NONPARAMETRIC PREDICTIVE REGRESSION

by

Xintian Yu

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

2016

Approved by:

________________________________________
Dr. Jiancheng Jiang

________________________________________
Dr. Yanqing Sun

________________________________________
Dr. Zhiyi Zhang

________________________________________
Dr. Juntao Guo
ABSTRACT

XINTIAN YU. Nonparametric Predictive Regression. (Under the direction of DR. JIANCHENG JIANG)

In financial time series nonlinear effects and time-varying effects are observed. In this dissertation we propose a predictive regression model with time varying coefficients and functional coefficients. It allows for nonstationary predictors. We establish asymptotic for the coefficient estimation and show oracle properties of the resulting estimators under stationary and nonstationary settings. Simulations demonstrate good finite sample performance of our estimators. A real example illustrates the use of our methodology.

In this dissertation, we investigate the generalized semiparametric varying-coefficient models for longitudinal data that can flexibly model three types of covariate effects: time-constant effects, time-varying effects, and covariate-varying effects, i.e., the covariate effects that depend on other possibly time-dependent exposure variables.

First, we consider the model that assumes the time-varying effects are unspecified functions of time while the covariate-varying effects are parametric functions of an exposure variable specified up to a finite number of unknown parameters. Second, we consider the model in which both time-varying effects and covariate-varying effects are completely unspecified functions. The estimation procedures are developed using multivariate local linear smoothing and generalized weighted least squares estimation techniques. The asymptotic properties of the proposed estimators are established. The simulation studies show that the proposed methods have satisfactory finite sam-
ple performance. ACTG 244 clinical trial of HIV infected patients are applied to examine the effects of antiretroviral treatment switching before and after HIV developing the 215-mutation. Our analysis shows benefit of treatment switching before developing the 215-mutation.

The proposed methods are also applied to the STEP study with MITT cases showing that they have broad applications in medical research.
ACKNOWLEDGMENTS

Upon the completion of this thesis I would sincerely gratefully express my thanks to many people. I would like to give my deepest gratitude to my dissertation advisor, Dr. Jiancheng Jiang for his guidance, insights and encouragement throughout my dissertation research process. His attitude to work and to life deeply engraved in my heart and memory.

I would also like to thank my committee members, Dr. Yanqing Sun and Dr. Zhiyi Zhang for their friendly guidance and thought provoking suggestions. I would like to thank Dr. Juntao Guo for agreeing to serve on my doctoral dissertation committee as an outside department committee member.

I would also like to thank Graduate School and Mathematics department for providing me financial support. I also need to offer special appreciation to Professor Joel Avrin and Shaozhong Deng, who served as graduate coordinator during my time in this program. Their insightful suggestions enabled me to successfully finish this program.

I dedicate my dissertation to my parents, Cunsheng Yu and Guangying Zhu, for their patience, generosity and sacrifice along my growing path.

Finally, I would like to thank my wife for the love, and support she has given over the years. I undoubtedly could not have done this without her support and motivation.
TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF TABLES ix

CHAPTER 1: INTRODUCTION 1

CHAPTER 2: MODEL WITH TIME VARYING AND NONLINEAR EFFECTS 5

CHAPTER 3: ESTIMATION 6

3.1. Orthogonal series estimation 6

3.2. Local smoother 8

CHAPTER 4: MODELS WITH STATIONARITY $X_I$ 10

4.1. Notations and conditions 10

4.2. Asymptotics 13

CHAPTER 5: MODEL WITH NONSTATIONARITY $X_I$ AND STATIONARY $Z_I$ 14

5.1. Notations and conditions 14

5.2. Asymptotics 17

CHAPTER 6: SIMULATIONS 19

6.1. Notations and conditions 19

6.2. Stationary $x_i$ 20

CHAPTER 7: REAL EXAMPLE 23

CHAPTER 8: DISCUSSION 30

REFERENCES 31

APPENDIX A: Sketch of Proofs 33
APPENDIX B: Sketch of Proofs
LIST OF FIGURES

FIGURE 1: Estimated function $\beta$ and $\gamma$, their mediums and 95% pointwise confidence intervals for Model 1.3 with $x$ is stationary and $n = 100$ 21

FIGURE 2: Estimated function $\beta$ and $\gamma$, their mediums and 95% pointwise confidence intervals for Model 1.3 with $x$ is stationary and $n = 400$ 22

FIGURE 3: Estimated functions from model 1.3 25

FIGURE 4: Estimated stock price and function of coefficient from model 1.3 26

FIGURE 5: Results from model 1.1 27

FIGURE 6: Results from model 1.2 28

FIGURE 7: $\hat{y}$ and One step forecast 1.2 29
TABLE 1: ADF test for stock price of Morgan Stanley, price of SP500 and 5 year daily Treasury yield rate
CHAPTER 1: INTRODUCTION

Nonlinear effects and time-varying effects exists widely in financial markets. For example, for the capital asset pricing model (CAPM) [see the books by Cochrane (2001) and TSAY (2002) for details], Blume (1975) suggested that beta coefficient changes over time, and Fabozzi (1978) revealed that many stocks’ beta coefficients move randomly through time rather than remain stable. The nonstationarity of beta and the time-varying behavior of equity return co-movements may exist, see Blume (1981), McDonald (1985), Lee (1986), Levy (1971), Rosenberg (1985), Kaplanis (1988), and Koch (1991). Another example is for the relationship between the electricity demand and other variables such as the income or production, the real price of electricity, and the temperature. Chang (2003) found that this relationship may change over time. These motivates us to consider the following time-varying coefficient model,

$$Y_t = \beta(t)^\top X_t + \epsilon_t,$$  \hspace{1cm} (1.1)

to fit financial data.

Horowitz (2004) considered the nonparametric estimation of an additive model with a link function, they derived estimators with oracle property by a two-step estimation method. They used basis functions to estimate the unknown link function in the first step. But their result was limited on additive model and $I(0)$ process.

It is well known that most data in financial industry is nonstationary. People are
interested in how models could be built in those nonstationary data. Granger (1981) and Engle (1987) introduced cointegration models in 1990s, which are built on nonstationary $x$ and nonstationary $y$. Cointegration models have attracted a amount of research attention in econometrics since then. The concept of cointegration provides an attractive and appealing characterization theoretically, but there is only a few evidences of cointegration found in empirical applications. This empirical consequence is probably due to constant parameters. The cointegrating parameters are constant in the cointegration model introduced by Engle (1987). However, cointegration characterizes long-term equilibrium relationships, and the exact quantitative relationship among economic variables may vary over time. Many financial and economic applications suggest that the value of cointegrating vector might be time-varying. For example, application of cointegration in investment analysis shows that frequent rebalancing is necessary to keep the portfolio in line with the index, indicating the value of cointegrating vector may be changing over time, see Xiao (2009).

A general conclusion of empirical studies is: cointegration relationships can not be found from these time series. Although the present value model suggests that asset prices are cointegrated with market fundamentals, empirically it is well known that stock prices are much more volatile than market fundamentals. Again, a plausible source of the additional volatility may come from time-varying cointegrating vectors, see Xiao (2009). Park (1999) considered the model where the cointegrating vector is a deterministic function of time $t$.

\[
Y_t = \beta(z_t)^\top X_t + \epsilon_t \tag{1.2}
\]
Cai and Park (2009) considered the functional-coefficient models, i.e. model 1.2, for nonstationary time series data. They derived their estimator based on the local linear estimation, but they found out a single value of bandwidth, $h$, can not make the estimation optimal in the sense of minimizing the asymptotic mean square error. They proposed a two-step estimation procedure to solve this problem. Xiao (2009) also considered 1.2 for nonstationary time series data and focused on inference procedures on both parameter instability and the hypothesis of cointegration.

In application with models (1.1)-(1.2), we predict the stock price of Morgan Stanley ($Y_t$) using the predictors, S&P 500 ($X_t$) and the 5 year daily Treasury yield rate ($z_t$). Both time-varying effects and nonlinearity effects are found. This motivates us to study the following model

$$y_i = \beta_0(t_i) + \{\beta_1(t_i) + \gamma(z_i)\}^T x_i + \varepsilon_i. \quad (1.3)$$

Since model (1.3) includes models (1.1) and (1.2) as specific examples, it can be used to validate if they are appropriate for fitting the above data. Where $x_i$ can be a $p$-dimensional $I(0)$ or $I(1), x_i$ does not involve constant, $t_i = i/n, z_i$ is $I(0), E(\varepsilon_i|x_i, z_i) = 0, \text{var}(\varepsilon_i|x_i, z_i) = \delta^2, \beta_0(t_i)$ is a $1 \times 1$ function, $\beta_1(t_i)$ is a $p \times 1$ function vector, $\gamma(z_i)$ is a $p \times 1$ function vector, $\varepsilon_i$ is a strictly $\alpha$-mixing stationary process. Without loss of generality, we assume that $E[\gamma(z_i)] = 0$. If $E[\gamma(z)]$ is not 0, we can take $\gamma(z) - E[\gamma(z)]$ as our new $\gamma(z)$.

