;

  end

  $CV_i \leftarrow \frac{1}{K} \sum_{k=1}^{K} E_k$

end

Find minimum CV_i and corresponding c_i

$C^*_{opt} \leftarrow c_i$

Remark

$\{c_1, c_2, \ldots, c_m\}$ are the empirical grid points. Let $\{s_{t,j}\}$ be a group of known stationary processes for $j = 1, 2, ..., N$. N is the number of stationary processes included in the group. We choose $c_1 = 0$ and $c_m = min\{C_j^* | max_{1 \le t \le n} n^{-1/2} log(n) |s_{t,j}| < C_j^*\}$ as the empirical values. Then $c_i = (i-1)\frac{c_m}{m-1}$ for $i = 1, 2, ...m$. A larger m provides higher accuracy of our classifier but slower search speed. A smaller m leads to faster search speed but lower accuracy.

CHAPTER 3: ASYMPTOTIC NORMALITIES OF ESTIMATORS

As mentioned before, with index vector $\mathcal{I}$, we regroup "stationary" parts of W_{t-1} as 1 and "nonstationary" parts as 0. Although, the proposed SM may make some mistake by chance. However, as $n \to \infty$, we have

$$P(\mathcal{I} = 1, \ldots, d_1) \to 1,$$
$$P(\mathcal{I}^c = d_1 + 1, \ldots, d_1 + d_2) \to 1,$$

where d_1 and d_2 are number of variables in $\gamma_1(z)$ and $\gamma_2(z)$ respectively. Hence, as $n \to \infty$, we also have

$$P\Big(\gamma_1(z_t) = \gamma_{\mathcal{I}}(z_t)\Big) \to 1,$$
$$P\Big(\gamma_2(z_t) = \gamma_{\mathcal{I}^c}(z_t)\Big) \to 1,$$

3.1 Asymptotic Normality of $\hat{\gamma}_1(z)$

Let $\hat{\gamma}_1(z)$ be the estimation of $\gamma_1(z)$ given by local smoother. To establish the asymptotic property of $\hat{\gamma}_1(z)$, we define $M_k(z) = E[W_{t-1,1}^{\otimes k} | z_t = z]$ for $k = 1, 2$; $\mu_i = \int_{-\infty}^{\infty} u^i K(u) du$; $\nu_i = \int_{-\infty}^{\infty} v^i K^2(v) dv$ and $\Sigma_{\gamma_1}(z) = \sigma_\epsilon^2 \nu_0(K) M_2(z)^{-1} / f_z(z)$, which is non-stochastic and is exactly the same as that in Cai et al. (2000).

Theorem 3.1 *(Theorem 2.1. of Cai and Li, 2009)*

Under regularity conditions, (A1)-(A.8) in the Appendix 8, with probability going to one, the asymptotic distribution of $\hat{\gamma}_1(z)$

$$\sqrt{nh_1}[\hat{\gamma}_1(z) - \gamma_1(z) - \frac{1}{2} h_1^2 \mu_2(K) \gamma_1^{(2)}(z)] \xrightarrow{D} N(0, \Sigma_{\gamma_1}(z)),$$

where h_1 is the bandwidth used at this step for estimating $\hat{\gamma}_1(z)$

To obtain the optimal bandwidth h_1 for $\hat{\gamma}_1(z)$, we looked into the integrated asymptotic mean squared error (IAMSE) for $\hat{\gamma}_1(z)$. It's easy to derive the IAMSE for $\hat{\gamma}_1(z)$, which is the same as (2.11) from Cai and Li (2009).

$$IAMSE(\hat{\gamma}_1(z)) = \int [\frac{h^4}{4}\mu_2^2(K)||\gamma_1^{(2)}(z)||^2 + \frac{tr(\Sigma_{\gamma_1,0}(z))}{nh}]q(z)dz, \tag{3.1}$$

By minimizing the IAMSE, with respect to h, optimal h_1 is obtained as

$$h_{1,opt} = \Big(\int tr(\Sigma_{\gamma_1,0}(z))q(z)dz\Big) \times \Big(\big[\mu_2^2(K)||\gamma_1^{(2)}(z)||^2 q(z)dz\big]^{-1/5} n^{-1/5}\Big), \tag{3.2}$$

For a simple representation, we may rewrite it into $h_{1,opt} = c_1 n^{-1/5}$. Given $h_1 = h_{1,opt}$, the $IAMSE(\hat{\gamma}_1(z)) = O(n^{-4/5})$. However, $h_{1,opt}$ can not make estimation for $\gamma_2(z)$ optimal in terms of Mean Squared Error and convergence rate. This motivates me to propose a second step estimation and corresponding asymptotic properties. Also, the $h_1 = h_{1,opt}$ is adopted in the simulation and real example parts.

3.2 Asymptotic Normality of $\hat{\gamma}_2(z)$

To derive the asymptotic property of $\hat{\gamma}_2(z)$, we need to prepare some ingredients, which plays an important role in establishing the theoretical results. Let x_t be an $I(1)$ processes, it can be expressed as $x_t = x_{t-1} + \eta_t = x_0 + \sum_{s=1}^t \eta_s(\ t \geq 1\ )$, where $\{\eta_s\}$ is an $I(0)$ process with mean zero and variance Υ_η.
Hence, we have

$$\frac{x_{[nr]}}{\sqrt{n}} \equiv \frac{x_t}{\sqrt{n}} = \frac{x_0}{\sqrt{n}} + \frac{\sum_{s=1}^t \eta_s}{\sqrt{n}} = \frac{x_0}{\sqrt{n}} + \frac{\sum_{s=1}^{[nr]} \eta_s}{\sqrt{n}}, \tag{3.3}$$

$[nr]$ denotes the integer part of nr and $r = t/n$.

For iid η_s and ρ-mixing η_s, as $n \to \infty$

$$\frac{x_{[nr]}}{\sqrt{n}} \xrightarrow{D} \mathcal{Y}_\eta(r) \, , \tag{3.4}$$

where $\mathcal{Y}_\eta(r)$ is a m-dimensional Brownian motion on $[0,1]$ with covariance matrix Υ_η. "$\xrightarrow{D}$" represents convergence in distribution. m is the dimension of x_t, Billingsley(1999). As demonstrated in Leamma A.1 from Hong and Jiang(2017), the weak convergence can be strengthened to a strong one, which will be employed to derive our the theoretic results below.

Theorem 3.2 *Under regularity conditions, (A1)-(A.8) in the Appendix 8, with probability going to one, the asymptotic distribution of* $\hat{\gamma}_2(z)$

$$a_n^{-1/2} b_n \Big(\hat{\gamma}_2(z) - \gamma_2(z) - \frac{h_2^2}{2} \mu_2(K) \gamma_2^{(2)}(z) \Big) \xrightarrow{D} N(0, \sigma_v^2 I_{d_2}),$$

where h_2 is the bandwidth used at this step for estimating $\hat{\gamma}_2(z)$;
$d_2 = \sum_{i=1}^{k+1} \mathbf{1}_{\{i \in \mathcal{I}^C\}}$;
$S_{nj} = \sum_{t=1}^{n} \Omega_t^* W_{t-1,2}^{\otimes 2} K_{h_2}(z_t - z)(\frac{z_t - z}{h_2})^j$ for $j = 0, 1, 2$;
$a_n = \sum_{t=1}^{n} \Big(K_{h_2}(z_t - z)\{S_{n2} - (\frac{z_t - z}{h}) S_{n1}\} \Omega_t^* W_{t-1,2} \Big)^{\otimes 2}$;
$b_n = \sum_{t=1}^{n} K_{h_2}(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2}) S_{n1}\} \Omega_t^* W_{t-1,2}^{\otimes 2} = S_{n0} S_{n2} - S_{n1} S_{n1}$;
Detailed proof of the following Theorem 3.2 is also provided in Appendix 8.

Remark

By Theorem 3.2, $\hat{\gamma}_2(z)$ is an estimator for $\gamma_2(z)$ with $\mathcal{B}_n(z) = b_n^{-1} \Big(\sum_{t=1}^{n} K_{h_2}(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2}) S_{n1}\} \Omega_t^* W_{t-1,2}^{\otimes 2} [\hat{\gamma}_2(z_t) - \gamma_2(z)] \Big)$ and $\mathcal{V}_n(z) = a_n^{-1/2} \Big\{ \sum_{t=1}^{n} K_{h_2}(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2}) S_{n1}\} \Omega_t^* W_{t-1,2} v_t^* \Big\}$. With the limiting distribution of $\hat{\gamma}_2(z)$, we will be able to construct its theoretical confidence interval.

3.3 Asymptotic Normality of $\hat{\gamma}(z)$

As results from Theorems 3.1 and 3.2, we have the following Bahadur representations of proposed estimators, $\hat{\gamma}_1(z)$ and $\hat{\gamma}_2(z)$

$$\begin{cases} \sqrt{nh_1}(\hat{\gamma}_1(z) - \gamma_1(z) - \frac{1}{2}h_1^2\mu_2(K)\gamma_1^{(2)}(z)) \\ \qquad\qquad = \sqrt{nh_1}M_2^{-1}(z)\sum_{t=1}^{n} n^{-1}W_{t-1,1}v_tK_{h_1}(z_t - z)\{1 + o_p(1)\} \\ a_n^{-1/2}b_n\left(\hat{\gamma}_2(z) - \gamma_2(z) - \frac{1}{2}h_2^2\mu_2(K)\gamma_2^{(2)}(z)\right) = a_n^{-1/2}c_nv^{**}\{1 + o_p(1)\} \end{cases} \tag{3.5}$$

where $M_2(z) = E[W_{t-1,1}^{\otimes 2}|z_t = z]$, $\Omega_t^{**} = K_{h_2}(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2})S_{n1}\}\Omega_t^*$, $c_n = \left(\Omega_1^{**}W_{1,2}, \Omega_2^{**}W_{2,2}, \ldots, \Omega_n^{**}W_{n,2}\right)$ and $v^{**} = (v_1^*, v_2^*, \ldots, v_n^*)^T$.

Define our final estimator as

$$\hat{\gamma}(z) = \left(\hat{\gamma}_1^T(z), \hat{\gamma}_2^T(z)\right)^T.$$

Theorem 3.3 *Under regularity conditions, (A1)-(A.8) in the Appendix 8, the asymptotic distribution of $\gamma(z)$*

$$\begin{pmatrix} \sqrt{nh_1} & 0 \\ 0 & a_n^{-1/2}b_n \end{pmatrix} \left[\hat{\gamma}(z) - \gamma(z) - \frac{1}{2}\begin{pmatrix} h_1^2\mu_2(K)(\gamma_1^{(2)}(z))^T \\ h_2^2\mu_2(K)(\gamma_2^{(2)}(z))^T \end{pmatrix}\right] \xrightarrow{D} MN(\Sigma_\gamma(z)),$$

where $\gamma(z) = \left(\gamma_1^T(z), \gamma_2^T(z)\right)^T$ and $\Sigma_\gamma(z)$ is the covariance matrix given by

$$\Sigma_\gamma(z) = \begin{pmatrix} \nu_0(K)M_2(z)^{-1}/f_z(z) & \Sigma_{1,2} \\ \Sigma_{2,1} & I_{d_2} \end{pmatrix} \sigma_v^2,$$

where $\Sigma_{1,2} = \Sigma_{2,1}^T = lim_{n\to\infty}\sum_{t=1}^{n}\sqrt{h_1/n}M_2^{-1}(z)W_{t-1,\mathcal{I}}(I_n - \mathcal{S})^T c_n^T(a_n^{-1/2})^T$; $\mathcal{S} = \left(s_{i,j}\right)$ and $s_{i,j} = \frac{1}{n}W_{i-1,\mathcal{I}}^T M_2^{-1}(z)W_{j-1,\mathcal{I}}K_{h_1}(z_j - z)$. Detailed proof of the following Theorem 3.3 is also provided in Appendix 9. To check the performance of our proposed estimators, I will test it on both simulated data and real data. Those will be covered in the next chapters.

CHAPTER 4: TESTING PREDICTABILITY

As we mentioned before, it's an important task to provide a unifying estimation and an inference tool for the varying-coefficients models. In the following part, an empirical likelihood method is introduced to construct a condfidence region for $\gamma(z)$ or test $H_0 : \gamma(z) = \gamma$. In parametric statistics, the parameters determine the distribution; nonparametric statistics, we estimate the CDF directly using the empirical CDF, which is also the nonparametric maximum likelihood estimator of F. Empirical likelihood, introduced by Owen(1988, 1990), is a nonparametric approach for constructing confidence region. It has several nice properties over the confidence region based directly on the asymptotic normal distribution of the estimator. In particular, the coverage error of the empirical likelihood confidence region is proved to be of the same order through the support of the regression function. "This is a significant improvement over the confidence region based directly on the asymptotic normal distribution of the local smoother, which have a larger order of coverage error near the boundary " (See Chen and Qin (2000)). In addition, some other appealing features of this method also include that it does not require explicit estimation of the stationary density of z_t and the conditional variance of the error v_t given z_t (See Jiang and Hong(2018), Owen (1990) and Chan, Li and Peng(2012)).

The profile likelihood method has been proven to be a powerful tool for constructing empirical likelihood regions, especially for varying-coefficient models (Fan and Huang, 2005). Let's assume that an initial estimation of $\gamma_2(z)$ is given before we run the step-one estimation. Firstly, we regress $y_t - W_{t-1,2}^T \hat{\gamma}_2(z_t)$ on $W_{t-1,1}$. Then regress $y_t - W_{t-1,1}^T \hat{\gamma}_1(z_t)$ on $W_{t-1,2}$. For the oracle property of local smoother, this method will output the same estimation as $\hat{\gamma}(z)$ from the step 1 in our proposed procedure

(Fan and Gijbels (1996)). Although this assumption will not make any difference in terms of estimation, it facilitates us to propose the following empirical likelihood approach.

For the local smoother in step-one, given $\gamma_2(z)$ and $W_{t-1,2}$, the corresponding profile local score equations are

$$\sum_{t=1}^{n} Q_t[y_t - W_{t-1,2}^{\top}\gamma_2(z) - W_{t-1,1}^{T}\{\xi_1 + \eta_1(z_t - z)\}]K_{h_1}(z_t - z) = 0, \qquad (4.1)$$

where $Q_t = (1, z_t - z)^T \otimes W_{t-1,1}$, with $A \otimes B$ denoting the Kronecker product of A and B. In (4.1), by eliminating η_1 and solving for ξ_1, we get the estimation equation:

$$\sum_{t=1}^{n} K_{h_1}(z_t - z)\{T_{n2} - (z_t - z)T_{n1}\}W_{t-1,1}\{y_t - W_{t-1,1}^{\top}\gamma_1(z) - W_{t-1,2}^{\top}\gamma_2(z)\} = 0. \quad (4.2)$$

Given $\gamma_1(z)$ and $W_{t-1,1}$, the profile weighted local score equations (WLSE) are

$$\sum_{t=1}^{n} Q_t^*[y_t - W_{t-1,1}^{\top}\gamma_1(z) - W_{t-1,2}^{T}\{\xi_2 + \eta_2(z_t - z)\}]K_{h_2}(z_t - z) = 0, \qquad (4.3)$$

where $Q_t^* = (1, z_t - z)^T \otimes (\Omega_t^* W_{t-1,2})$. In (4.2) by eliminating η_2 and solving for ξ_2, we get the estimation equation:

$$\sum_{t=1}^{n} K_{h_2}(z_t - z)\{S_{n2} - (z_t - z)S_{n1}\}\Omega_t^* W_{t-1,2}\{y_t - W_{t-1,1}^{\top}\gamma_1(z) - W_{t-1,2}^{\top}\gamma_2(z)\} = 0. \quad (4.4)$$

By combining (4.2), (4.4), the estimation equations are obtained

$$\begin{cases} Z_{1t}(\gamma(z)) = 0 \\ Z_{2t}(\gamma(z)) = 0 \end{cases}$$

where h_1, h_2 are the corresponding bandwidths in our estimation procedure;

$S_{nj} = \sum_{t=1}^{n} \Omega_t^* W_{t-1,2}^{\otimes 2} K_{h_2}(z_t - z)(z_t - z)^j$;

$T_{nj} = \sum_{t=1}^{n} W_{t-1,1}^{\otimes 2} K_{h_1}(z_t - z)(z_t - z)^j$;

$Z_{1t}(\gamma(z)) = K_{h_1}(z_t - z)\{T_{n2} - (z_t - z)T_{n1}\}W_{t-1,1}\{y_t - W_{t-1,1}^{\top}\gamma_1(z) - W_{t-1,2}^{\top}\gamma_2(z)\}$;

$Z_{2t}(\gamma(z)) = K_{h_2}(z_t - z)\{S_{n2} - (z_t - z)S_{n1}\}\Omega_t^* W_{t-1,2}\{y_t - W_{t-1,1}^{\top}\gamma_1(z) - W_{t-1,2}^{\top}\gamma_2(z)\}$.

Therefore, we define the empirical likelihood ratio(See Chen and Keilegom,2009):

$$L_n(\gamma(z)) = \sup\Big\{\prod_{t=1}^{n}(np_t) : p_t \geq 0, \sum_{t=1}^{n} p_t = 1, \sum_{t=1}^{n} p_t Z_t(\gamma(z)) = 0\Big\},$$

where $Z_t(\gamma(z)) = \Big(Z_{1t}(\gamma(z)), Z_{2t}(\gamma(z))\Big)$.

By using Lagrange multiplier technique, we obtain $p_t = n^{-1}\{1 + \lambda^{\top} Z_t(\gamma(z)\}^{-1}$. Hence, the log empirical likelihood ratio is

$$\ell_n(\gamma(z)) = -2\log L(\gamma(z)) = 2\sum_{t=1}^{n} \log\{1 + \lambda^{\top} Z_t(\gamma(z))\}, \tag{4.5}$$

where $\lambda = \lambda(\xi)$ satisfies $\sum_{t=1}^{n} Z_t(\gamma(z))/\{1 + \lambda^{\top} Z_t(\gamma(z))\} = 0$. It's not expensive to evaluate the log empirical likelihood ratio that the objective function in terms of computational cost, since $\ell_n(\gamma(z))$ is concave in λ. Our next theorem demonstrates that the Wilks result holds for the above empirical likelihood ratio.

Theorem 4.1 *Under regularity conditions, (A1)-(A.8) in the Appendix A and* $E|v_t|^3 <$

∞, *The* $\ell_n(\gamma(z))$ *converges in distribution to* $\chi^2(k+1)$*, a chi-squared distribution with degrees of freedom* $k+1$*, as* $n \to \infty$.

By Theorem 4.1, we can construct a $100(1-\alpha)\%$ confidence region for $\gamma(z)$ as $\ell_n(\gamma(z)) = \{\gamma(z) : \ell_n(\gamma(z)) \leq \chi^2_{k+1,\alpha}\}$.

Detailed proof of the Theorem 4.1 is provided in Appendix 10.

CHAPTER 5: SIMULATIONS

To investigate performance of the proposed Two-step estimation procedure, 1000 simulations were conducted based on the following model.

$$y_t = x_{t-1}^{\top}\,\beta(z_t) + \theta(z_t) + v_t, \quad x_t = (x_{t,1}, \ldots, x_{t,k})^{\top},$$

$$x_{t,i} = \rho_i x_{t-1,i} + u_{t,i}, \;\; u_{t,i} \sim N(0,1),$$

which is equivalent to

$$y_t = W_{t-1}^{\top}\,\gamma(z_t) + v_t, \tag{5.1}$$

$$W_{t,i} = \rho_i W_{t-1,i} + u_{t,i}, \;\; u_{t,i} \sim N(0,1),$$

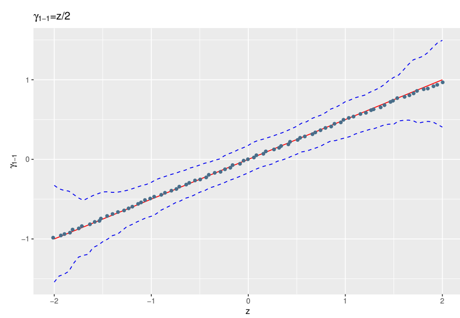
where $W_{t-1} = (x_{t-1}^{\top}, 1)^{\top}$ and $\gamma(z_t) = (\beta(z_t)^{\top}, \theta(z_t))^{\top}$, which is a vector of smooth functions of z_t. The model is tested on the following two different settings.

Example 5.1

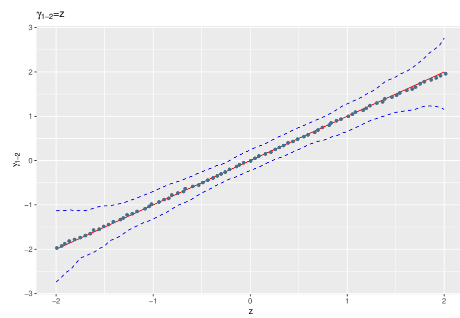
We set $\gamma(z) = [\gamma_1(z), \gamma_2(z), \gamma_3(z), \gamma_4(z)]^T = [z/2, z, sin(z), cos(z)]^T$, $W_{t,1} = 0.8W_{t-1,1} + u_{t,1}$, $W_{t,2} = 0.5W_{t-1,2} + u_{t,2}$, $W_{t,3} = W_{t-1,3} + u_{t,3}$, $W_{t,4} = W_{t-1,4} + u_{t,4}$, $z_t = 0.5z_{t-1} + u_t$, $v_t \sim N(0,1)$, $u_{t,1}$, $u_{t,2}$, $u_{t,3}$, $u_{t,4}$, u_t are $\sim N(0,1)$. $n = 300$ with 1000 replications.

Our simulation involves the choices of kernel function $K(.)$ and bandwiths h_1 and h_2, which needs to be specified in step-one and step-two estimations. One can use any data-driven method to select h_1 and h_2 optimally. For simplicity, Gaussian kernel and rule-of-thumb bandwidths $h_1 = cn^{-1/5}(c > 0)$, $h_2 = 1.06Sn^{-1/3}$ are adopted here. where S is the sample standard deviation of z_t and c is a tuning parameter which can be chosen by cross-validation.

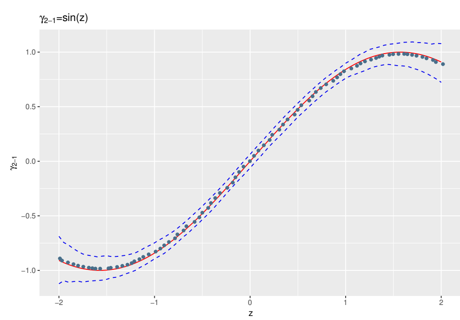
A sample of size $n = 300$ is drawn for both Example 4.1. In Figure 5.1, Red solid lines show true curve for the intended varying-coefficients. Dark blue dash dot lines are the pointwise median among the 1000 simulations. Light blue dash dot lines show the the pointwise 2.5% and 97.5% percentiles among 1000 simulations.



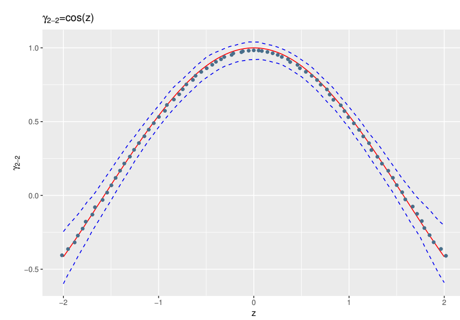
$\gamma_1(z) = z/2$



$\gamma_2(z) = z$



$\gamma_3(z) = sin(z)$



$\gamma_4(z) = cos(z)$

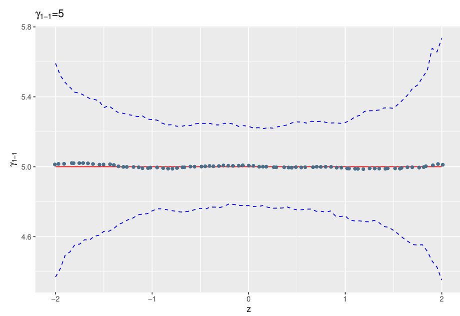
Figure 5.1: Example 4.1: Estimated Varying-coefficients $\gamma(z)$ with $z \in [-2, 2]$.

Example 5.2

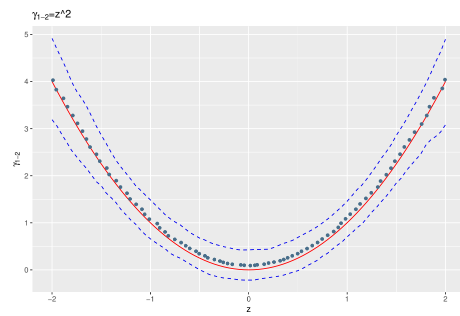
We set $\gamma(z) = [\gamma_1(z), \gamma_2(z), \gamma_3(z), \gamma_4(z)]^T = [5, z^2, sin(z)/2, cos(z)/2]^T$, $W_{t,1} = 0.8W_{t-1,1} + u_{t,1}$, $W_{t,2} = 0.5W_{t-1,2} + u_{t,2}$, $W_{t,3} = W_{t-1,3} + u_{t,3}$, $W_{t,4} = W_{t-1,4} + u_{t,4}$, $z_t = 0.5z_{t-1} + u_t$, $v_t \sim t(3)$, $u_{t,1}$, $u_{t,2}$, $u_{t,3}$, $u_{t,4}$, u_t are $\sim N(0, 1)$. $n = 300$ with 1000 replications.

Our simulation involves the choices of kernel function $K(.)$ and bandwiths h_1 and h_2, which needs to be specified in step-one and step-two estimations. One can use any data-driven method to select h_1 and h_2 optimally. For simplicity, Gaussian kernel and rule-of-thumb bandwidths $h_1 = cn^{-1/5}(c > 0)$, $h_2 = 1.06Sn^{-1/3}$ are adopted here. where S is the sample standard deviation of z_t and c is a tuning parameter which can be chosen by cross-validation.