We fit our models (1.3) with real data and compare the goodness of fit with those from models (1.1) and (1.2) in the section 8.

We propose a two-step estimation method to estimate the time-varying and nonlin-
ear coefficients for stationary or nonstationary explanatory variables. We show that our estimators are "oracle" in the sense that their asymptotic distributions are the same as the case with a known unknown variables.

The rest of this dissertation is organized as follows: Chapter 2 shows the model we consider in this dissertation. Chapter 3 gives a brief introduction of the two-step estimation procedure. Chapter 4 considers the case when $x_i$ is stationary. The asymptotic results for stationary $x_i$ are shown here. Chapter 5 considers the case when $x_i$ is nonstationary. The asymptotic results for nonstationary $x_i$ are shown here. Chapter 6 gives the simulation results for both stationary $x_i$ and nonstationary $x_i$. Chapter 7 is a real example for the model we consider. Concluding remarks are presented in Chapter 8. Proofs are contained in the Appendix.
CHAPTER 2: MODEL WITH TIME VARYING AND NONLINEAR EFFECTS

Assume a sample \( \{y_i\}_{i=1}^n \) are generated from

\[
y_i = \beta_0(t_i) + \{\beta_1(t_i) + \gamma(z_i)\}^\top x_i + \varepsilon_i, \tag{2.1}
\]

where \( x_i \) can be a \( p \)-dimensional \( I(0) \) or \( I(1) \), \( x_i \) does not involve constant, \( t_i = i/n \), \( z_i \) is \( I(0) \), \( E(\varepsilon_i|x_i, z_i) = 0 \), \( \text{var}(\varepsilon_i|x_i, z_i) = \delta^2 \beta_0(t_i) \) is a \( 1 \times 1 \) function, \( \beta_1(t_i) \) is a \( p \times 1 \) function vector, \( \gamma(z_i) \) is a \( p \times 1 \) function vector, \( \varepsilon_i \) is a strictly \( \alpha \)-mixing stationary process. Without loss of generality, we assume that \( E[\gamma(z_i)] = 0 \). If \( E[\gamma(z)] \) is not 0, we can take \( \gamma(z) - E[\gamma(z)] \) as our new \( \gamma(z) \).

When \( x_i \) and \( y_i \) both are nonstationary, \( \varepsilon_i \) is stationary, we say that \( x_i \) and \( y_i \) are cointegrated with a varying coefficient cointegration vector \( \beta_1(t_i) + \gamma(z_i) \), which is a vector of smooth functions of \( z_i \) and time \( t_i \). This setting is more general than the usual assumption that cointegration vector is constant.
Horowitz (2004) proposed a two-step estimation method to estimate the unknown link function. In the rst step, least squares is used to obtain a series approximation to each unknown function. The rst-step estimates are inputs to the second stage. Cai and Park (2009) shows local linear method can be used to estimate the unknown functions even thought the independent variables and dependent variable are non-stationary. Their estimators have good properties. Xiao (2009) shows local polynomial can be used to estimate the unknown functions in the same situation. However, we can not estimate all our unknown functions by local linear or local polynomial method at the same time because of the two variables $t_i$ and $z_i$. If we know the functions of $t_i$, we could estimate the unknown functions of $z_i$ by local linear method. If we know the functions of $z_i$, we could estimate the unknown functions of $t_i$ by local linear method. That is the basic idea of our two-step estimation method. If we could get good estimators in the first step estimation, we should expect the estimators by local linear method in the second step have the same properties as those in Cai and Park (2009) and Xiao (2009).

3.1 Orthogonal series estimation

Without loss of generality, we assume that the support of $z_t$ is $Z = [-1, 1]$. We normalize $\gamma(\cdot)$ as $E\gamma(z) = 0$ so that we could identify $\gamma(z)$. Let $\{p_k(\cdot), k = 1, 2, \ldots\}$
be a standard orthogonal basis for smooth functions on \([-1, 1]\) and satisfy that
\[
\int_{-1}^{1} p_k(x) \, dx = 0, \quad \text{and} \quad \int_{-1}^{1} p_k(x)p_j(x) \, dx = \begin{cases} 1 & \text{if } k = j; \\ 0 & \text{otherwise}. \end{cases}
\]

One choice of the orthogonal basis is orthogonal spline basis. Let
\[
P_{\kappa 0}(t) = [1, p_1(t), \ldots, p_\kappa(t)]^\top, \quad P_{\kappa 1}(t, z) = [1, p_1(t), \ldots, p_\kappa(t), p_1(z), \ldots, p_\kappa(z)]^\top, \quad \Theta_{\kappa 0} = (\theta_0, \theta_1, \ldots, \theta_\kappa)^\top, \quad \text{and} \quad \Theta_{\kappa d} = (\theta_{0,d}, \theta_{11,d}, \ldots, \theta_{1\kappa,d}, \theta_{21,d}, \ldots, \theta_{2\kappa,d})^\top,
\]
for \(d = 1, \ldots, p\). Then \(P_{\kappa 0}(t)\Theta_{\kappa 0}\) is a series approximation to \(\beta_0(t)\), and \(P_{\kappa 1}(t, z)\Theta_{\kappa d}\) is a series approximation to the \(d\)th component of \(\beta_1(t) + \gamma(z)\). Our orthogonal series estimator of \(\beta_0(t)\) and the \(d\)th component \(\beta_{1d}(t) + \gamma_d(z)\) of \(\beta_1(t) + \gamma(z)\) are respectively defined as
\[
\tilde{\beta}_0(t) = P_{\kappa 0}(t)^\top \hat{\Theta}_{\kappa 0} \quad \text{and} \quad \tilde{\beta}_{1d}(t) + \tilde{\gamma}_d(z) = P_{\kappa 1}(t, z)^\top \hat{\Theta}_{\kappa d},
\]
where
\[
(\hat{\Theta}_{\kappa 0}, \hat{\Theta}_{\kappa d}) = \arg \min_{\Theta_{\kappa j}} \sum_{i=1}^{n} \left\{ y_i - P_{\kappa 0}(t_i)\Theta_{\kappa 0} - \sum_{d=1}^{p} \Theta_{\kappa d}^\top P_{\kappa 1}(t_i, z_i)x_{i,d} \right\}^2, \quad (3.1)
\]
where \(x_{i,d}\) is the \(d\)th component of \(x_i\).

Define following notations:
\[
B = [\Theta_{\kappa 0}^\top, \Theta_{\kappa 1}^\top, \Theta_{\kappa 2}^\top, \ldots, \Theta_{\kappa p}^\top]^\top
\]
\[
\hat{B} = [\hat{\Theta}_{\kappa 0}^\top, \hat{\Theta}_{\kappa 1}^\top, \hat{\Theta}_{\kappa 2}^\top, \ldots, \hat{\Theta}_{\kappa p}^\top]^\top
\]
\[
A_i = [P_{\kappa 0}(t_i)^\top, P_{\kappa 1}(t_i, z_i)^\top x_{i,1}, P_{\kappa 2}(t_i, z_i)^\top x_{i,2}, \ldots, P_{\kappa p}(t_i, z_i)^\top x_{i,p}]^\top
\]

Equation 3.1 can be written as
\[
\hat{B} = \arg \min_{\Theta_{\kappa j}} \sum_{i=1}^{n} \left\{ y_i - A_i^\top B \right\}^2, \quad (3.2)
\]
\( \hat{B} \) can be found. \( \hat{B} = \left( \sum_{i=1}^{n} A_i A_i^T \right)^{-1} \left( \sum_{i=1}^{n} y_i A_i \right) \). In order to keep \( E\gamma(z) = 0 \), we have to centralize \( \tilde{\gamma}(z) \). Denote \( \tilde{\gamma}_d^*(z) = \tilde{\gamma}_d(z) - E\tilde{\gamma}_d(z) \), \( \tilde{\beta}^*_{1d}(t) = \tilde{\beta}_{1d}(t) + E\tilde{\gamma}_d(z) \). Then \( E\tilde{\gamma}_d^*(z) = 0, \tilde{\gamma}_d^*(z) \) and \( \tilde{\beta}^*_{1d}(t) \) are first step estimators.

The orthogonal series estimators will be employed as initial estimators for regression components in the second-stage estimation introduced below. The orthogonal series estimators are used to ensure that the biases of the first-stage estimators converge to zero rapidly.