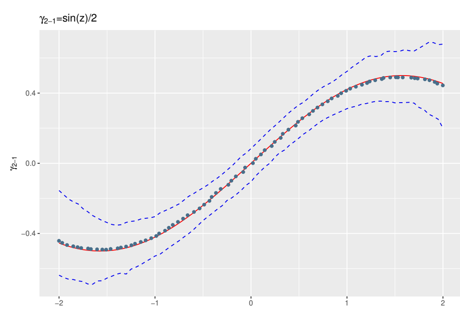
A sample of size $n = 300$ is drawn Example 4.2. In Figure 5.2, Red solid lines show true curve for the intended varying-coefficients. Dark blue dash dot lines are the pointwise median among the 1000 simulations. Light blue dash dot lines show the the point-wise 2.5% and 97.5% percentiles among 1000 simulations.



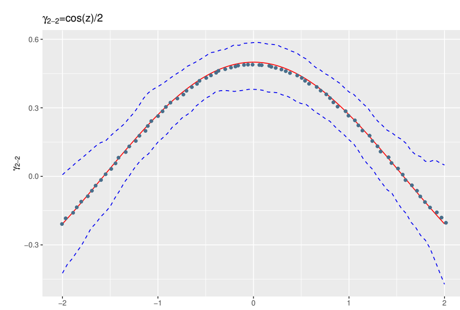
$\gamma_1(z) = 5$



$\gamma_2(z) = z^2$



$\gamma_3(z) = sin(z)/2$



$\gamma_4(z) = cos(z)/2$

Figure 5.2: Example 4.2:Estimated Varying-coefficients $\gamma(z)$ with $z \in [-2, 2]$.

We noticed that the proposed unifying two-step method gives very precise estimation. The median nearly coincide with the true curve. And the 95% confidence interval covers the true curve when z values vary. To check the performance of estimators in a more practical setting, we would like to test it on a real example in the Chapter 5.

Example 5.3

To investigate the performance of our proposed empirical likelihood method, we run 1000 simulations for the following model

$$y_t = W_{t-1}^{\top}\gamma(z_t) + v_t,$$

$W_{t-1} = (x_{t-1}^{\top}, 1)^{\top}$ and $\gamma(z_t) = (\beta(z_t)^{\top}, \theta(z_t))^{\top}$, which is a vector of smooth functions of z_t. The data generating process is identical to Example 5.1 in the simulation section with sample size $n = 300$. We set $\gamma(z) = [\gamma_1(z), \gamma_2(z), \gamma_3(z), \gamma_4(z)]^T = [z/2, z, sin(z), cos(z)]^T$, $W_{t,1} = 0.8W_{t-1,1} + u_{t,1}$, $W_{t,2} = 0.5W_{t-1,2} + u_{t,2}$, $W_{t,3} = W_{t-1,3} + u_{t,3}$, $W_{t,4} = W_{t-1,4} + u_{t,4}$, $z_t = 0.5z_{t-1} + u_t$, $v_t \sim N(0,1)$, $u_{t,1}$, $u_{t,2}$, $u_{t,3}$, $u_{t,4}$, u_t are $\sim N(0,1)$. $n = 300$ with 1000 replications.

We set $h_1 = 1.02n^{-1/5a}$, $h_2 = 1.06Sn^{-1/3}$ or $h_1 = 1.02n^{-1/5}$, $h_2 = 1.06Sn^{-1/3a}$, where S is the sample standard deviation of z_t. The performance of our proposed estimator is directly influenced by the bias, which is controlled by the choice of bandwidths h_1 and h_2. To investigate into this, we chose three levels of bandwidth a, where $a = 1, \frac{1}{2}$ and $\frac{1}{4}$.

Tables 5.1 present summarized simulated results of the three bandwidth levels at $z = 1$, with average, bias of $\hat{\gamma}(1)$ and the coverage probability of the 95% EL-based confidence regions among the 1000 simulations. Note that true value $\gamma(1) = (1/2, 1, sin(1), cos(1))^T = (0.5, 1, 0.8414, 0.5403)^T$.

Table 5.1: Simulation Results for $\gamma(1)$

		$h_1 = 1.02n^{-1/5a}$, $h_2 = 1.06Sn^{-1/3}$	$h_1 = 1.02n^{-1/5}$, $h_2 = 1.06Sn^{-1/3a}$
	ave	(0.4916, 0.9926, 0.8288, 0.5291)	(0.4916, 0.9926, 0.8288, 0.5291)
a = 1	bias	(-0.0084, -0.0074, -0.0126, -0.0112)	(-0.0084, -0.0074, -0.0126, -0.0112)
	sd	(0.1462, 0.2004 0.0657, 0.0467)	(0.1462, 0.2004, 0.0667, 0.0467)
	cp	94.5%	94.5%
	ave	(0.4897, 0.9826, 0.8282, 0.5282)	(0.4916, 0.9926, 0.7711, 0.4917)
a=1/2	bias	(-0.0103, -0.0174, -0.0132, -0.0121)	(-0.0084, -0.0074, -0.0703, -0.0486)
	sd	(0.1343, 0.1909, 0.0661, 0.0471)	(0.1462, 0.2004, 0.0518, 0.0351)
	cp	91.4%	70.8%
	ave	(0.4863, 0.9777, 0.8276, 0.5277)	(0.4916, 0.9926, 0.7073, 0.4575)
a=1/4	bias	(-0.0137, -0.0223, -0.0138, -0.0126)	(-0.0084, -0.0074, -0.1341, -0.0828)
	sd	(0.1412, 0.2059, 0.0662, 0.0474)	(0.1462, 0.2004, 0.0617, 0.0412)
	cp	87.4%	39.7%

Tables 5.2 present summarized simulated results of the three bandwidth levels at $z = -1$, with average, bias of $\hat{\gamma}(-1)$ and the coverage probability of the 95% EL-based confidence regions among the 1000 simulations. Note that true value $\gamma(-1) = (-1/2, -1, sin(-1), cos(-1))^T = (-0.5, -1, -0.8415, 0.5403)^T$.

Table 5.2: Simulation Results for $\gamma(-1)$

		$h_1 = 1.02n^{-1/5a}$, $h_2 = 1.06Sn^{-1/3}$	$h_1 = 1.02n^{-1/5}$, $h_2 = 1.06Sn^{-1/3a}$
	ave	(-0.4971, -0.9797, -0.8206, 0.5242)	(-0.4971, -0.9797, -0.8206, 0.5242)
a = 1	bias	(0.0029, 0.0203, 0.0208, -0.0161)	(0.0029, 0.0203, 0.0208, -0.0161)
	sd	(0.1409, 0.2260, 0.0727, 0.0486)	(0.1409, 0.2260, 0.0727, 0.0486)
	cp	94.3%	94.3%
	ave	(-0.4874, -0.9714, -0.8189, 0.5229)	(-0.4971, -0.9796, -0.7598, 0.4867)
a=1/2	bias	(0.0126, 0.0286, 0.0225, -0.0174)	(0.0029, 0.0204, 0.0816, -0.0536)
	sd	(0.1273, 0.1959, 0.0722, 0.0474)	(0.1409, 0.2260, 0.0536, 0.0354)
	cp	90.9%	72.1%
	ave	(-0.4865, -0.9682, -0.8184, 0.5221)	(-0.4971, -0.9796, -0.7001, 0.4561)
a=1/4	bias	(0.0135, 0.0318, 0.0230, -0.0182)	(0.0029, 0.0204, 0.1413, -0.0842)
	sd	(0.1378, 0.2055, 0.0726, 0.0475)	(0.1409, 0.2260, 0.0642, 0.0421)
	cp	88.5%	51.2%

We observe that as the bandwidth increases the coverage probability decreases. We also observe a positive relationship between coverage probability and the bias. Our rule of thumb bandwidths $h_1 = 1.02n^{-1/5}$, $h_2 = 1.06Sn^{-1/3}$ perform the best in terms of coverage probability and bias. This reflects that the performance of our proposed empirical likelihood method relies on a proper choice of bandwidths.

CHAPTER 6: REAL EXAMPLE

To check performance of the proposed estimators in a practical setting, we consider a real example here. We download the 5-year daily treasury yield rate, the 6-month daily treasury yield rate, the stock price of Morgan Stanley and the price of S&P 500 from Yahoo Finance. All data are from Nov. 12th, 2012 to Nov. 12th, 2016 with 1004 data points. We consider a sample with size n = 700 for estimation, then forecast 1-day forward for next 304 trading days. We build our model based on Capital Asset Pricing Model (CAPM):

$$\begin{aligned} y_t =& W_{t-1}^{\top}\gamma(z_t) + v_t \\ =& \gamma_1(z_t)W_{1,t-1} + \gamma_2(z_t)W_{2,t-1} + \gamma_3(z_t)W_{3,t-1} + \gamma_4(z_t)W_{4,t-1} + v_t, \end{aligned} \tag{6.1}$$

where y_t is the S&P 500 index. We choose logarithmic difference of the 5-year daily treasury yield rate and logarithmic difference the 6-month daily treasury yield rate as $W_{1,t-1}$, $W_{2,t-1}$ respectively. These two predictors can be treated as $I(0)$ processes. Then we choose the S&P 500 index lagged by 1-day as $W_{4,t-1}$ and the stock price for Morgan Stanley as $W_{3,t-1}$, which are $I(1)$ processes. Let $z_t = log(R_{1,t}) - log(R_{2,t})$ be the spread in the logged interest rates. $R_{1,t}$ is the 10 year daily treasury yield rate. $R_{2,t}$ is the 2 year daily treasury yield rate. As mentioned in Tsay (1998), the magnitude of z_t may reflect the status of the United States economy, so it is reasonable to use z_t as a threshold variable. To check if z_t is stationary, we conduct the following ADF test. Table 6.1 reports the testing results, which indicates the stationarity of z_t.

Table 6.1: Augmented Dickey Fuller (ADF) Test

	ADF Test	P-Value
Yield Spread	-12.212	$<$0.01

We always split the sample into two parts, training sample and test sample. The first part with sample size $n = 700$ is for estimation. Then we use the built model for the 1-step ahead forecast. Then we compare the estimated S&P 500 index with historical data. We repeat the process for the next 300 trading days, from Aug 23rd, 2015 to Nov 7th, 2016.

Figure 6.1 shows the relationship between historical data of S&P 500 index and the forecast based on our proposed unifying two-step estimation procedure. The red line shows the true curve for S&P 500 index, and black solid dots are the estimated the values. Results indicated that the proposed unifying estimation models the relationship between S&P 500 index and predictors very well. The new procedure is promising and may be applied to many different models.

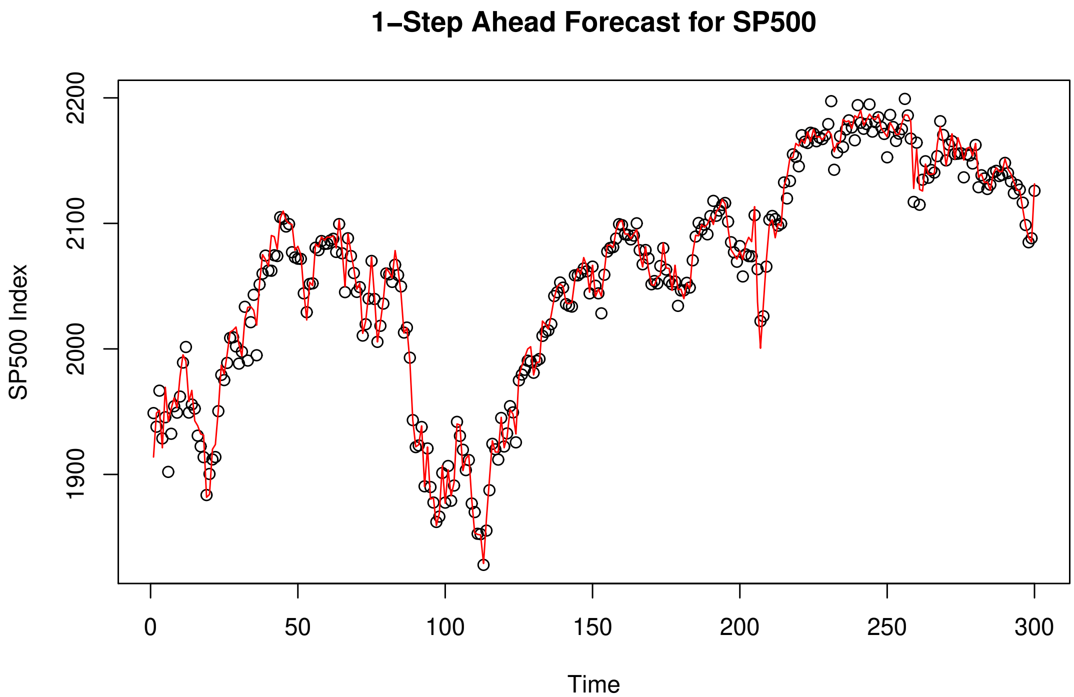
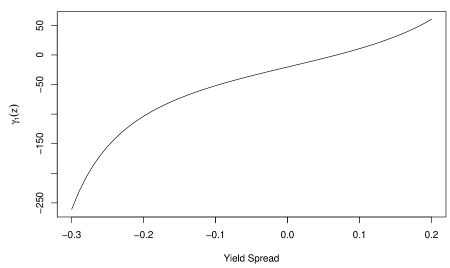
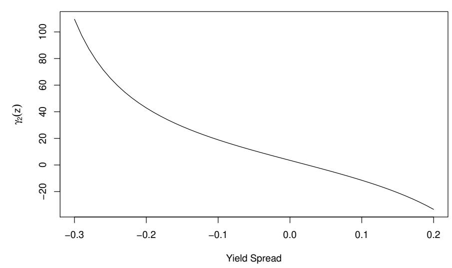


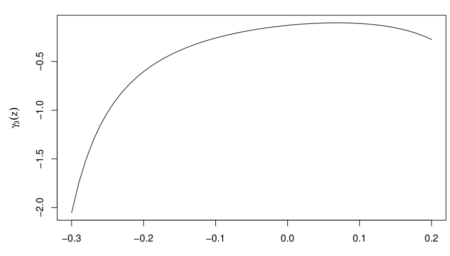
Figure 6.1: 1-Step Ahead Forecast for S&P 500 Index



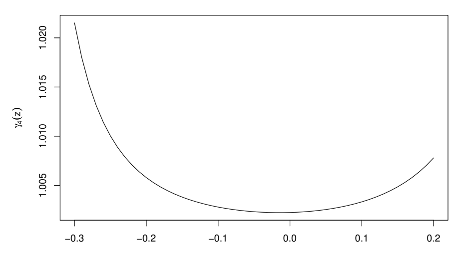
Log Difference 5-year T-bill



Log Difference 6-month T-bill



Morgan Stanley



S&P 500 Index - lagged by 1 day

Figure 6.2: Estimated Varying-coefficients for Model

CHAPTER 7: DISCUSSION

In this dissertation, we have proposed a unifying two-step estimation procedure for varying-coefficients models, which facilitates the unifying inference for coefficients. The proposed two-step procedure provides a unifying estimation procedure for the varying-coefficients models. The asymptotic joint distribution of the proposed estimators has been established, which provides a Wald type of confidence regions for the coefficient functions. However, this confidence region does not work well when the conditional variance of the error changes. To solve this problem, we have proposed an empirical likelihood inference tool for the coefficient functions.

In our future research, it's interesting to investigate the following topics related to this dissertation. Firstly, it's very important to discuss about the optimal bandwidths selection theoretically and empirically. Secondly, it's an interesting extension to generalize the asymptotic analysis of this dissertation to the case when z_t is nonstaionary. Last but not the least, for $W_{t-1,2}$, nonstationary part of W_{t-1}, one can extend it from $I(1)$ process to $I(2)$, $I(3)$ or even $I(p)$ process.

REFERENCES

[1] L. M. Viceira, "Testing for structural change in the predictability of asset returns," *Manuscript, Harvard University*, vol. 4, no. 3.5, pp. 3–0, 1997.

[2] Y. Amihud and C. M. Hurvich, "Predictive regressions: A reduced-bias estimation method," *Journal of Financial and Quantitative Analysis*, vol. 39, no. 4, pp. 813–841, 2004.

[3] R. S. Tsay, "Testing and modeling multivariate threshold models," *journal of the american statistical association*, vol. 93, no. 443, pp. 1188–1202, 1998.

[4] G. Elliott and J. H. Stock, "Inference in time series regression when the order of integration of a regressor is unknown," *Econometric theory*, vol. 10, no. 3-4, pp. 672–700, 1994.

[5] J. Y. Campbell and M. Yogo, "Efficient tests of stock return predictability," *Journal of financial economics*, vol. 81, no. 1, pp. 27–60, 2006.

[6] C. Polk, S. Thompson, and T. Vuolteenaho, "Cross-sectional forecasts of the equity premium," *Journal of Financial Economics*, vol. 81, no. 1, pp. 101–141, 2006.

[7] C. L. Cavanagh, G. Elliott, and J. H. Stock, "Inference in models with nearly integrated regressors," *Econometric theory*, vol. 11, no. 5, pp. 1131–1147, 1995.

[8] W. Torous, R. Valkanov, and S. Yan, "On predicting stock returns with nearly integrated explanatory variables," *The Journal of Business*, vol. 77, no. 4, pp. 937–966, 2004.

[9] Z. Cai and Y. Wang, "Testing predictive regression models with nonstationary regressors," *Journal of Econometrics*, vol. 178, pp. 4–14, 2014.

[10] P. C. Phillips *et al.*, *Spectral regression for cointegrated time series*. Cowles Foundation for Research in Economics at Yales University, 1988.

[11] J. Jiang and Y. Mack, "Robust local polynomial regression for dependent data," *Statistica Sinica*, vol. 11, no. 3, pp. 705–722, 2001.

[12] Z. Xiao, "Quantile cointegrating regression," *Journal of Econometrics*, vol. 150, no. 2, pp. 248–260, 2009.

[13] J. Fan, Q. Yao, and Z. Cai, "Adaptive varying-coefficient linear models," *Journal of the Royal Statistical Society: series B (statistical methodology)*, vol. 65, no. 1, pp. 57–80, 2003.

[14] I. Berkes and L. Horváth, "Convergence of integral functionals of stochastic processes," *Econometric Theory*, vol. 22, no. 2, pp. 304–322, 2006.

[15] P. J. Bickel, "One-step huber estimates in the linear model," *Journal of the American Statistical Association*, vol. 70, no. 350, pp. 428–434, 1975.

[16] R. J. Carroll, J. Fan, I. Gijbels, and M. P. Wand, "Generalized partially linear single-index models," *Journal of the American Statistical Association*, vol. 92, no. 438, pp. 477–489, 1997.

[17] M. Jansson and M. J. Moreira, "Optimal inference in regression models with nearly integrated regressors," *Econometrica*, vol. 74, no. 3, pp. 681–714, 2006.

[18] N. H. Chan, C.-Z. Wei, *et al.*, "Asymptotic inference for nearly nonstationary ar (1) processes," *The Annals of Statistics*, vol. 15, no. 3, pp. 1050–1063, 1987.

[19] S. Datta, "On asymptotic properties of bootstrap for ar (1) processes," *Journal of Statistical Planning and Inference*, vol. 53, no. 3, pp. 361–374, 1996.

[20] J. Fan and Q. Yao, "Nonlinear time series: Nonparametric and parametric methods springer-verlag," *New York*, 2003.

[21] P. Hall and B.-Y. Jing, "Comparison of bootstrap and asymptotic approximations to the distribution of a heavy-tailed mean," *Statistica Sinica*, pp. 887–906, 1998.

[22] J. Lederer and C. Müller, "Don't fall for tuning parameters: tuning-free variable selection in high dimensions with the trex," 2015.

[23] Y. Fan and C. Y. Tang, "Tuning parameter selection in high dimensional penalized likelihood," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 75, no. 3, pp. 531–552, 2013.

[24] P. Hall, *The bootstrap and Edgeworth expansion.* Springer Science & Business Media, 2013.

[25] S. Ling, "Self-weighted least absolute deviation estimation for infinite variance autoregressive models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 3, pp. 381–393, 2005.

[26] J. Fan, *Local polynomial modelling and its applications: monographs on statistics and applied probability 66.* Routledge, 2018.

[27] J. Fan and J. Jiang, "Nonparametric inferences for additive models," *Journal of the American Statistical Association*, vol. 100, no. 471, pp. 890–907, 2005.

[28] J. Y. Park and S. B. Hahn, "Cointegrating regressions with time varying coefficients," *Econometric Theory*, vol. 15, no. 5, pp. 664–703, 1999.

[29] Q.-M. SHAO and C.-R. LIU, "Strong approximations for partial sums of weakly dependent random variables," *Science in China Series A-Mathematics, Physics, Astronomy & Technological Science*, vol. 30, no. 6, pp. 575–587, 1987.

[30] R. S. Tsay, *Analysis of financial time series*, vol. 543. John Wiley & Sons, 2005.

[31] B. Rosenberg, "Prediction of common stock betas," *The Journal of Portfolio Management*, vol. 11, no. 2, pp. 5–14, 1985.

[32] S. Zhou, X. Shen, D. Wolfe, *et al.*, "Local asymptotics for regression splines and confidence regions," *The annals of statistics*, vol. 26, no. 5, pp. 1760–1782, 1998.

[33] J. L. Horowitz and V. G. Spokoiny, "An adaptive, rate-optimal test of linearity for median regression models," *Journal of the American Statistical Association*, vol. 97, no. 459, pp. 822–835, 2002.

[34] Z. Cai, J. Fan, and Q. Yao, "Functional-coefficient regression models for nonlinear time series," *Journal of the American Statistical Association*, vol. 95, no. 451, pp. 941–956, 2000.

[35] P. C. Phillips and C. Han, "Gaussian inference in ar (1) time series with or without a unit root," *Econometric Theory*, vol. 24, no. 3, pp. 631–650, 2008.

[36] P. D. Koch and T. W. Koch, "Evolution in dynamic linkages across daily national stock indexes," *Journal of International Money and Finance*, vol. 10, no. 2, pp. 231–251, 1991.

[37] Z. Cai, Q. Li, and J. Y. Park, "Functional-coefficient models for nonstationary time series data," *Journal of Econometrics*, vol. 148, no. 2, pp. 101–113, 2009.

[38] P. Hall and C. Heyde, "Martingale limit theory and its applications (1980)."

[39] P. Ravikumar, J. Lafferty, H. Liu, and L. Wasserman, "Sparse additive models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 71, no. 5, pp. 1009–1030, 2009.

APPENDIX 8: PROOF OF THEOREMS IN SECTIONS 3.1, 3.2

Notation and Conditions:

For convenience, some notations from Hong and Jiang (2017) and Li and Cai(2009) adopted here. Let $U_{n,i}(r) = n^{-1/2}x_{[nr],i}$, where $r = t/n$ and $[x]$ denotes the integer part of x. Under some regularity conditions on $u_{t,i}$ in Philips(1988), it will lead to

$$U_{n,i}(r) \xrightarrow{D} U_{\gamma_i}(r) \tag{8.1}$$

as $n \to \infty$ and $U_{\gamma_i}(r) = \int_0^r exp[(r-s)\gamma_i]dW_u^{(i)}(s)$, which is a diffusion process and $W_u^{(i)}(s)$ is a Brownian motion. $Var[W_u^{(i)}(s)] = Var(u_{1,i}) + 2\sum_{s=2}^{\infty} E[u_{1,i}u_{s,i}]$.

The following notation and regularity conditions are needed for our asymptotic results. We make the following assumptions.

Assumptions:

A1. $\gamma(z)$ is twice continuously differentiable in z for all $z \in R$.

A2. The kernel function $K(.)$ is a symmetric and continuous density function, supported by $[-1, 1]$. And let $\mu_i = \int_{-1}^1 u^i K(u)du$ and $\nu_i = \int_{-1}^1 u^i K^2(u)du$

A3. The bandwith h satisfies $h \to 0$ and $nh \to \infty$.

A4. v_t has a finite foutth moment, $E(v_t|x_t, z_t) = 0$ and $E(v_t^2|x_t, z_t) = \sigma_v^2$ is a positive constant.

A5. Time series z_t is stationary and has continuous stationarity density $f(z)$ with bounded support $supp(f)$.

A6. If $x_{t,i}$ is stationary or an $AR(1)$ process with $|\rho_i| < 1$, then the condition expectation $\alpha_i(z) = E(x_{t,i}|z_t = z)$ has continous second order derivative and the conditional variance $\sigma_i(z) = Var(x_{t,i}|z_t = z)$ is continuous on $z \in supp(f)$. Furthermore, $E|x_{t,i}|^4 < \infty$

A7. If $x_{t,i}$ is an $AR(1)$ process, then $E(u_{0,i}) = 0$, $E|u_{t,i}|^{k_1+k_2} < \infty$ for some $k_1 > 2$ and $k_2 > 0$ and $\{u_{t,i}\}_{t=0}^{\infty}$ is α-mixing with mixing coefficients $\alpha_i(s)$ satisfying $\sum_{s=1}^{\infty}\{\alpha_i(s)\}^{1-2/k_1}$.

A8. If $x_{t,i}$ is not an $AR(1)$ process, we assume it's an ρ-mixing process with mixing coefficients $\rho_i^*(s)$ satisfying $\sum_l \rho_i^*(l) < \infty$ or is an α-mixing process with mixing coefficients local smoother$\alpha_i^*(s)$ satisfying $\sum_l l^p \alpha_i^*(l)^q < \infty$, for some $0 < q < 1$ and $p > q$.

Conditions(A1.)-(A.4) are standard in local smoothing, and conditions and used in Cai and Li(2009), conditions (A5.)-(A7.) are used in Shaoxin(2018), condition (A8.) is a general condition used for a stationary process and also used in Jiang and Mack(2001).