3.2 Local smoother

It is well known that for any \( t \in [0, 1] \) and \( t_i \) in the neighborhood of \( t \), by Taylor’s expansion,

\[
\hat{\beta}_k(t_i) \approx \beta_k(t) + \beta'_k(t)(t_i - t) \equiv a_k + b_k(t_i - t). \quad k = 0, 1
\]

and for any \( z_i \) in the neighborhood of \( z \),

\[
\gamma(z_i) \approx \gamma(z) + \gamma'(z)(z_i - z) \equiv a_2 + b_2(z_i - z)
\]

Note that \( a_1, b_1, a_2 \) and \( b_2 \) are unknown vectors for every \( t \) and \( z \) and \( a_0 \) and \( b_0 \) are two unknown constants for every \( t \). In the second step, we minimize

\[
\sum_{i=1}^{n} \left[ y_i - \hat{\beta}_0'(t_i) - \hat{\gamma}^*(z_i) \right]^\top x_i - \left\{ a_2 + b_2(z_i - z) \right\}^\top x_i \right]^2 K_{h_1}(z_i - z) \quad (3.3)
\]

and get the minimizer \( \hat{a}_2 \) which estimates \( \gamma(z) \) denoted by \( \hat{\gamma}(z) \), where \( K_{h_1}(\cdot) = \frac{1}{h_1} K(\frac{\cdot}{h_1}) \). Similarly, we minimize

\[
\sum_{i=1}^{n} \left[ y_i - \hat{\beta}_0'(t_i) - \hat{\gamma}^*(z_i) \right]^\top x_i - \left\{ a_1 + b_1(t_i - t) \right\}^\top x_i \right]^2 K_{h_2}(t_i - t), \quad (3.4)
\]
and
\[ \sum_{i=1}^{n} [y_i - \tilde{\beta}_1^*(t_i) x_i - \tilde{\gamma}_1^*(z_i) x_i - \{a_0 + b_0(t_i - t)\}] K_{ho}(t_i - t), \quad (3.5) \]
get the minimizer \( \hat{a}_0 \) and \( \hat{a}_1 \) which estimates \( \beta_0(t) \) and \( \beta_1(t) \) denoted by \( \hat{\beta}_0(t) \) and \( \hat{\beta}_1(t) \).

We could see from equation 3.3,3.4 and 3.5 that we estimate the unknown functions each time as if we have already known the other unknown functions. We show oracle properties for our estimators in two cases: stationary \( x_i \) and nonstationary \( x_i \) in Chapter 4. It is easy to derive close form of our estimators in the following.

Define following notations:
\[ P_{k1}(t) = (1, p_1(t), \ldots, p_k(t))^\top, \quad P_{k1}(z) = (p_1(z), \ldots, p_k(z))^\top, \]
\[ \Theta_{kd1} = (\theta_{0,d}, \theta_{11,d}, \ldots, \theta_{11,k,d})^\top, \quad \Theta_{kd2} = (\theta_{21,d}, \ldots, \theta_{21,k,d})^\top \]
It can be easily check that \( P_{k1}(t, z) = [P_{k1}(t), P_{k1}(z)]^\top, \quad \Theta_{kd} = [\Theta_{kd1}^\top, \Theta_{kd2}^\top]^\top \)
We have the folllowing notations:
\[ \tilde{\Theta}_t = [\hat{\theta}_{k0}^\top, \hat{\theta}_{k11}^\top, \hat{\theta}_{k12}^\top, \ldots, \hat{\theta}_{kpt}^\top]^\top \]
\[ P_{k}(t_i, x_i) = [P_{k0}(t_i)^\top, P_{k1}(t_i)^\top x_{i,1}, P_{k1}(t_i)^\top x_{i,2}, \ldots, P_{k1}(t_i)^\top x_{i,p}]^\top \]
\[ W_{ih}(z) = (1, \frac{z_i - z}{h}) \otimes x_i^\top, \quad A^* = \sum_{i=1}^{n} W_{ih}(z)^\top W_{ih}(z) K_h(z_i - z) \]
\[ B^* = [\sum_{i=1}^{n} (y_i - \tilde{\beta}_0^*(t_i) - \tilde{\beta}_1^*(t_i)^\top x_i) W_{ih}(z)^\top K_h(z_i - z)] \]
It is easy to check that \( \hat{\theta}_{kd} \), which is the estimator of \( \theta_{kd} \), is the first \( k + 1 \) elements of \( \hat{\Theta}_{kd} \). We will have the following solution for \( (a_2, h_2) \):
\[
\begin{bmatrix}
\hat{a}_2 \\
h_2
\end{bmatrix} = [A^*]^{-1} B^*
\]
Similarly, \( \hat{a}_0 \) and \( \hat{a}_1 \) could be easily determined.
4.1 Notations and conditions

Some notations:

$A_{ik}$ denotes the kth component of $A_i$.

$\hat{Q}_\kappa = n^{-1} \sum_{i=1}^{n} A_i A_i^\top$. Then $Q_\kappa = E\hat{Q}_\kappa$. Let $Q_{ij}$ denote the (i,j) element of $Q_\kappa$.

$Z_k$ is a $d(\kappa) \times n$ matrix whose ith column is $A_i$.

$E_{p_i(t_k)p_j(z_k)} = C_{ij}$ for all $i,j$ from 1 to $\kappa$ and $k$ from 1 to $n$.

for $j \leq 0$, $\mu_j(K) = \int_{-\infty}^{\infty} v^j K(v)dv$, and $\nu_j(K) = \int_{-\infty}^{\infty} v^j K^2(v)dv$. $S = E(x_i x_i^\top | z_i = z)$,$S_0 = E(x_i x_i^\top | t_i = t)$

$\gamma^{(s)}(z) = d^s \gamma(z)/dz^s$ for $s = 1$ and 2

$R(z_i) = \gamma(z_i) - \gamma(z) - \gamma^{(1)}(z)(z_i - z)$

$\overline{A}(t_i) = [P_{n0}(t_i)^\top, P_{n1}(t_i)^\top x_{i1}, P_{n1}(z_i)^\top \cdot 0, P_{n1}(t_i)^\top x_{i2}, P_{n1}(z_i)^\top \cdot 0, \cdots, P_{n1}(t_i)^\top x_{ip}, P_{n1}(z_i)^\top]$. 0$]^\top$

$S_{nk}(B) = \frac{1}{n} \sum_{i=1}^{n} (y_i - A_i^\top B)^2$

$S_k(B) = E(S_{nk}(B))$

$\theta_{0} = \arg \min S_k(B)$

$b_{\kappa0}(i) = \beta_0(t_i) + (\beta_1(t_i) + \gamma(z_i))^\top x_i - A_i^\top \theta_{\kappa0}$

$\overline{b}_{\kappa0}(i) = \beta_0(t_i) + \beta_1(t_i)^\top x_i - \overline{A}_i^\top \theta_{\kappa0}$

$\overline{\theta}_k = [\theta_{\kappa0}, \theta_{\kappa1}, \theta_{\kappa1z} * 0, \theta_{\kappa2}, \theta_{\kappa2z} * 0, \cdots, \theta_{\kappa dt}, \theta_{\kappa dz} * 0]^\top$
$$\Sigma_{0} = \begin{pmatrix}
\nu_0(K)S & \nu_1(K)S \\
\nu_1(K)S & \nu_2(K)S
\end{pmatrix}$$

The following conditions are needed to derive the asymptotic properties of the proposed estimators.

(A1) $x_i$ is a $p$-dimensional $I(0)$. Let $x_{i,j}$ is the $j$th component of $x_i$, without loss of generality, assume $x_{i,j} = b_j x_{i-1,j} + \delta_{i,j}$, where $1 \leq i \leq n$, $1 \leq j \leq p$. $\delta_{i,j}$ is independent with $E\delta_{i,j} = 0$ and $Var\delta_{i,j} = \zeta_j^2$. $E(\delta_{i,j}\delta_{i,k}) = \zeta_{j,k}$ for $1 \leq k \neq j \leq q$ so $E(x_{i,j}x_{i,k}) = \zeta_{j,k}$.

(A2) $t_i = i/n, \varepsilon_i$ has finite fourth moment, $E(\varepsilon_i | x_i, z_i) = 0$, $var(\varepsilon_i | x_i, z_i) = \delta^2$ is a positive constant.

(A3) $z_i$ is $I(0), f(z)$ is continuously differentiable in a neighborhood of $z$ and $f_z(z) > 0$.

(A4) (i) Without loss of generality, we assume that $E[\gamma(z_i)] = 0$

(ii) $\gamma(z)$ is twice continuously differentiable in $z$ for all $z \in [-C, C]$, where $C$ is any constant in $\mathbb{R}$.

(iii) $\beta(t)$ is twice continuously differentiable in $t$ for all $t \in [0, 1]$, $S$ is positive-definite and continuous in a neighborhood of $z$.

(A5) There are constants $C_Q < \infty$ and $c_\lambda > 0$ such that $|Q_{ij}| \leq C_Q$ and $\lambda_{\kappa,\min} > c_\lambda$ for all $\kappa$ and all $i, j = 1, ..., d(\kappa)$.

(A6) Assume $b_{\kappa_0}(i) = O(\kappa^{-2})$ for all $i$ from 1 to $n$

(A7) (i) Assume $h_1 = C_{h_1}n^{-1/5}$, $h_2 = C_{h_2}n^{-1/5}$ and $h_0 = C_{h_0}n^{-1/5}$ for some constant $C_{h_1}, C_{h_2}$ and $C_{h_0}$ satisfying $0 < C_{h_1} < \infty$, $0 < C_{h_2} < \infty$ and $0 < C_{h_0} < \infty$.
(ii) $\kappa = C_n n^\nu$ for some constant $C_\kappa$ satisfying $0 < C_\kappa < \infty$ and some $\nu$ satisfying $\frac{1}{5} < \nu < \frac{3}{10}$

(A8) Assume $\sup_{t_i, z_i} \|p_\kappa(t_i, z_i)\| = O(\kappa^{1/2})$

(A9) Assume the kernel function $K(\cdot)$ is a symmetric and continuous density function, supported by $[-1, 1]$ and $\mu_0(K) = 1, \mu_1(K) = 0$

We give some comments on the above conditions. $x_i$ is stationary. However, we have assumptions A1 to make the proof can be done easily. Assumptions A2 and A3 are regularity conditions. Assumption A4 defines the sense in which $\gamma(z_i)$ and $\beta(t)$ must be smooth. Assumption A4(i) is needed for identification. Assumptions A4(ii) and A4(iii) are smoothness conditions. Assumption A5 insures the existence and nonsingularity of the covariance matrix of the asymptotic form of the first-stage estimator. This is analogous to assuming that the information matrix is positive definite in parametric maximum likelihood estimation, see Horowitz (2004). Assumption A6 bounds the magnitudes of the basis functions and insure that the errors in the series approximations to the $\gamma(z)$ converge to zero sufficiently rapidly as $\kappa \to \infty$. Assumption A7 states the rates at which $\kappa \to \infty$ and $h \to \infty$ as $n \to \infty$. The assumed rate of convergence of $h$ is well known to be asymptotically optimal for kernel regression when the conditional mean function is twice continuously differentiable. The required rate for $\kappa$ insures that the asymptotic bias and variance of the first stage estimator are sufficiently small to achieve an $n^{-2/5}$ rate of convergence in the second stage. See Horowitz (2004). Assumption A8 helps the second-stage estimator avoid the curse of dimensionality. These conditions are satisfied by splines and the Fourier basis. To simplify the proofs of the theoretical results, $K(\cdot)$ is assumed to have a compact support.
It can be relaxed to allow kernels with noncompact support if we put restrictions on the tail of $K(\cdot)$, see Jiang J. (2008)

4.2 Asymptotics

In this section, we establish the asymptotic normality of the two-step estimators when $x_i$ is stationary. Detailed proof of the following Theorems are provided in Appendix.

Theorem 4.1. Under conditions (A1) $\sim$ (A9),
\[
\sqrt{nh_1}[\hat{\gamma}(z) - \gamma(z) - \frac{h^2}{2} \mu_2(K)r^{(2)}(z)\{1 + o_p(1)\}] \overset{d}{\rightarrow} N\{0, f_z(z)^{-1}\delta^2S^{-1}\nu_0(K)\}
\]

Theorem 4.2. Under conditions (A1) $\sim$ (A9),
\[
\sqrt{nh_2}[\hat{\beta}_1(t) - \beta_1(t) - \frac{h^3}{2} \mu_2(K)\beta_1^{(2)}(t)\{1 + o_p(1)\}] \overset{d}{\rightarrow} N\{0, \delta^2S^{-1}\nu_0(K)\}
\]

Theorem 4.3. Under conditions (A1) $\sim$ (A9),
\[
\sqrt{nh_0}[\hat{\beta}_0(t) - \beta_0(t) - \frac{h^3}{2} \mu_2(K)\beta_0^{(2)}(t)\{1 + o_p(1)\}] \overset{d}{\rightarrow} N\{0, \delta^2\nu_0(K)\}
\]

Above theorems can be extended to that $X_i, Z_i, \varepsilon, \zeta$ is a strictly $\alpha$-mixing stationary process with more than second moment. See assumption A6 in Cai and Park (2009). Theorem 4.1 is exactly the same as that in Cai (2000). The bandwidth is taken to be of the order $n^{-1/5}$ so that $\hat{\gamma}(z) - \gamma(z), \hat{\beta}_1(t) - \beta_1(t)$ and $\hat{\beta}_0(t) - \beta_0(t)$ reach the optimal convergent rate.
5.1 Notations and conditions

$x_i$, which is a vector of $I(1)$ process, can be expressed as $x_i = x_{i-1} + \delta_i = x_0 + \sum_{s=1}^{i} \delta_s (i \geq 1)$, where $\delta_s$ is an $I(0)$ process with mean zero and variance $\Omega_\delta$.

\[
\frac{x_{[nr]}}{\sqrt{n}} = \frac{x_i}{\sqrt{n}} = \frac{x_0}{\sqrt{n}} + \frac{1}{\sqrt{n}} \sum_{s=1}^{i} \delta_s = \frac{x_0}{\sqrt{n}} + \frac{1}{\sqrt{n}} \sum_{s=1}^{[nr]} \delta_s \tag{5.1}
\]

where $r = i/n$ and $[x]$ denotes the integer part of $x$, see Cai and Park (2009). Under some regularity conditions, Donsker’s theorem; see Theorems 14.1 and 19.2 in Billingsley (1999) for iid $\delta_i$ and $\rho$-mixing $\delta_i$, respectively, generalizes in an obvious way to the multivariate cases and leads to

\[
\frac{x_{[nr]}}{\sqrt{n}} \Rightarrow W_\delta(r) \quad \text{as } n \to \infty \tag{5.2}
\]

where $W_\delta(\cdot)$ is a $p$-dimensional Brownian motion on $[0, 1]$ with covariance matrix $\Sigma_\delta$.

For any Borel measurable and totally Lebesgue integrable function $\Gamma(\cdot)$, one has

\[
\frac{1}{n} \sum_{i=1}^{n} \frac{\Gamma(x_{[nr]})}{\sqrt{n}} \to d \int_{0}^{1} \Gamma(W_\delta(s))ds \quad \text{as } n \to \infty \tag{5.3}
\]

where $\to d$ denotes the convergence in distribution, so that, for $l = 1, 2$,

\[
\frac{1}{n} \sum_{i=1}^{n} \left(\frac{x_{[nr]}}{\sqrt{n}}\right)^{\otimes \ell} \to d \int_{0}^{1} W_\delta(s)^{\otimes \ell}ds \equiv W_{\delta,\ell} \quad \text{as } n \to \infty \tag{5.4}
\]

see Theorem 1.2 in Berkes (2006) and Cai and Park (2009).
Define $\hat{Q}_κ^* = n^{-2} \sum_{i=1}^{n} A_i A_i^\top$ and $Q_κ^* = E\hat{Q}_κ^*$. Let $Q_κ^*_{ij}$ denote the $(i,j)$ element of $Q_κ^*$.

(B1) $x_i$ is a $p$-dimensional $I(1)$, $x_{i,j}$ is the $j$th component of $x_i$. Without loss of generality, assume $x_{i,j} = x_{i-1,j} + \delta_{i,j}$, where $1 \leq i \leq n$, $1 \leq j \leq p$, $\delta_{i,j}$ is independent with $E\delta_{i,j} = 0$, $Var\delta_{i,j} = \zeta_j^2$.

(B2) (i) Assume $h_1 = C_{h_1} n^{-2/5}$ for some constant $C_{h_1}$ satisfying $0 < C_{h_1} < \infty$, $\kappa = C_κ n^\nu$ for some constant $C_κ$ satisfying $0 < C_κ < \infty$ and some $\nu$ satisfying $\frac{3}{20} < \nu < \frac{7}{20}$.

(ii) Assume $h_2 = C_{h_2} n^{-2/5}$ for some constant $C_{h_2}$ satisfying $0 < C_{h_2} < \infty$, $\kappa = C_κ n^\nu$ for some constant $C_κ$ satisfying $0 < C_κ < \infty$ and some $\nu$ satisfying $\frac{3}{20} < \nu < \frac{7}{20}$.

(iii) Assume $h_0 = C_{h_0} n^{-1/5}$ for some constant $C_{h_0}$ satisfying $0 < C_{h_0} < \infty$, $\kappa = C_κ n^\nu$ for some constant $C_κ$ satisfying $0 < C_κ < \infty$ and some $\nu$ satisfying $\frac{1}{5} < \nu < \frac{3}{10}$.