Lemma A.1 (Theorem 3.3 of Hansen(1992)) Suppose $U_n \xrightarrow{D} U$ in $\mathcal{D}_{Mkm}[0,1]$ and $U(.)$ is a.s. continous. For a random sequence e_j and a sequence of nondecreasing σ-field $\mathcal{F}_j^e$ to which e_j is adapted, assume that $sup_j E|E(e_j|\mathcal{F}_{j-m}^e)| \to 0$ as $m \to \infty$. Then

$$sup_{0\leq s\leq 1}|n^{-1}\sum_{j=1}^{[ns]} U_{nj}e_j| \xrightarrow{p} 0$$

Lemma A.2 (Lemma A.3 of Cai(2009)) Let $w_t = h^{1/2}K_h(z_t - z)\epsilon_t(h^{-1}(z_t - z))^j$, and $U_{nt}^* = n^{-1/2}W_{t-1,\mathcal{I}^c}$. Let $\mathcal{F}_t = \sigma(U_{ni}^*, w_i : i \leq t)$ to be the smallest σ - field containing the past history of (U_{nt}^*, w_t) for all n and $i \leq t$. For any $0 \leq r \leq 1$, define $\mathcal{L}_n(r) \equiv n^{-1}\sum_{i=1}^{[nr]}\eta_i\xi_i - n^{-1}U_{[nr]}\xi_{[nr]+1}$, where $\xi_i = \sum_{k=1}^{\infty} E_i(w_{i+k})$ and η_t is from $W_{t-1,\mathcal{I}^c} = \sum_{s=1}^{t}\eta_s$. Then, we have

$$sup_{0\leq r\leq 1}|\mathcal{L}_n(r)| = o_p(1)$$

.

Proof of Theorem 3.1 Firstly, we define $D_n = diag\{I_{d_1}, \sqrt{n}I_{d_2}\}$, where d_1, d_2 are the dimensions of $W_{t-1,\mathcal{I}^c}$ and $W_{t-1,\mathcal{I}}$. Given D_n, we define $\mathcal{A}_n = diag\{1, h\} \otimes D_n$. By rewrite the estimators in step-one, we obtain the following expression.

$$\begin{aligned}\mathcal{A}_n\begin{pmatrix}\hat{\gamma}(z)\\ \hat{\gamma}^{(1)}(z)\end{pmatrix}^T =&\mathcal{A}_n\Big[\sum_{t=1}^{n}\begin{pmatrix}W_{t-1}\\ (z_t-z)W_{t-1}\end{pmatrix}^{\otimes 2}K_h(z_t-z)\Big]^{-1}\\ &\times\Big[\sum_{t=1}^{n}\begin{pmatrix}W_{t-1}\\ (z_t-z)W_{t-1}\end{pmatrix}y_tK_h(z_t-z)\Big]\\ =&G_n(z)^{-1}n^{-1}\Big[\sum_{t=1}^{n}K_h(z_t-z)\Big]\\ &\times y_t\begin{pmatrix}1\\ h^{-1}(z_t-z)\end{pmatrix}\otimes(D_n^{-1}W_{t-1})\end{aligned}\tag{8.2}$$

where

$$\begin{aligned}G_n(z) =&n^{-1}\sum_{t=1}^{n}K_h(z_t-z)\begin{pmatrix}1\\ h^{-1}(z_t-z)\end{pmatrix}^{\otimes 2}\otimes(D_n^{-1}W_{t-1})^{\otimes 2}\\ =&\begin{pmatrix}G_{n,0}(z) & G_{n,1}(z)\\ G_{n,2}(z) & G_{n,3}(z)\end{pmatrix}\end{aligned}\tag{8.3}$$

with $j = 0, 1, 2, 3, 4$

$$G_{n,j}(z) = n^{-1}\sum_{t=1}^{n}K_h(z_t-z)\Big(h^{-1}(z_t-z)\Big)^j\otimes(D_n^{-1}W_{t-1})^{\otimes 2}$$

In order to analyze $G_{n,j}(z)$, we express it as below,

$$G_{n,j}(z) = \begin{pmatrix}G_{n,j,0}(z) & G_{n,j,1}(z)\\ G_{n,j,1}(z) & G_{n,j,2}(z)\end{pmatrix}$$

where

$$G_{n,j,0}(z) = \frac{1}{n}\sum_{t=1}^{n}\Big(h^{-1}(z_t-z)\Big)^jW_{t-1,\mathcal{I}}W_{t-1,\mathcal{I}}^TK_h(z_t-z)$$

$$G_{n,j,1}(z) = \frac{1}{n}\sum_{t=1}^{n}\Big(h^{-1}(z_t - z)\Big)^j W_{t-1,\mathcal{I}} W^T_{t-1,\mathcal{I}^c} n^{-1/2} K_h(z_t - z)$$

$$G_{n,j,2}(z) = \frac{1}{n}\sum_{t=1}^{n}\Big(h^{-1}(z_t - z)\Big)^j (n^{-1/2} W_{t-1,\mathcal{I}^c})^{\otimes 2} K_h(z_t - z)$$

And

$$G^*_{n,j,1}(z) = \frac{1}{n}\sum_{t=1}^{n}\Big(h^{-1}(z_t - z)\Big)^j W_{t-1,\mathcal{I}} K_h(z_t - z)$$

$$G^*_{n,j,2}(z) = \frac{1}{n}\sum_{t=1}^{n}\Big(h^{-1}(z_t - z)\Big)^j W^{\otimes 2}_{t-1,\mathcal{I}} K_h(z_t - z)$$

By Taylor's expansion argument, we have

$$E[G^*_{n,j,1}(z)] = E[(h^{-1}(z_t - z))^j W_{t-1,\mathcal{I}} K_h(z_t - z)] = f_z(z) M_1(z) \mu_j(K) + o(1) \quad (8.4)$$

$$E[G^*_{n,j,2}(z)] = E[(h^{-1}(z_t - z))^j W^{\otimes 2}_{t-1,\mathcal{I}} K_h(z_t - z)] = f_z(z) M_2(z) \mu_j(K) + o(1) \quad (8.5)$$

We can easily show the following properties (See Theorem 1 of Cai, 2000)

$$Var[G^*_{n,j,1}(z)] = Var[G^*_{n,j,2}(z)] = O((nh)^{-1}) = o(1) \quad (8.6)$$

Therefore, by summarizing properties above

$$G^*_{n,j,l}(z) = f_z(z) M_l(z) \mu_j(K) + o_p(1),\ for\ l = 1, 2 \quad (8.7)$$

Hence,

$$G_{n,j,0}(z) = f_z(z) M_2(z) \mu_j(K) + o_p(1) \quad (8.8)$$

Then, we define $e_t = K_h(z_t - z)(h^{-1}(z_t - z))^j W_{t-1,\mathcal{I}} - E[K_h(z_t - z)(h^{-1}(z_t - z))^j W_{t-1,\mathcal{I}}]$. It's easy to obtain the following property,

$$sup_{s\geq 0} Var\Big(\sum_{t=s+1}^{s+m} e_t\Big) = O(m/h), \quad (8.9)$$

for any $m \geq 1$.

Recall that $U^*_{nt} = n^{-1/2} W_{t-1,\mathcal{I}^c}$ and $U^*_n(r) = U^*_{n,[nr]}$ for any $r \in [0,1]$. For any arbitrary small δ, $0 \leq \delta \leq 1$. Set $N = [1/\delta]$, $t_k = [kn/N]$,$t^*_k = t_{k+1} - 1$, and $t^{**}_k = min\{t^*_k, n\}$. So that,

$$\begin{aligned}\Big|\frac{1}{n}\sum_{j=1}^{[ns]} U^*_{nj} e_t\Big| =&\Big|\frac{1}{n}\sum_{k=0}^{N-1}\sum_{t=t_k}^{t^{**}_k} U^*_{nj} e_t\Big| \\ \leq&\Big|\frac{1}{n}\sum_{k=0}^{N-1} U^*_{nt_k}\sum_{t=t_k}^{t^{**}_k} e_t\Big| + \Big|\frac{1}{n}\sum_{k=0}^{N-1}\sum_{t=t_k}^{t^{**}_k}[U^*_{nj} - U^*_{nt_k}]e_t\Big| \\ \leq&\frac{1}{n}\sum_{k=0}^{N-1}|U^*_{nt_k}|\Big|\sum_{t=t_k}^{t^{**}_k} e_t\Big| + \frac{1}{n}\sum_{k=0}^{N-1}\sum_{t=t_k}^{t^{**}_k}[U^*_{nj} - U^*_{nt_k}]|e_t| \\ \leq& sup_{0\leq s\leq 1}|U^*_n(s)|\frac{1}{n}\sum_{k=0}^{N-1}\Big|\sum_{t=t_k}^{t^{**}_k} e_t\Big| + sup_{|r-s|\leq\delta}|U^*_n(r) - U^*_n(s)|\frac{1}{n}\sum_{t=1}^{n}|e_t|\end{aligned} \tag{8.10}$$

It's easy to verify that $sup_{0\leq s\leq 1}|U^*_n(s)| = O_p(1)$ and $\sum_{t=1}^n |e_t|/n = o_p(1)$, which is same proof as ... And

$$\begin{aligned}E\Big[\frac{1}{n}\sum_{k=0}^{N-1}\Big|\sum_{t=t_k}^{t^{**}_k} e_t\Big|\Big] \leq&\frac{N}{n} sup_{0\leq k\leq N-1} E\Big|\sum_{t=t_k}^{t^{**}_k} e_t\Big| \\ \leq& sup_{t\leq n} E\Big|\frac{1}{\delta n}\sum_{i=t}^{t+\delta n} e_i\Big| \leq C(\delta n h)^{-1/2} \to 0\end{aligned} \tag{8.11}$$

As $n \to 0$

$$\frac{1}{n}\sum_{j=1}^{n} U^*_{nj} e_t = o_p(1) + sup_{|r-s|\leq\delta}|U^*_n(r) - U^*_n(s)| O_p(1) \tag{8.12}$$

And as $\delta \to 0$.

$$sup_{|r-s|\leq\delta}|U^*_n(r) - U^*_n(s)| \xrightarrow{D} sup_{|r-s|\leq\delta}|\mathcal{Y}_{\eta,2}(r) - \mathcal{Y}_{\eta,2}(s)| \xrightarrow{P} 0 \tag{8.13}$$

Hence,

$$\frac{1}{n}\sum_{j=1}^{n} U_{nj}^{*} e_t = o_p(1) \tag{8.14}$$

Note that,

$$n^{-1/2}U_{[nr]} \xrightarrow{D} \mathcal{Y}_{\eta,2}(r)$$

$$\frac{1}{n}\sum_{j=1}^{n} \mathcal{E}(n^{-1/2}U_{[nr]}) \xrightarrow{D} \int_0^1 \mathcal{E}(\mathcal{Y}_{\eta,2}(s))ds \tag{8.15}$$

where $\mathcal{E}(.)$ is a Borel measurable and totally Lebesgue integrable function. So that, for $l = 1, 2$, as $n \to \infty$

$$\frac{1}{n}\sum_{j=1}^{n} (n^{-1/2}W_{t-1,\mathcal{I}^c})^{\otimes l} \xrightarrow{D} \int_0^1 (\mathcal{Y}_{\eta,2}(s))^{\otimes l} ds \equiv \mathcal{Y}_{\eta,2}^{(l)} \tag{8.16}$$

By (8.8), (8.16), and Lemma 2, we have,

$$G_{n,j,1}(z) = f_z(z)M_1(z)\mu_j(K)\mathcal{Y}_{\eta,2}^{(1)} + o_p(1) \tag{8.17}$$

Similarly

$$G_{n,j,2}(z) = f_z(z)\mu_j(K)\mathcal{Y}_{\eta,2}^{(2)} + o_p(1) \tag{8.18}$$

By plugging (8.8), (8.17) and (8.20) into $G_{n,j}(z)$,

$$G_{n,j}(z) = f_z(z)\mu_j(K)G(z) + o_p(1) \tag{8.19}$$

Given Gaussian Kernel, $\mu_0(K) = 1$, $\mu_1(K) = 0$, we get

$$G_n(z) = f_z(z)\begin{pmatrix} 1 & 0 \\ 0 & \mu_2(K) \end{pmatrix} \otimes G(z) + o_p(1) \tag{8.20}$$

Define

$$\mathcal{K}_n(z)^{-1} = f_z^{-1}(z)G^{-1}(z) + o_p(1)$$

$$\mathcal{B}_n(z) = \frac{1}{n}\sum_{t=1}^{n} K_h(z_t - z)\mathcal{A}_n^{-1} W_{t-1,\mathcal{I}} W_{t-1,\mathcal{I}}^T \times (\gamma(z_t) - \gamma(z) - (z_t - z)\gamma^{(1)}(z)),$$

Given defintion above, we have

$$D_n[\hat{\gamma}(z) - \gamma(z)] \equiv \mathcal{K}_n(z)^{-1}\mathcal{B}_n(z) + \mathcal{K}_n(z)^{-1} n^{-1}\sum_{t=1}^{n} K_h(z_t - z)\mathcal{A}_n^{-1} W_{t-1,\mathcal{I}} \tag{8.21}$$