(B3) Assume $\sup_x |b_κ(0)| = O(κ^{-2})$

(B4) Assume $\sup_{t_i, z_i} \|p_κ(t_i, z_i)\| = O(κ^{1/2})$ for all $i$ from 1 to $n$

(B5) There are constants $C_Q^* < \infty$ and $c_\lambda^* > 0$ such that $|Q_κ^*_{ij}| \leq C_Q^*$ and $\lambda_{κ,min} > c_\lambda^*$ for all $κ$ and all $i, j = 1, \ldots, d(κ)$.

(B6) Assume $t_i = i/n, z_i$ is stationary, $\varepsilon_i$ has a finite fourth moment, $E(\varepsilon_i|X_t, Z_t) = 0$, $var(\varepsilon_i|X_t, Z_t) = \delta^2_i$, $\beta_0(t_i)$ is a $1 \times 1$ function, $\varepsilon_i$ is a strictly $\alpha$-mixing stationary process.

(B7) Without loss of generality, we assume that $E[\gamma(z_i)] = 0$, $\gamma(z)$ is twice continuously differentiable in $z$ for all $z \in \mathbb{R}$, $\beta(t)$ is twice continuously differentiable in $t$ for all $t \in [0, 1]$. 

(B8) Assume the kernel function $K(\cdot)$ is a symmetric and continuous density function supported by $[-1, 1]$ and $\mu_0(K) = 1, \mu_1(K) = 0$

We give some comments on the above conditions. Assumption B1 makes the proof can be done easily. Assumption B2 states the rates at which $\kappa \to \infty$ and the bandwidth converges to 0 as $n \to \infty$. It requires the first-stage estimator to be undersmoothed. Undersmoothing is needed to insure the sufficiently rapid convergence of the bias of the orthogonal series estimator. We will show the asymptotic normality of the two-stage estimator does not depend on the choice of $\kappa$ if B2 is satisfied. Optimizing the choice of $\kappa$ would require a rather complicated higher-order theory and is beyond the scope of this dissertation, see Jiang J. (2008). Assumption B3 bounds the magnitudes of the basis functions and insures that the errors in the series approximations to the $\gamma(z)$ converge to zero sufficiently rapidly as $\kappa \to \infty$. See Horowitz (2004). Assumption B4 helps the second-stage estimator avoid the curse of dimensionality. These conditions are satisfied by splines and the Fourier basis. $\alpha$-mixing is one of the weakest mixing conditions for weakly dependent stochastic processes. Stationary linear and nonlinear time series or Markov chains fulfilling certain (mild) conditions are $\alpha$-mixing with exponentially decaying coefficients; see discussions and examples in Cai (2002), Carrasco (2002) and Chen (2005). The conditional homoscedastic error. Assumption B5 insures the existence and nonsingularity of the covariance matrix of the asymptotic form of the first-stage estimator, see Horowitz (2004). Assumption B6 can be relaxed to allow for conditional heteroscedasticity of the form $\text{var}(\varepsilon_t|X_t, Z_t) = \delta^2(Z_t)$, i.e. the conditional variance is only a function of the stationary covariates $(Z_t)$. However, it is technically difficult to let it also be a
function of the nonstationary covariate $X_t$; see Cai and Park (2009). Assumption B7 are smoothness conditions. Assumption B8 that $K(\cdot)$ be compactly supported is imposed for the sake of brevity of proofs, and can be removed at the cost of lengthier arguments.

5.2 Asymptotics

In this section, we establish the asymptotic normality of the two-step estimators when $x_i$ is nonstationary. Detailed proof of the following Theorems are provided in Appendix.

**Theorem 5.1.** Under conditions $(B1),(B2)(i),(B3)\sim(B8)$

$$n\sqrt{h_1}[\hat{\gamma}(z) - \gamma(z) - \frac{h^2}{2}\mu_2(K)r^{(2)}(z)\{1 + o_p(1)\}] \overset{d}{\to} MN(\Sigma_\delta(z))$$

where $MN(\Sigma_\delta(z))$ is a mixed normal distribution with mean zero and conditional covariance matrix given by $\Sigma_\delta(z) = \delta^2\nu_0(K)W_{\delta,2}^{-1}/f_z(z)$

Here, a mixed normal distribution is defined as follows. Conditional on the random variable that appears at the asymptotic variance, the estimator has an asymptotic normal distribution, see Phillips (1989) and Phillips (1998) for a formal definition of a mixed normal distribution Cai and Park (2009).

We have similar results for $\beta_1(t)$ and $\beta_0(t)$

**Theorem 5.2.** Under conditions $(B1),(B2)(ii),(B3)\sim(B8)$

$$n\sqrt{h_2}[\hat{\beta}_1(t) - \beta_1(t) - \frac{h^3}{2}\mu_2(K)\beta^{(2)}_1(t)\{1 + o_p(1)\}] \overset{d}{\to} MN(\Sigma_\delta(t))$$

where $MN(\Sigma_\delta(t))$ is a mixed normal distribution with mean zero and conditional covariance matrix given by $\Sigma_\delta(t) = \delta^2\nu_0(K)W_{\delta,2}^{-1}$
Theorem 5.3. Under conditions (B1), (B2)(iii), (B3) ∼ (B8),

\[ \sqrt{n \theta_0} [\hat{\beta}_0(t) - \beta_0(t) - \frac{h_2^2}{2} \mu_2(K) \beta_0^{(2)}(t) \{1 + o_p(1)\}] \overset{d}{\to} N\{0, \delta^2 \nu_0(K)\} \]

The rate of convergence in Theorem 5.1 and Theorem 5.2 is \(n \sqrt{h}\), which is the same as those in Cai and Park (2009) and Xiao (2009) for nonstationary \(x_i\) case. It implies that our estimators, \(\hat{\gamma}(z)\) and \(\hat{\beta}_1(t)\), are "oracle" in the sense that their asymptotic distribution are the same as the case with a known \(\beta_1(t)\) and \(\gamma(z)\). The bandwidth is taken to be of the order \(n^{-2/5}\) so that \(\hat{\gamma}(z) - \gamma(z)\) and \(\hat{\beta}_1(t) - \beta_1(t)\) reach the optimal convergent rate. The bandwidth is taken to be of the order \(n^{-1/5}\) so that \(\hat{\beta}_0(t) - \beta_0(t)\) reach the optimal convergent rate.
We did simulations to demonstrate that the proposed two step estimators could give an accurate approximation to the unknown functions. Since B-spline method (1978) is efficient in digital computation and functional approximation, we here use the B-spline basis in the first-step estimation. $\kappa$ is chosen to be 8. Smaller number $\kappa$ or larger number $\kappa$ will not have a big effect on the results. We choose standard normal kernel as our kernel function used in the simulation. We consider the following model.

\[
y_i = \beta_0(t_i) + (\beta_{11}(t_i) + \gamma_1(z_i))x_{i,1} + (\beta_{12}(t_i) + \gamma_2(z_i))x_{i,2} + \varepsilon_i
\]

\[
e^{-3t_i} + 20t_i + (-40t_i^2 - 20t_i + 1.5z_i^2 - 7.5z_i - 8)x_{i,1} + (e^{4t_i} + 2t_i + 3z_i^2 - 6z_i - 16)x_{i,2} + \varepsilon_i
\]

(6.1)

where $\gamma_1(z_i) = 1.5z_i^2 - 7.5z_i - 8$ and $\gamma_2(z_i) = 3z_i^2 - 6z_i - 16$. $p = 2$ in this example.

We assume that $\varepsilon \sim N(0, 0.25)$, $z_i = 0.3z_{i-1} + U_i$ and $U_i \sim \text{Uniform} (-4, 4)$ in the above model. The initial value for the first component of $x$ is denoted by $x_{1,1}$, the first component of $x$ at time $i$ is denoted by $x_{i,1}$, the initial values for the second component of $x$ is denoted by $x_{1,2}$, the second component of $x$ at time $i$ is denoted by $x_{i,2}$. Note that we choose those $\gamma(\cdot)$ functions so that $E(1.5z^2 - 7.5z - 8) = 0$ and $E(3z^2 - 6z - 16) = 0$. 


6.2 Stationary $x_i$

Example 1:

Choose $x_{1,1} = x_{1,2} = 0$, $x_{i,1} = 0.9x_{i-1,1} + \delta_{1i}$, where $\delta_{1i} \sim t(3)$, $x_{i,2} = 0.6x_{i-1,2} + \delta_{2i}$, where $\delta_{2i} \sim t(7)$. $y$ is generated from above model (6.1). So that $x_1, x_2, y$ all are stationary. 10 grid points are chosen with 500 simulation at each grid points.

Simulation results are shown in Figure 1 for $n = 100$.

Simulation results are shown in Figure 2 for $n = 400$.

The solid lines are true lines of $\beta_0$, $\beta_1$, $\beta_2$, $\gamma_1$ and $\gamma_2$ functions in Figure 1 and Figure 2. The middle dash dot lines in Figure 1 and Figure 2 are the median of the estimators. The upper and lower dot lines in Figure 1 and Figure 2 are 2.5% and 97.5% quantile of the estimators.

You should see from Figure 1 and Figure 2 that the estimation is very good. The solid lines almost cover the middle dash dot lines.
Figure 1: Estimated function $\beta$ and $\gamma$, their mediums and 95% pointwise confidence intervals for Model 1.3 with $x$ is stationary and $n = 100$. 

$x$ is stationary
Figure 2: Estimated function $\beta$ and $\gamma$, their mediums and 95% pointwise confidence intervals for Model 1.3 with $x$ is stationary and $n = 400$
CHAPTER 7: REAL EXAMPLE

We consider a real application here. We download 5 year daily Treasury yield rate, 6 month daily Treasury yield rate, stock price of Morgan Stanley and price of SP500 from website. All data are from the Jan. 2nd, 2003 to Dec. 12th, 2015. The sample size is 3249. It can be easily seen from table 1 that stock price of Morgan Stanley and price of SP500 are nonstationary by ADF test, however, log "difference" of 5 year daily Treasury yield rate and 6 month daily Treasury yield rate could be treat as stationary, see AN APPLICATION in Jiang (2014). We build our model based on CAPM model: \( y_i = \beta_0(t_i) + \gamma_0(z_i) + (\beta_1(t_i) + \gamma_1(z_i))x_i + \epsilon_i \). We choose log "difference" of 5 year daily Treasury yield rate and 6 month daily Treasury yield rate as \( z \), stock price of Morgan Stanley as \( y \), price of SP500 as \( x \). It is well known that return of SP500 and Morgan Stanley, which is stationary, is \( x_i \) and \( y_i \), in traditional CAPM. We can see that \( x_i \) and \( y_i \) are nonstationary and are the price of SP500 and Morgan Stanley, respectively, in our model. That is different with those in the traditional CAPM. The coefficients are constant in traditional CAPM. But they are not constant in above model. We split the sample into two part: training sample, the first 3000 data and testing sample, the remaining 249 data. Define one step forecast \( \hat{y}_j \) for \( y \) at time \( j \) as following: \( \hat{y}_j = \hat{\beta}_0(t_{j-1}) + \hat{\gamma}_1(z_{j-1}) + (\hat{\beta}_1(t_{j-1}) + \hat{\gamma}_1(z_{j-1}))x_{j-1} \), where \( j \) from 3001 to 3249 and \( \hat{\beta}_0, \hat{\gamma}_0, \hat{\beta}_1 \) and \( \hat{\gamma}_1 \) are estimated by the proposed two-step estimation method using only the data from 1 to \( j - 1 \).
Figure 4 shows the cointegration relationship between Stock price of Morgan Stanley and price of SP500 as well as the estimated $\beta_0$ function. The functions in 3 is estimated by the 3249 data. We could see that $\beta_0$ and $\beta_1$ change with time and $\gamma_0$ and $\gamma_1$ change with yield rate. 4 shows that the positive relationship between Stock price of Morgan Stanley and price of SP500 is increasing from 2003 to 2006 and it is relatively high before 2008. This implies that the market is bull at that time. However, this positive relationship is decreasing during the crisis. It reaches the bottom at 2011. After that it begins to increase. This is coincident with what we have observed in the financial market now. Financial market begins to recover after 2011.

We compare our model with model 1.1 in Figure 5 and model 1.2 in Figure 6. We can see that model 1.1 works well here. The residual is stationary from ADF test. Estimated stock price is close to the true stock price over time. I calculate the variance of the residual, which is 13.380. That is larger than the variance of the residual in our model 1.3, which is 5.859. We believe our model 1.3 is better than model 1.1. It is easily to see that our model 1.3 is better than model 1.2 from 6 too. The variance of the residual in 1.2 is 68.504.

We test the $\hat{\epsilon}$ by ADF test. The test statistic is -2.9771, which has a p-value 0.01. ADF test rejects the null hypothesis that $\hat{\epsilon}$ is nonstationary. So $\hat{\epsilon}$ is stationary, which implies that $y_i$ and $x_i$ are cointegrated. 7 shows $\hat{y}$ from the first 3000 data and the one-step forecast. We can see that the estimation error is small and the forecast is very well.
| Stock price of Morgan Stanley | -0.5896 | 0.4284 |
| Price of SP500 are nonstationary | 1.3095 | 0.9522 |
| Treasury yield rate from Jiang (2014) | -3.3093 | 0.01 |

Table 1: ADF test for stock price of Morgan Stanley, price of SP500 and 5 year daily Treasury yield rate

![Graphs](#)

Figure 3: Estimated functions from model 1.3
Figure 4: Estimated stock price and function of coefficient from model 1.3
Figure 5: Results from model 1.1
Figure 6: Results from model 1.2
Figure 7: $\hat{y}$ and One step forecast 1.2
CHAPTER 8: DISCUSSION

In this dissertation, we studied the varying coefficient model with both nonlinear effects and time-varying effects for stationary and nonstationary data. We suggested using the proposed two-step method to estimate the unknown coefficient functions and derived the asymptotic properties of the proposed estimators. Our estimation method could be extended to the function coefficient model with more than two variables in coefficient. We would like to mention three interesting future research topics related to this dissertation. First, it would be very useful and important to discuss how to select data-driven (optimal) bandwidths theoretically and empirically. Secondly, an important extension would be to generalize the asymptotic analysis of this dissertation to the case where both \( z_i \) and \( x_i \) are nonstationary. Further, we can consider an extension of the test in Xiao (2009) so that we could test not only \( I(1) \) process but also \( I(2), I(3) \) or even \( I(p) \) process. We are currently exploring these extension.
REFERENCES


APPENDIX A: Sketch of Proofs

Theorem 4.1

This section begins with lemmas that are used to prove Theorem 4.1.

Lemma A.1. If $A$ and $B$ are nonnegative matrices, then

$(a) \lambda_{\min}(A)Tr(B) \leq Tr(AB) \leq \lambda_{\max}(A)Tr(B)$;
$(b) \lambda_{\min}(A)\lambda_{\max}(B) \leq \lambda_{\max}(AB) \leq \lambda_{\max}(A)\lambda_{\max}(B)$.

Proof of lemma A.1

Part(a) is the lemma 6.5 of Zhou S (1998). Part(b) is a basic inequality.

Lemma A.2. Let $\Omega_{\kappa 1} = Q_{\kappa}^{-1}E(A_{i}^{\otimes 2}\epsilon_{i}^{2})Q_{\kappa}^{-1}$ and $\Omega_{\kappa 2} = Q_{\kappa}^{-1}E(A_{i}^{\otimes 2})Q_{\kappa}^{-1}$, by (A3) and (A6), the largest eigenvalues of $E(A_{i}^{\otimes 2}\epsilon_{i}^{2}), E(A_{i}^{\otimes 2})$, $\Omega_{\kappa 1}$ and $\Omega_{\kappa 2}$ are bounded for all $\kappa$.

Proof of lemma A.2

This result holds from the same argument as for Lemma 2 of Jiang J. (2008).

Lemma A.3. If condition (A1) - (A9) hold, then

$(a) \|\hat{Q}_{\kappa} - Q_{\kappa}\|^2 = O_p(\kappa^2/n)$
$(b) \|\hat{Q}_{\kappa}^{-1}\|^2 = O_p(\kappa,\|Q_{\kappa}^{-1}\|)^2 = O_p(\kappa)$
$(c) \|Q_{\kappa}^{-1}(Q_{\kappa} - \hat{Q}_{\kappa})\|^2 = O_p(\kappa^2/n)$

Proof of lemma A.3

$(a) E\|\hat{Q}_{\kappa} - Q_{\kappa}\|^2 = \sum_{k=1}^{d(\kappa)} \sum_{j=1}^{d(\kappa)} E\left(n^{-1} \sum_{i=1}^{n} A_{ik}A_{ij} - Q_{kj}\right)^2$
$= \sum_{k=1}^{d(\kappa)} \sum_{j=1}^{d(\kappa)} \left(E\frac{n^{-2}}{} \sum_{i=1}^{n} \sum_{\ell=1}^{n} A_{ik}A_{\ell k}A_{\ell j}A_{ij} - Q_{kj}^2\right)$
Define $M_{k_1,j_1} = En^{-2} \sum_{i=1}^{n} \sum_{\ell=1}^{n} A_{ik_1} A_{\ell k_1} A_{ij_1} A_{\ell j_1} - Q^2_{k_1 j_1}$, where $A_{ik_1}$ are from $P_{km_1}(t_i, z_i)\top x_{i,m_1}$ and $A_{ij_1}$ are from $P_{km_2}(t_i, z_i)\top x_{i,m_2}$ for any $m_1$ and $m_2$ from 1 to $p$ and $m_1 \neq m_2$.