So that,

$$\mathcal{B}_n(z) = \begin{pmatrix} \mathcal{X}_{n,0}(z) + \mathcal{X}_{n,1}(z) \\ \mathcal{X}_{n,2}(z) + \mathcal{X}_{n,3}(z) \end{pmatrix} \tag{8.22}$$

where,

$$\mathcal{X}_{n,0}(z) = \frac{1}{n}\sum_{t=1}^{n} K_h(z_t - z) W_{t-1,\mathcal{I}}^{\otimes 2} \times (\gamma_1(z_t) - \gamma_1(z) - (z_t - z)\gamma_1^{(1)}(z))$$

$$\mathcal{X}_{n,1}(z) = \frac{1}{n}\sum_{t=1}^{n} K_h(z_t - z) W_{t-1,\mathcal{I}} (W_{t-1,\mathcal{I}^c} n^{-1/2})^T \times (\gamma_2(z_t) - \gamma_2(z) - (z_t - z)\gamma_2^{(1)}(z))$$

$$\mathcal{X}_{n,2}(z) = \frac{1}{n}\sum_{t=1}^{n} K_h(z_t - z) W_{t-1,\mathcal{I}^c} W_{t-1,\mathcal{I}}^T \times (\gamma_1(z_t) - \gamma_1(z) - (z_t - z)\gamma_1^{(1)}(z))$$

$$\mathcal{X}_{n,3}(z) = \frac{1}{n}\sum_{t=1}^{n} K_h(z_t - z) (W_{t-1,\mathcal{I}} n^{-1/2})^{\otimes 2} n^{1/2} \times (\gamma_2(z_t) - \gamma_2(z) - (z_t - z)\gamma_1^{(1)}(z))$$

By Taylor expansion, it's straightforward to show following properties,

$$\mathcal{X}_{n,0}(z) = h^2 f_z(z) M_2(z)\Big[\frac{\mu_2(K)}{2}\gamma_1^{(2)}(z)\Big]\{1 + o_p(1)\}$$

$$\mathcal{X}_{n,1}(z) = h^2 f_z(z) M_1(z) \mathcal{Y}_{\eta,2}^{(1)T}\Big[\frac{\mu_2(K)}{2}\gamma_2^{(2)}(z)\Big]\{1 + o_p(1)\}$$

$$\mathcal{X}_{n,2}(z) = h^2 f_z(z) M_1(z) \mathcal{Y}_{\eta,2}^{(1)} n^{1/2}\Big[\frac{\mu_2(K)}{2}\gamma_1^{(2)}(z)\Big]\{1 + o_p(1)\}$$

$$\mathcal{X}_{n,3}(z) = h^2 f_z(z) \mathcal{Y}_{\eta,2}^{(2)} n^{1/2}\Big[\frac{\mu_2(K)}{2}\gamma_2^{(2)}(z)\Big]\{1 + o_p(1)\}$$

We can rewrite $\mathcal{B}_n(z)$ by plugging expressions above into (8.22)

$$\mathcal{B}_n(z) = h^2 f_z(z)G(z)D_n\Big[\frac{\mu_2(K)}{2}\gamma^{(2)}(z)\Big]\{1 + o_p(1)\} \tag{8.23}$$

By plugging into (8.23) into (8.21), we have,

$$\sqrt{nh}D_n[\hat{\gamma}(z) - \gamma(z) - h^2 B_\gamma(z) + o_p(h^2)] = \mathcal{K}_n(z)^{-1}\mathcal{T}_n(z) \tag{8.24}$$

where,

$$\mathcal{T}_n(z) = \begin{pmatrix} \mathcal{T}_{n,1}(z) \\ \mathcal{T}_{n,2}(z) \end{pmatrix}$$

with

$$\mathcal{T}_{n,1}(z) = \sqrt{\frac{h}{n}}\sum_{t=1}^{n} K_h(z_t - z)\epsilon_t W_{t-1,\mathcal{I}}$$

and

$$\mathcal{T}_{n,2}(z) = \sqrt{\frac{h}{n}}\sum_{t=1}^{n} K_h(z_t - z)\epsilon_t (W_{t-1,\mathcal{I}^c} n^{-1/2})$$

.

To establish asymptotic normality, we need to analyze $\mathcal{T}_{n,1}(z)$ and $\mathcal{T}_{n,2}(z)$. See Theorem 2 of Cai et al.(2000),

$$\mathcal{T}_{n,1}(z) \xrightarrow{\mathcal{D}} N(0, \sigma_\epsilon^2 \nu_0(K) f_z(z) M_2(z)) = \sqrt{\nu_0(K) f_z(z)}\mathcal{Y}_\epsilon(1) \tag{8.25}$$

$$\mathcal{T}_{n,2}(z) \xrightarrow{\mathcal{D}} \sqrt{\nu_0(K) f_z(z)} \int_0^1 \mathcal{Y}_{\eta,2}(r) d\mathcal{Y}_{\epsilon,1}(r) \tag{8.26}$$

Hence, combining (8.25) and (8.26) leads to

$$\mathcal{T}_n(z) \xrightarrow{\mathcal{D}} \sqrt{\nu_0(K) f_z(z)} \begin{pmatrix} \mathcal{Y}_\epsilon(1) \\ \int_0^1 \mathcal{Y}_{\eta,2}(r) d\mathcal{Y}_{\epsilon,1}(r) \end{pmatrix} \tag{8.27}$$

Since conditional variance of $\begin{pmatrix} \mathcal{Y}_{\epsilon}(1) \\ \int_0^1 \mathcal{Y}_{\eta,2}(r) d\mathcal{Y}_{\epsilon,1}(r) \end{pmatrix}$ is

$$\sigma_{\epsilon}^2 \begin{pmatrix} M_2(z) & M_1(z)(\mathcal{Y}_{\eta,2}^{(1)})^T \\ \mathcal{Y}_{\eta,2}^{(1)}(M_1(z))^T & \mathcal{Y}_{\eta,2}^{(2)} \end{pmatrix} = \sigma_{\epsilon}^2 G(z)$$

$$\sqrt{nh} D_n[\hat{\gamma}(z) - \gamma(z) - h^2 B_{\gamma}(z) + o_p(h^2)] \xrightarrow{\mathcal{D}} f_z(z)^{-1/2} \nu_0^{1/2} G_z^{-1} \begin{pmatrix} \mathcal{Y}_{\epsilon}(1) \\ \int_0^1 \mathcal{Y}_{\eta,2}(r) d\mathcal{Y}_{\epsilon,1}(r) \end{pmatrix} \tag{8.28}$$

Proof of Theorem 3.2 Let $S_{nj} = \sum_{t=1}^n \Omega_t^* W_{t-1,\mathcal{I}^c}^{\otimes 2} K_h(z_t - z)(\frac{z_t - z}{h_2})^j$ for $j = 0, 1, 2$. Given S_{nj}, we define $a_n = \sum_{t=1}^n \Big(K_h(z_t - z)\{S_{n2} - (\frac{z_t - z}{h}) S_{n1}\} \Omega_t^* W_{t-1,\mathcal{I}^c} \Big)^{\otimes 2}$, $b_n = \sum_{t=1}^n K_h(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2}) S_{n1}\} \Omega_t^* W_{t-1,\mathcal{I}^c}^{\otimes 2} = S_{n0} S_{n2} - S_{n1} S_{n1}$,

$$\mathcal{B}_n(z) = b_n^{-1} \Big(\sum_{t=1}^n K_h(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2}) S_{n1}\} \Omega_t^* W_{t-1,\mathcal{I}^c}^{\otimes 2} [\gamma_2(z_t) - \gamma_2(z)] \Big)$$

.

By the defintion of $\hat{\gamma_2}(z)$ below,

$$(\hat{\gamma_2}(z), \hat{\gamma_2}^{(1)}(z))^T = A_n^{-1} B_n \tag{8.29}$$

where

$$A_n = \sum_{t=1}^n \{(1, z_t - z)^\top (1, z_t - z)\} \otimes (\Omega_t^* W_{t-1,\mathcal{I}^c}^{\otimes 2}) K_h(z_t - z)$$

and

$$B_n = \sum_{t=1}^n Q_t^* y_t^* K_h(z_t - z).$$

,

$$Q_t^* = (1, z_t - z)^\top \otimes (\Omega_t^* W_{t-1,\mathcal{I}^c})$$

with

$$\Omega_t^* = diag\{w_{t,1}, \ldots, \omega_{t,d_2}\}$$

, $w_{t,i} = (1 + ||W_{t-1,\mathcal{I}^c}||^2)^{-1/2}$. We have,

$$\hat{\gamma_2}(z) - \gamma_2(z) = \mathcal{B}_n(z) + b_n^{-1} a_n^{1/2} \mathcal{V}_n(z) \tag{8.30}$$

where $\mathcal{V}_n(z) = a_n^{-1/2} \Big\{ \sum_{t=1}^n K_h(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2}) S_{n1}\} \Omega_t^* W_{t-1,\mathcal{I}^c} v_t^* \Big\}$

Under some regularity conditions, we have following multivariate cases property. See Theorem 14.1 and 19.2 of Billingsley(1999) for iid η_s and ρ-mixing η_s. $[nr]$ denotes the integer part of nr and $r = t/n$, as $n \to \infty$

$$\frac{U_{[nr]}}{\sqrt{n}} \xrightarrow{D} \mathcal{Y}_\eta(r) \tag{8.31}$$

where $\mathcal{Y}_\eta(r)$ is a m-dimensional Brownian motion on $[0, 1]$ with covariance matrix Υ_η. "$\xrightarrow{D}$" represents convergence in distribution. m is the dimension of $W_{t-1,\mathcal{I}^c}$

To establish asymptotic normality, it's necessary to analyze S_{nj}, $\mathcal{B}_n(z)$ as wells as $\mathcal{V}_n$. Assume that $\Omega_t^* W_{t-1,\mathcal{I}^c} \xrightarrow{D} \frac{\mathcal{B}(r)}{||\mathcal{B}(r)||}$ as $n \to \infty$ and z_t is stationary. By combing (8.17), Lemma A.1. and (8.4),

$$\begin{aligned} S_{nj} =& \sum_{t=1}^n \Omega_t^* W_{t-1,\mathcal{I}^c}^{\otimes 2} K_h(z_t - z)(\frac{z_t - z}{h_2})^j \\ =& n^{1/2} \sum_{t=1}^n (\Omega_t^* W_{t-1,\mathcal{I}^c})(n^{-1/2} W_{t-1,\mathcal{I}^c}) K_h(z_t - z)(\frac{z_t - z}{h_2})^j \\ =& E[K_h(z_t - z)(\frac{z_t - z}{h_2})^j (\Omega_t^* W_{t-1,\mathcal{I}^c})] \frac{1}{n} \sum_{j=1}^n U_{nj}^* + \frac{1}{n} \sum_{j=1}^n U_{nj}^* e_t \\ =& f_z(z) M_0(z) \mu_j(K) \mathcal{Y}_{\eta,2}^{(1)} + o_p(1) \end{aligned} \tag{8.32}$$

where $\mathcal{Y}_{\eta,2}^{(l)} = \int_0^1 \mathcal{Y}_{\eta,2}^{\otimes l}(r)dr$,

$$\mathcal{B}(r) = \int_0^1 \{U_\gamma^*(r) - U_\gamma^*(s)\}ds$$

and $U_\gamma^*(r) = \{U_{\gamma_1}^*(r), U_{\gamma_2}^*(r), \ldots, U_{\gamma_d}^*(r)\}$

By using (8.32)and Taylor's expansion argument, we can rewrite $Bias_n(z)$ into following expressions,

$$\begin{aligned}
\mathcal{B}_n(z) &= b_n^{-1}\Big(\sum_{t=1}^n K_h(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2})S_{n1}\}\Omega_t^* W_{t-1,\mathcal{I}^c}^{\otimes 2}[\gamma_2(z_t) - \gamma_2(z)]\Big) \\
&= (S_{n0}S_{n2} - S_{n1}S_{n1})^{-1} \\
&\quad \times \Big\{\sum_{t=1}^n K_h(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2})S_{n1}\}\Omega_t^* W_{t-1,\mathcal{I}^c}^{\otimes 2}[\gamma_2(z_t) - \gamma_2(z)]\Big\} \\
&= (S_{n0}S_{n2} - S_{n1}S_{n1})^{-1}(S_{n1}S_{n2} - S_{n1}S_{n2})\gamma^{(1)}(z)h_2 \\
&\quad + (S_{n0}S_{n2} - S_{n1}S_{n1})^{-1}(S_{n2}S_{n2} - S_{n1}S_{n3})\gamma^{(2)}(z)\frac{h_2^2}{2} + O_p(h_2^3) \\
&= (S_{n0}S_{n2} - S_{n1}S_{n1})^{-1}(S_{n2}S_{n2} - S_{n1}S_{n3})\gamma^{(2)}(z)\frac{h_2^2}{2} + O_p(h_2^3) \\
&= \frac{\mu_2^2(K) - \mu_1(K)\mu_3(K)}{\mu_2(K)\mu_0(K) - \mu_1^2(K)}\gamma^{(2)}(z)\frac{h_2^2}{2} + O_p(h_2^3) \\
&= \mu_2^2(K)\gamma^{(2)}(z)\frac{h_2^2}{2} + O_p(h_2^3)
\end{aligned} \tag{8.33}$$

Note that

$$\begin{aligned}
\mathcal{V}_n &= a_n^{-1/2}\Big\{\sum_{t=1}^n K_h(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2})S_{n1}\}\Omega_t^* W_{t-1,\mathcal{I}^c} v_t^*\Big\} \\
&= a_n^{-1/2}\Big(\Omega_1^{**}U_1, \Omega_2^{**}U_2, \ldots, \Omega_n^{**}U_n\Big)(v_1^*, v_2^*, \ldots, v_n^*)^T \\
&= a_n^{-1/2} c_n v^{**}
\end{aligned} \tag{8.34}$$

where $\Omega_t^{**} = K_h(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2})S_{n1}\}\Omega_t^*$, $c_n = \Big(\Omega_1^{**}U_1, \Omega_2^{**}U_2, \ldots, \Omega_n^{**}U_n\Big)$ and $v^{**} = (v_1^*, v_2^*, \ldots, v_n^*)^T$.