Define $N_{k_2,j_2} = En^{-2} \sum_{i=1}^{n} \sum_{\ell=1}^{n} A_{ik_2} A_{\ell k_2} A_{ij_2} A_{\ell j_2} - Q^2_{k_2 j_2}$, where $A_{ik_2}$ and $A_{ij_2}$ are from $P_{km}(t_i, z_i)\top x_{i,m}$ for any $m$ from 1 to $p$. Then $E\|\hat{Q}_\kappa - Q_\kappa\|^2 = \sum_{k_1} \sum_{j_1} M_{k_1,j_1} + \sum_{k_2} \sum_{j_2} N_{k_2,j_2}$

In the following, we will prove $M_{k_1,j_1} = O(n^{-1})$ and $N_{k_2,j_2} = O(n^{-1})$.

With out loss of generality, assume $A_{ik_2} = p_1(t_i)x_{i,1}$, $A_{ij_2} = p_1(z_i)x_{i,1}$ and $Ex_{i,1} = 0$

It is easy to check that $EA_{ik_2} A_{ij_2} = C_{11}\zeta^2_1/(1 - b_1^2)$ and $Q^2_{k_2 j_2} = C^2_{11}\zeta_1^4(1 - b_1^2)^{-2}$

$EA_{ik_2} A_{ij_2}, A_{i+1k_2} A_{i+1j_2} = E(p_1(t_i)x_{i,1}p_1(z_i)x_{i,1}p_1(t_{i+1})(b_1x_{i,1} + \delta_{i+1,1})p_1(z_{i+1})(b_1x_{i,1} + \delta_{i+1,1}))$

$= C^2_{11} b_1^4 Ex_{i,1} + C^2_{11}\zeta_1^4(1 - b_1^2)^{-1}$

$EA_{ik_2} A_{ij_2}, A_{i+2k_2} A_{i+2j_2} = C_{11}^2 b_1^4 E x_{i,1}^4 + C^2_{11}\zeta_1^4(1 - b_1^2)^{-1}(b_1^2 + 1)$

Similar arguments yield that

$E \sum_{i=1}^{n} \sum_{\ell=1}^{n} A_{ik_2} A_{ij_2} A_{\ell k_2} A_{\ell j_2}$

$= C_{11}^2 E x_{i,1}^4[n + \sum_{m=1}^{n-1} 2(n - m)b_1^{-2m} + C_{11}^2 \zeta_1^4(1 - b_1^2)^{-1} \sum_{m=1}^{n-1} [2(n - m)(\sum_{s=1}^{m} b_1^2 s^{-1})]$

It is easy to check that $C_{11}^2 E x_{i,1}^4[n + \sum_{m=1}^{n-1} 2(n - m)b_1^{-2m}] = O(n)$.

Note that $\lim_{n \to \infty} n^2 b_1^{2(n-1)} = 0$

$\sum_{m=1}^{n-1} [2(n - m)(\sum_{s=1}^{m} b_1^{2(s-1)})] = \sum_{m=1}^{n-1} [2b_1^{2(m-1)} \sum_{s=m}^{n-1} (n - s)]$

$= \sum_{m=1}^{n-1} b_1^{2(m-1)}(n-m)(n-m+1)$

$= n^2 \sum_{m=1}^{n-1} b_1^{2(m-1)} - n^2 \sum_{m=1}^{n-1} b_1^{2(m-1)}(2m-1) + \sum_{m=1}^{n-1} b_1^{2(m-1)}(m^2 + m)$

$\to n^2(1 - b_1^2)^{-1} + O(n)$

$N_{k_2,j_2} = n^{-2}[C_{11}^2\zeta_1^4(1 - b_1^2)^{-2}n^2 + O(n)] - C_{11}^2\zeta_1^4(1 - b_1^2)^{-2} = O(n^{-1})$. 
With out loss of generality, assume $A_{ik_1} = p_1(t_i)x_{i,1}$, $A_{ij_1} = p_1(z_i)x_{i,2}$ and $Ex_i,1 = Ex_i,2 = 0$

$Q_{k_1j_1} = \frac{1}{n} \sum_{i=1}^{n} Ep_1(t_i)p_1(z_i)x_{i,1}x_{i,2} = C_{11} \frac{\zeta_1}{1-b_1b_2}$ so that $Q_{k_1j_1}^2 = C_{11}^2 \frac{\zeta_1^2}{(1-b_1b_2)^2}$

It is easy to check that $Ex_i,1^2 = \zeta_1^2/(1-b_1^2)$ and $Ex_i,2^2 = \zeta_2^2/(1-b_2^2)$.

$EA_{ik_1}A_{ij_1}A_{i+1k_1}A_{i+1j_1} = E(p_1(t_i)x_{i,1}p_1(z_i)x_{i,2}p_1(t_{i+1})x_{i+1,1}p_1(z_{i+1})x_{i+1,2})$

$= b_1b_2E(p_1(t_i)p_1(z_i)p_1(t_{i+1})p_1(z_{i+1}))E(x_{i,1}^2)(x_{i,2}^2)+E(p_1(t_i)p_1(z_i)p_1(t_{i+1})p_1(z_{i+1}))Ex_{i,1}x_{i,2}\delta_{i+1,1}\delta_{i+1,2}$

$= b_1b_2C_{11}^2\zeta_1^2(1-b_1^2)^{-1}\zeta_2^2(1-b_2^2)^{-1} + C_{11}^2 \frac{\zeta_1^2 \zeta_2^2}{(1-b_1b_2)^2}$.

Following the arguments of $N_{k_2,j_2}$.

Note that $C_{11}^2 \zeta_1^2(1-b_1^2)^{-1}\zeta_2^2(1-b_2^2)^{-1}[n + 2\sum_{m=1}^{n-1}(n-m)b_1^mb_2^n] = O(n)$

Therefore, $M_{k_1,j_1} = n^{-2}[C_{11}^2 \frac{\zeta_1^2 \zeta_2^2}{(1-b_1b_2)^2}n^2 + O(n)] - C_{11}^2 \frac{\zeta_1^2 \zeta_2^2}{(1-b_1b_2)^2} = O(n^{-1})$ (b) follow lemma 3 of Jiang J. (2008).

(c) $\|Q_\kappa^{-1}(Q_\kappa - \hat{Q}_\kappa)\| = Tr\{(Q_\kappa - \hat{Q}_\kappa)Q_\kappa^{-2}(Q_\kappa - \hat{Q}_\kappa)\} = Tr\{Q_\kappa^{-2}(Q_\kappa - \hat{Q}_\kappa)^2\} \leq \lambda_{max}(Q_\kappa^{-1}) \cdot \|Q_\kappa - \hat{Q}_\kappa\|^2 = O_p(\kappa^2/n)$

**Lemma A.4.** By condition (A1)-(A9)

(a) $\|n^{-1}\hat{Q}_k^{-1}\sum_{i=1}^{n} A_i b_k(i)\| = O_p(\kappa^{-2})$

(b) $\|n^{-1}\hat{Q}_k^{-1}\sum_{i=1}^{n} A_i z_i\| = O_p(\kappa^{1/2}n^{-1/2})$

Proof of lemma A.4

(a)define $\var = [b_k(1), b_k(2), \ldots, b_k(n)]^\top$ and $\Lambda = [A_1, A_2, \ldots, A_n]$, by condition (A6)

$\|n^{-1}\hat{Q}_k^{-1}\sum_{i=1}^{n} A_i b_k(i)\|^2 = \|\hat{Q}_k^{-1}\Lambda \var/n\|^2 = n^{-2} \var^\top \Lambda^\top \hat{Q}_k^{-2} \Lambda \var \leq n^{-2} \lambda_{max}(\hat{Q}_k^{-2}) \var^\top \Lambda^\top \Lambda \var = n^{-1}\lambda_{max}(\hat{Q}_k^{-2}) \var^\top \hat{Q}_k \var \leq n^{-1}\lambda_{max}(\hat{Q}_k^{-1})^2 \lambda_{max}(\hat{Q}_k) \var^\top \var = O(\kappa^{-4})$

(b)follow lemma 5 of Horowitz (2004).
Lemma A.5. By condition (A1)-(A9)

\[ \hat{B} - \theta_{k_0} = n^{-1}Q_k^{-1}\sum_{i=1}^{n} A_i \varepsilon_i + n^{-1}Q_k^{-1}\sum_{i=1}^{n} A_i b_{k_0}(i) + R_n \text{ where } \|R_n\| = O_p(\kappa^{3/2}n^{-1}) \]

Proof of lemma A.5 

Define \( M_i = \beta_0(t_i) + \{\beta_1(t_i) + \gamma(z_i)\}^{\top} x_i \) and \( \eta_i = A_i^{\top}(\hat{B} - M_i) = A_i^{\top}(\hat{B} - \theta_{k_0}) - b_{k_0}(i) \), so that \( A_i^{\top}\hat{B} = \eta_i + M_i \).