It's easy to establish that

$$E(v^{**}v^{**T}|\mathcal{F}) = \sigma_v^2 I_n\{1+o(1)\} \tag{8.35}$$

See (B.11) and (B.21) of Fan and Jiang(2005).

Therefore, it is easy to show that

$$E(\mathcal{V}_n|\mathcal{F}) = 0 \tag{8.36}$$

and

$$\begin{aligned} E(\mathcal{V}_n^{\otimes 2}|\mathcal{F}) &= \sigma_v^2 a_n^{-1/2} \sum_{t=1}^{n} \Big(K_h(z_t - z)\{S_{n2} - (\frac{z_t - z}{h_2})S_{n1}\}\Omega_t^* W_{t-1,\mathcal{I}^c} \Big)^{\otimes 2} a_n^{-1/2}\{1+o(1)\} \\ &= \sigma_v^2 I_k + o(1) \end{aligned} \tag{8.37}$$

Since $\mathcal{V}_n = a_n^{-1/2} c_n v^{**} \equiv p_n^T v^{**} \equiv \sum_{t=1}^n p_{n_i} v_i^{**}$, where $p_n = (p_{n_1}, p_{n_2}, \ldots, p_{n_n})^t$. By martingale limit theorem(MLT) (Hall and Heyde, 1980), it's easy to show that for any $k \times 1$ vector $\mathcal{R}$,

$$\mathcal{R}^T \mathcal{V}_n \xrightarrow{D} N(0, \sigma_v^2 \mathcal{R}^T \mathcal{R}) \tag{8.38}$$

By the Wald device,

$$\mathcal{V}_n \xrightarrow{D} N(0, \sigma_v^2 I_k)$$

By applying the Slutsky theorem,

$$\mathcal{V}_n(z) = a_n^{-1/2} b_n \Big(\hat{\gamma}_2(z) - \gamma_2(z) - \mathcal{B}_n(z) \Big) \rightarrow N(0, \sigma_v^2 I_k) \tag{8.39}$$

By combining (8.39) and (8.33), we finished proof of Theorem 3.2

APPENDIX 9: PROOF OF THEOREM IN SECTION 3.3

Proof of Theorem 3.3 By Theorem 3.1, 3.2, we have following Bahadur representations of proposed estimators,

$$\sqrt{nh_1}[\hat{\gamma}_{\mathcal{I}}(z) - \gamma_{\mathcal{I}}(z) - \frac{1}{2}h_1^2\mu_2(K)\gamma_{\mathcal{I}}^{(2)}(z)] \\ = \sqrt{nh_1}M_2^{-1}(z)\sum_{t=1}^{n} n^{-1}W_{t-1,\mathcal{I}}v_t K_{h_1}(z_t - z) + o_p(1) \tag{9.1}$$

and

$$a_n^{-1/2}b_n\Big(\hat{\gamma}_2(z) - \gamma_2(z) - \frac{1}{2}h_2^2\mu_2(K)\gamma_2^{(2)}(z)\Big) = a_n^{-1/2}c_n v^{**}1 + o_p(1) \tag{9.2}$$

where $M_2(z) = E[W_{t-1,\mathcal{I}}^{\otimes 2}|z_t = z]$, $\Omega_t^{**} = K_{h_2}(z_t - z)\{S_{n2} - (\frac{z_t-z}{h_2})S_{n1}\}\Omega_t^*$, $c_n = \Big(\Omega_1^{**}W_{1,\mathcal{I}^c}, \Omega_2^{**}W_{2,\mathcal{I}^c}, \ldots, \Omega_n^{**}W_{n,\mathcal{I}^c}\Big)$ and $v^{**} = (v_1^*, v_2^*, \ldots, v_n^*)^T$. $a_n = \frac{1}{n}\sum_{t=1}^n \Big(K_{h_2}(z_t - z)\{S_{n2} - (\frac{z_t-z}{h})S_{n1}\}\Omega_t^* W_{t-1,\mathcal{I}^c}\Big)^{\otimes 2}$; $b_n = \sum_{t=1}^n K_{h_2}(z_t - z)\{S_{n2} - (\frac{z_t-z}{h_2})S_{n1}\}\Omega_t^* W_{t-1,\mathcal{I}^c}^{\otimes 2} = S_{n0}S_{n2} - S_{n1}S_{n1}$;

Firstly, we define $\psi_n = \sqrt{nh_1}M_2^{-1}(z)\sum_{t=1}^n n^{-1}W_{t-1,\mathcal{I}}K_{h_1}(z_t - z)v_t$; $\mathcal{V}_n = a_n^{-1/2}c_n v^{**}$. Given $v_t^* = v_t - W_{t-1,1}^T[\hat{\gamma}_1(z_t) - \gamma_1(z_t)]$, we can rewrite ψ_n and $mathcalV_n$ into following forms.

$$\psi_n = \sqrt{nh_1}M_2^{-1}(z)\frac{1}{n}\{W_{1,\mathcal{I}}K_{h_1}(z_1 - z), \ldots, W_{n,\mathcal{I}}K_{h_1}(z_n - z)\}v \tag{9.3}$$

$$\mathcal{V}_n = a_n^{-1/2}c_n(I_n - \mathcal{S})v \tag{9.4}$$

where $v = (v_1, \ldots, v_n)^T$ and $\mathcal{S} = \Big(s_{i,j}\Big)$, $s_{i,j} = \frac{1}{n}W_{i-1,\mathcal{I}}^T M_2^{-1}(z)W_{j-1,\mathcal{I}}K_{h_1}(z_j - z)$.

$$Cov(\psi_n, \mathcal{V}_n) = E[\psi_n\mathcal{V}_n^T] - E[\psi_n]E[\mathcal{V}_n^T]$$

Since $E[\psi_{n1}] = 0$ and $E[\psi_{n2}] = 0$, by (A.9), (A.11), (A.12) of (Cai and Li, 2009) and

Lemma(A.1), Lemma(A.2) we have

$$\begin{aligned}
Cov(\psi_n, \mathcal{V}_n) =& E[\psi_n \mathcal{V}_n^T] \\
=& E[\sqrt{nh_1} M_2^{-1}(z) \frac{1}{n} \{W_{1,\mathcal{I}} K_{h_1}(z_1 - z), \ldots, W_{n,\mathcal{I}} K_{h_1}(z_n - z)\} \\
& (I_n - \mathcal{S})^T c_n^T (a_n^{-1/2})^T] \sigma_v^2 \\
=& E[\sqrt{nh_1} M_2^{-1}(z) \frac{1}{n} \{W_{1,\mathcal{I}} K_{h_1}(z_1 - z), \ldots, W_{n,\mathcal{I}} K_{h_1}(z_n - z)\} \\
& c_n^T (a_n^{-1/2})^T] \sigma_v^2 \\
& - E[\sqrt{nh_1} M_2^{-1}(z) \frac{1}{n} \{W_{1,\mathcal{I}} K_{h_1}(z_1 - z), \ldots, W_{n,\mathcal{I}} K_{h_1}(z_n - z)\} \mathcal{S}^T \\
& c_n^T (a_n^{-1/2})^T] \sigma_v^2 \\
=& lim_{n \to \infty} \sum_{t=1}^{n} \sqrt{h_1/n} M_2^{-1}(z) W_{t-1,\mathcal{I}} (I_n - \mathcal{S})^T c_n^T (a_n^{-1/2})^T \sigma_v^2 = \Sigma_{1,2} \sigma_v^2
\end{aligned} \tag{9.5}$$

By combining (9.5), Theorem 3.1 and Theorem 3.2, we have

$$\begin{pmatrix} \sqrt{nh_1} & 0 \\ 0 & a_n^{-1/2} b_n \end{pmatrix} \left[\hat{\gamma^*}(z) - \gamma^*(z) - \frac{1}{2} \begin{pmatrix} h_1^2 \mu_2(K) (\gamma_1^{(2)}(z))^T \\ h_2^2 \mu_2(K) (\gamma_2^{(2)}(z))^T \end{pmatrix} \right] \xrightarrow{D} MN(\Sigma_\gamma(z))$$

where $\gamma^*(z) = \left(\gamma_1^T(z), \gamma_2^T(z)\right)^T$ and $\Sigma_{\gamma^*}(z)$ is the covariance matrix given by

$$\Sigma_{\gamma^*}(z) = \begin{pmatrix} \nu_0(K) M_2(z)^{-1} / f_z(z) & \Sigma_{1,2} \\ \Sigma_{2,1} & I_{d_2} \end{pmatrix} \sigma_v^2$$

We finished proof of Theorem 3.3.

APPENDIX 10: PROOF OF THEOREM IN SECTION 4.1

Lemma C.1. Without loss of generality, assume that $W_{t-1,\mathcal{I}} = (X_{t,1}^T, X_{t,2}^T)^T$, where $X_{t,2}$ is a $d\times 1$ vector of $NI(1)$ or $I(1)$ processes and $X_{t,1}$ is a $(k-d)\times 1$ vector of stationary or $I(0)$ processes. Let $\alpha(z) = E(X_{t,2}|z_t = z)$, $U_\gamma^*(r) = \{U_{\gamma_1}^*(r), U_{\gamma_2}^*(r), \ldots, U_{\gamma_d}^*(r)\}$, $\mathcal{B}(r) = \int_0^1 (U_\gamma^*(r) - U_\gamma^*(s))ds$, and $\mathcal{B}^*(r) \equiv \mathcal{B}(r)/||\mathcal{B}(r)||$. Assume some regularity conditions and $E|v_t^3| < \infty$. Then we have following results with probability goes to one.

(i) $V_n \equiv \frac{1}{n}\sum_{t=1}^n Z_t^{\otimes 2}(\gamma(z)) \xrightarrow{p} V$, where $V = diag(V_1, V_2)$, $V_1 = \sigma_v^2 E(X_{t,1} - \alpha(z_t))^{\otimes 2}$ and $V_2 = \sigma_v^2 \int_0^1 \mathcal{B}^*(r)^{\otimes 2} dr$.

(ii)$Z_n^* = max_{1\le t\le n}||Z_t(\gamma(z)|| = o_p((nh)^{1/2})$.

(iii)if $nh^5 = O(1)$, then $||\bar{Z}|| = ||(nh)^{-1}\sum_{t=1}^n Z_t(\gamma(z))|| = O_p((nh)^{-1/2})$, where $||\bar{Z}|| = (nh)^{-1}\sum_{t=1}^n Z_t(\gamma(z))$

(iv)$(nh)^{-1}\sum_{t=1}^n ||Z_t(\gamma(z))||^3 = o_p((nh)^{1/2})$

Proof. of Lemma C.1 For $i = 1, \ldots, d$, $x_{t,i}$ is $NI(1)$ or $I(1)$. For $\frac{t-1}{n} \le r \le \frac{t}{n}$define $U_{n,i}(r) = U_{ni,t} = n^{-1/2}x_{t-1,i}$. Then

$$\begin{aligned}\sum_{s=1}^n \phi_s(z)x_{s-1,i} =& n^{1/2}\{\sum_{s=1}^n \omega_t(z)\}^{-1}\sum_{s=1}^n K_h(z_s - z)\{S_{n,2}(z) - h^{-1}(z_s - z)S_{n,1}(z)\}U_{ni,s} \\ =& \{\sum_{s=1}^n \{n^{-1}\omega_t(z)\}^{-1}S_{n,2}F_{n,0}(z) - S_{n,1}(z)F_{n,.1}(z)\end{aligned} \tag{10.1}$$

where for $j = 0, 1, F_{n,j}(z) = n^{-1}\sum_{s=1}^n K_h^{(j)}(z_s - z)U_{ni,s}$, with $K_h^{(j)}(z_s - z) = h^{-j}(z_s - z)K_h(z_s - z)$. Using the same argument $F_{n,j}$ in equation (A.11) of Cai, Li and Park (2009), we have $F_{n,j}(z) = f(z)\mu_j W_\mu^{(i)} + o_p(1)$, where $W_\mu^{(i)} = \int_0^1 U_{\gamma_i}(r)dr$.

By definition of $Z_t(\gamma(z))$ and (1.1), we have

$$Z_t(\gamma(z)) = \Big(Z_{1t}(\gamma(z)), Z_{2t}(\gamma(z))\Big) \tag{10.2}$$

$$\begin{aligned} Z_{1t}(\gamma(z)) =& K_{h_1}(z_t - z)\{T_{n2} - (z_t - z)T_{n1}\}W_{t-1,\mathcal{I}}\{y_t - W_{t-1,\mathcal{I}}^\top \gamma_1(z) - W_{t-1,\mathcal{I}^c}^\top \gamma_2(z)\} \\ =& K_{h_1}(z_t - z)\{T_{n2} - (z_t - z)T_{n1}\}W_{t-1,\mathcal{I}}v_t \end{aligned} \tag{10.3}$$

and

$$\begin{aligned} Z_{2t}(\gamma(z)) =& K_{h_2}(z_t - z)\{S_{n2} - (z_t - z)S_{n1}\}\Omega_t^* W_{t-1,\mathcal{I}^c}\{y_t - W_{t-1,\mathcal{I}}^\top \gamma_1(z) - W_{t-1,\mathcal{I}^c}^\top \gamma_2(z) \\ =& K_{h_2}(z_t - z)\{S_{n2} - (z_t - z)S_{n1}\}\Omega_t^* W_{t-1,\mathcal{I}^c}v_t \end{aligned} \tag{10.4}$$

Then

$$\begin{aligned} Z_t^{\otimes 2}(\gamma(z)) =& \Big(Z_{1t}(\gamma(z)), Z_{2t}(\gamma(z))\Big)^{\otimes 2} \\ =& \Big(\phi_{1t}v_t, \phi_{1t}v_t\Big)^{\otimes 2} \\ =& \begin{pmatrix} \phi_{1t}^{\otimes 2}v_t^2 & \phi_{1t}\phi_{2t}v_t^2 \\ \phi_{2t}\phi_{1t}v_t^2 & \phi_{2t}^{\otimes 2}v_t^2 \end{pmatrix} \end{aligned} \tag{10.5}$$

where $\phi_{1t} = \Big(K_{h_1}(z_t - z)\{T_{n2} - (z_t - z)T_{n1}\}W_{t-1,\mathcal{I}}$; $\phi_{2t} = K_{h_2}(z_t - z)\{S_{n2} - (z_t - z)S_{n1}\}\Omega_t^* W_{t-1,\mathcal{I}^c}$.

By using the similar arguments (8.10), (A.15) and (A.20) in (Hong and Jiang, 2018), we obtain the following results.

$$\begin{aligned} V_n \equiv& \frac{1}{n}\sum_{t=1}^n Z_t^{\otimes 2}(\gamma(z)) \\ =& \frac{1}{n}\sum_{t=1}^n \begin{pmatrix} \phi_{1t}^{\otimes 2}v_t^2 & \phi_{1t}\phi_{2t}v_t^2 \\ \phi_{2t}\phi_{1t}v_t^2 & \phi_{2t}^{\otimes 2}v_t^2 \end{pmatrix} \xrightarrow{p} V, \end{aligned} \tag{10.6}$$

where $V = diag(V_1, V_2)$, where $V_1 = \sigma_v^2 E(X_{t,1} - \alpha(z_t))^{\otimes 2}$ and $V_2 = \sigma_v^2 \int_0^1 \mathcal{B}^*(r)^{\otimes 2} dr$ We complete the proof for Lemma C.1 (i) .

(ii) Note that $Z_t(\gamma(z)) = \Big(Z_{1t}(\gamma(z)), Z_{2t}(\gamma(z))\Big)$; $\phi_{1t} = \Big(K_{h_1}(z_t - z)\{T_{n2} - (z_t -$

$z)T_{n1}\}W_{t-1,\mathcal{I}}$; $\phi_{2t} = K_{h_2}(z_t - z)\{S_{n2} - (z_t - z)S_{n1}\}\Omega_t^* W_{t-1,\mathcal{I}^c}$. it suffices to show that

$$\eta_{n,j} \equiv (nh_j)^{-1/2} max_{1\leq t\leq n}||\phi_{jt}|| = o_p(1)$$

and

$$\eta^*_{n,j} \equiv (nh_j)^{-1/2} max_{1\leq t\leq n}||\phi_{jt}v_t|| = o_p(1)$$

For $j = 1$, since $lim_{n\to\infty} P(\Omega_t = I_{d_2}) = 1$ using the (A.10) in (Hong and Jiang, 2017), we obtain the following

$$\eta_{n,1} \leq (nh_1)^{-1/2} max_{1\leq t\leq n}||\phi_{1t}|| = o_p(1)$$

and

$$\eta^*_{n,1} \leq (nh_1)^{-1/2} max_{1\leq t\leq n}||\phi_{1t}v_t|| = o_p(1)$$

For $j = 2$ with probability tending to one, that $\eta_{n,2} \leq (nh_2)^{-1/2} max_{1\leq t\leq n} = o(1)$ and $\eta^*_{n,2} \leq (nh_2)^{-1/2} max_{1\leq t\leq n}|v_t| = o_p(1)$. We proved $Z^*_n = max_{1\leq t\leq n}||Z_t(\gamma(z)|| = o_p((nh)^{1/2})$. $h = min\{h_1, h_2\}$.

iii) By definition, we have

$$\begin{aligned}(nh)^{1/2}\bar{Z} &= (nh)^{-1/2}\sum_{t=1}^{n} Z_t(\gamma(z)) \\ &= (nh)^{-1/2}\sum_{t=1}^{n}\Big(Z_{1t}(\gamma(z)), Z_{2t}(\gamma(z))\Big) \\ &= (nh)^{-1/2}\sum_{t=1}^{n}\Big(\phi_{1t}v_t, \phi_{2t}v_t\Big) \equiv M_n\end{aligned}$$

Note that $E(M_n) = 0$ and $Var(M_n) = \sigma_v^2 E\{\Big(\phi_{1t}v_t, \phi_{2t}v_t\Big)^{\otimes 2}\} = O(1)$ Then $M_n = O_p(1)$. Hence $(nh)^{1/2}\bar{Z} = O_p(1)$, $\bar{Z} = O_p((nh)^{-1/2})$

iv) Since $||\phi_{2t}|| \leq \sqrt{d1}$ and $\phi_{1t} = X_t$ with probability tending to one and $E|v_t^3 < \infty$ we have $(nh)^{-1}\sum_{t=1}^{n}||Z_t(\gamma(z))||^3 = o_p((nh)^{1/2})$.

Proof of Theorem 4.1. Note that $p_t = \frac{1}{n}\frac{1}{1+\lambda^\top Z_t(\gamma_{\mathcal{I}}(z))}$, where λ satisfies that

$$g(\lambda) \equiv n^{-1}\sum_{t=1}^{n}\frac{Z_t(\gamma(z))}{1+\lambda^\top Z_t(\gamma(z))} = 0$$

It follows that,

$$\begin{aligned} g(\lambda) &= n^{-1}\sum_{t=1}^{n} Z_t(\gamma(z))[1 - \frac{\lambda^\top Z_t(\gamma(z))}{1+\lambda^\top Z_t(\gamma(z))}] \\ &= n^{-1}\sum_{t=1}^{n} Z_t(\gamma(z)) - n^{-1}\sum_{t=1}^{n}\frac{Z_t^{\otimes 2}(\gamma(z))}{1+\lambda^\top Z_t(\gamma(z))}\lambda \\ &= \bar{Z} - \tilde{V}_n\lambda \\ &= 0 \end{aligned}$$

Hence,

$$\bar{Z} = \tilde{V}_n\lambda \tag{10.7}$$

Since every $p_t > 0$, we have $1 + \lambda^\top Z_t(\gamma(z)) > 0$ and therefore,

$$\begin{aligned} \lambda^\top V_n\lambda &= \lambda^\top n^{-1}\sum_{t=1}^{n}\frac{Z_t^{\otimes 2}(\gamma(z))}{1+\lambda^\top Z_t(\gamma_{\mathcal{I}}(z))}[1 + \lambda^\top Z_t(\gamma(z))]\lambda \\ &\leq \lambda^\top n^{-1}\sum_{t=1}^{n}\frac{Z_t^{\otimes 2}(\gamma(z))}{1+\lambda^\top Z_t(\gamma(z))}\{1 + ||\lambda|| max_{1\leq t\leq n}||Z_t(\gamma(z))||\}\lambda \\ &= \lambda^\top\tilde{V}_n\lambda(1 + ||\lambda||Z_n^*) \end{aligned} \tag{10.8}$$

Where $V_n = n^{-1}\sum_{t=1}^{n} Z_t^{\otimes 2}(\gamma(z))$ and $Z_n^* = max_{1\leq t\leq n}||Z_t(\gamma_{\mathcal{I}}(z)||$.

Let $\lambda = \rho\eta$ and $\eta \in \mathcal{R}^K$, $||\eta|| = 1$.

Then $||\eta|| = \rho$. Given(10.7) and (10.8), we can obtain the following results,

$$0 \leq \rho\eta^T V_n\eta \leq \rho\eta^T\tilde{V}_n\eta(1 + \rho Z_n^*) = \eta^T\bar{Z}(1 + \rho Z_n^*)$$

By Lemma C.1(i), we have follows,

$$\begin{aligned}\eta^T \bar{Z} &\geq \frac{\rho}{1+\rho Z_n^*}\eta^T V_n \eta \\ &= \frac{\rho}{1+\rho Z_n^*}(\eta^T V \eta + o_p(1))\end{aligned}$$

Therefore,

$$\rho[min\, eig(V) + o_p(1) - \eta^{'} \bar{Z} Z_n^*] \leq \eta^{'} \bar{Z} \tag{10.9}$$

By Lemma C.1(ii), $Z_n^* = o_p((nh)^{1/2})$ and $||\bar{Z}|| = O_p((nh)^{-1/2})$.We have follows with probability goes to one,

$$||\lambda|| = \rho = O_p((nh)^{-1/2})$$

and

$$max_{1\leq t\leq n}|\lambda Z_t(\gamma_{\mathcal{I}}(z))| \leq ||\lambda|| Z_n^* = O_p((nh)^{-1/2}) \cdot o_p((nh)^{-1/2}) = o_p(1)$$

Given results above, we can rewrite $g(\lambda)$ as follows,

$$\begin{aligned}g(\lambda) &= n^{-1}\sum_{t=1}^{n} Z_t(\gamma(z))[1 - \lambda^\top Z_t(\gamma(z)) + \frac{\lambda^\top Z_t^{\otimes 2}(\gamma(z))\lambda}{1+\lambda^\top Z_t(\gamma(z))}] \\ &= \bar{Z} - V_n \lambda + n^{-1}\sum_{t=1}^{n} \frac{\lambda^\top Z_t^{\otimes 2}(\gamma(z))\lambda Z_t(\gamma(z))}{1+\lambda^\top Z_t(\gamma(z))} \\ &= 0\end{aligned} \tag{10.10}$$

By Lemma C.1 and (10.9), the last term in (10.10) is bounded by,

$$\begin{aligned}||\lambda||^2 n^{-1}\sum_{t=1}^{n}\frac{||Z_t(\gamma(z)||^3}{1-||\lambda||Z_n^*} &= O_p((nh)^{-1})\{1+o_p(1)\}n^{-1}\sum_{t=1}^{n}||Z_t(\gamma(z)||^3 \\ &= O_p((nh)^{-1})\{1+o_p(1)\}o_p((nh)^{1/2}) \\ &= o_p((nh)^{-1/2})\end{aligned}$$

Hence, by (10.10), we have follows,

$$\lambda = V_n^{-1}\bar{Z} + o_p((nh)^{-1/2}) \tag{10.11}$$

Since that $max_{1\le t\le n}|\lambda Z_t(\gamma(z))| = o_p(1)$. By Taylor's expansion, we obtain the follows,

$$log\{1+\lambda^\top Z_t(\gamma(z))\} = \lambda^\top Z_t(\gamma(z)) - \frac{1}{2}\{\lambda^\top Z_t(\gamma(z))\}^2 + \phi_t \tag{10.12}$$

where for some finite $M > 0$,

$$P\{|\phi_t| \le M|\lambda^\top Z_t(\gamma(z))|^3, 1 \le t \le n\} \to 1, as\ \ n \to \infty$$

It follows from (10.12) that

$$\begin{aligned}\ell_n(\gamma(z)) &= 2\sum_{t=1}^{n}\log\{1+\lambda^\top Z_t(\gamma(z)\} \\ &= 2\sum_{t=1}^{n}\lambda^\top Z_t(\gamma(z)) - \sum_{t=1}^{n}\lambda^\top Z_t^{\otimes 2}(\gamma(z))\lambda + 2\sum_{t=1}^{n}\phi_t\end{aligned}$$

By combining results above, we have

$$\begin{aligned}|\sum_{t=1}^{n}\phi_t| &\le M||\lambda||^3\sum_{t=1}^{n}||Z_t(\gamma(z))||^3 \\ &= O_p((nh)^{-3/2})\cdot o_p((nh)^{3/2}) \\ &= o_p(1)\end{aligned} \tag{10.13}$$

Hence,

$$\ell_n(\gamma(z)) = 2\sum_{t=1}^{n}\lambda^\top Z_t(\gamma(z)) - \sum_{t=1}^{n}\lambda^\top Z_t^{\otimes 2}(\gamma(z))\lambda + o_p(1) \tag{10.14}$$

Then by (10.10) and (10.11),

$$\ell_n(\gamma(z)) = n\bar{Z}^\top V_n^{-1}\bar{Z} + o_p(1). \tag{10.15}$$

Applying the CLT, the following result is obtained,

$$\begin{aligned} n^{1/2} V_n^{-1/2} \bar{Z} &= V_n^{-1/2} n^{-1/2} \sum_{t=1}^{n} Z_t(\gamma(z)) \\ &\to N(0, I_{k+1}) \end{aligned} \tag{10.16}$$

where I_{k+1} is the $(k+1) \times (k+1)$ identity matrix. Therefore, $\ell_n(\gamma(z)) \xrightarrow{p} \chi^2_{k+1}$.