From (3.3), we know that \( \sum_{i=1}^{n} (y_i - A_i^{\top}\hat{B}) A_i = 0 \Rightarrow \sum_{i=1}^{n} (M_i + \varepsilon_i - M_i - \eta_i) A_i = 0 \)

\[ \Rightarrow \sum_{i=1}^{n} (\varepsilon_i - \eta_i) A_i = 0 \Rightarrow \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i A_i = \frac{1}{n} \sum_{i=1}^{n} A_i A_i^{\top}(\hat{B} - \theta_{k_0}) - \frac{1}{n} \sum_{i=1}^{n} A_i b_{k_0}(i) \]

\[ \Rightarrow \hat{B} - \theta_{k_0} = \frac{1}{n} \hat{Q}_k^{-1} \sum_{i=1}^{n} \varepsilon_i A_i + \frac{1}{n} \hat{Q}_k^{-1} \sum_{i=1}^{n} A_i b_{k_0}(i) \]

\[ \Rightarrow \hat{B} - \theta_{k_0} = n^{-1}Q_k^{-1}\sum_{i=1}^{n} A_i \varepsilon_i + n^{-1}Q_k^{-1}\sum_{i=1}^{n} A_i b_{k_0}(i) + n^{-1}(\hat{Q}_k^{-1} - Q_k^{-1}) \sum_{i=1}^{n} A_i \varepsilon_i + n^{-1}(\hat{Q}_k^{-1} - Q_k^{-1}) \sum_{i=1}^{n} A_i b_{k_0}(i) = J_n + J_{n_2} + J_{n_3} + J_{n_4} \]

\[ \|J_n\| = ||Q_k^{-1}(\hat{Q}_k - Q_k)n^{-1}\hat{Q}_k^{-1}\sum_{i=1}^{n} A_i \varepsilon_i|| \leq ||Q_k^{-1}(\hat{Q}_k - Q_k)|| \cdot ||n^{-1}\hat{Q}_k^{-1}\sum_{i=1}^{n} A_i \varepsilon_i|| = O_p(\kappa^{1/2}n^{-1/2}) = O_p(\kappa^{3/2}n^{-3/2}) \]

\[ \|J_{n_2}\| = ||Q_k^{-1}(\hat{Q}_k - Q_k)n^{-1}\hat{Q}_k^{-1}\sum_{i=1}^{n} A_i b_{k_0}(i)|| \leq ||Q_k^{-1}(\hat{Q}_k - Q_k)|| \cdot ||n^{-1}\hat{Q}_k^{-1}\sum_{i=1}^{n} A_i b_{k_0}(i)|| = O_p(\kappa^{1/2}n^{-1/2}) = O_p(\kappa^{3/2}n^{-3/2}) \]

Lemma A.6. \( \sqrt{\frac{n}{h}} \sum_{i=1}^{n} \varepsilon_i K_h(z_i - z)x_i \to N(0, f_x(z)\nu_0(K)\delta^2S) \)

Proof of lemma A.6 \( E(\sum_{i=1}^{n} \varepsilon_i K_h(z_i - z)x_i)^2 = E(\sum_{i=1}^{n} \varepsilon_i^2 K_h^2(z_i - z)x_i x_i^{\top} + o_p(1) \]

\[ = nh^{-1}\delta^2\nu_0(K)f_x(z)E(x_i x_i^{\top} | z_i = z) \]

Define \( F_t = \sigma(x_i, z_i, \varepsilon_{i-1}, i \leq t) \). By martingale central limit theorem, \( \sqrt{\frac{n}{h}} \sum_{i=1}^{n} \varepsilon_i K_h(z_i - z)x_i \) goes to Normal Distribution.

Lemma A.7. By condition (A1)-(A9)

(a) \( n^{-1} \sum_{i=1}^{n} x_i x_i^{\top} K_h(z_i - z)(\frac{z_i - z}{h})^{\ell} = f_z(z)\mu_{\ell}(K)S \)

(b) \( n^{-1} \sum_{i=1}^{n} R(z_i)^{\top} x_i K_h(z_i - z)(\frac{z_i - z}{h})^{\ell} x_i^{\top} = \frac{1}{2} h^2 S \gamma^{(2)}(z) f_z(z)\mu_{2+\ell} \)
Proof of lemma A.7

(a) could be easily proof by change-of-variable, the kernel theory and an application of Taylor’s expansion.

(b) Note that \( R(1)(z_i | z_i = z) = 0 \) and \( R(2)(z_i | z_i = z) = \gamma(2)(z) \).

Proof of lemma A.8

\[
\| \frac{1}{n} \sum_{i=1}^{n} K_h(z_i - z) x_i \| = O_p(1)
\]

The result holds from the same argument as for Lemma A.7 in M. Define

\[
G(z) = E \{ x_i x_{i,1} | z_i = z \} f(z), \quad \xi_i = K_h(z_i - z) x_i P_{\kappa_1}^T(t) x_{i,1}, \quad C(z) = \int E \{ x_i x_{i,1} | z_i = z \} P_{\kappa_1}^T(t) f(z) dt
\]

For each \( z \in [-C, C] \), the components of \( C(z) \) include the Fourier coefficients of a function that is bounded uniformly over \( z \). Therefore, by Bessel’s inequality, there exists some finite constant \( M \) for all \( \kappa \), such that \( C^\top(z) C(z) \leq M \).

The arguments similar to those used to prove \( E \| r_{n1} \|^2 = \frac{1}{n^2} \{ E \| \sum_{i=1}^{n} \xi_i \|^2 - E \| \xi_i \|^2 \} = O_p(\frac{n}{nh}) = O_p(1) \)

By the definitions of \( C(z) \) and \( \xi_i \),

\[
r_{n2} = E K_h(z_i - z) x_i P_{\kappa_1}^T(t) x_{i,1} - \int G(z) P_{\kappa_1}^T(t) dt = \int [ \int \{ G(z + \mu h) K(\mu) - G(z) K(\mu) \} d\mu ] P_{\kappa_1}^T(t) dt
\]
\[ f \left[ \int \{ \frac{\partial G(z+\Delta)}{\partial z} \mu hK(\mu) \} d\mu \right] P_{n1}^T(t) dt \text{(Dominated convergence theorem)} = h \int \frac{\partial G(z)}{\partial z} P_{n1}^T(t) dt (1 + o_p(1)) \text{ where } \Delta \text{ is between 0 and } \mu h. \text{ Therefor, we obtain that } \|r_{n2}\|^2 = O(\kappa h^2) = O_P(1).

so that \( \frac{1}{n} \sum_{i=1}^{n} K_h(z_i - z) x_i P_{n1}^T(t)x_{i,1} = C(z) + r_{n1} + r_{n2} = O_P(1) \)

**Lemma A.9.** \( \frac{1}{n} \sum_{i=1}^{n} \{ \hat{\beta}_0^*(t_i) + \hat{\beta}_1^*(t_i) \top x_i - \beta_0(t_i) - \beta_1(t_i) \top x_i \} K_h(z_i - z) x_i = o_p(h^2) \)

**Proof of Lemma A.9**

By Lemma A.5, \( \frac{1}{n} \sum_{i=1}^{n} \{ \hat{\beta}_0^*(t_i) + \hat{\beta}_1^*(t_i) \top x_i - \beta_0(t_i) - \beta_1(t_i) \top x_i \} K_h(z_i - z) x_i \)

\[ = \frac{1}{n} \sum_{i=1}^{n} \{ \bar{\mathcal{A}}(t_i) \top \hat{B} - \beta_0(t_i) - \beta_1(t_i) \top x_i \} K_h(z_i - z) x_i \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \{ \bar{\mathcal{A}}(t_i) \top (\hat{B} - \theta_{k0}) - \bar{b}_{k0}(i) \} K_h(z_i - z) x_i \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \left[ x_i \bar{\mathcal{A}}(t_i) \top K_h(z_i - z) \left( \frac{1}{n} Q_k^{-1} \sum_{j=1}^{n} \varepsilon_j A_j \right) \right] + \frac{1}{n} \sum_{i=1}^{n} \left[ x_i \bar{\mathcal{A}}(t_i) \top K_h(z_i - z) \frac{1}{n} Q_k^{-1} \sum_{j=1}^{n} A_j b_{k0}(j) \right] - \frac{1}{n} \sum_{i=1}^{n} \left[ x_i \bar{b}_{k0}(i) K_h(z_i - z) \right] + \frac{1}{n} \sum_{i=1}^{n} \left[ x_i \bar{\mathcal{A}}(t_i) \top K_h(z_i - z) R_n \right] = C_{n1} + C_{n2} + C_{n3} + C_{n4} \]

Arguments like those used to prove Lemma 7 in Horowitz (2004) show that, \( \|C_{n1}\| = o_p(h^2) \)

Arguments like those used to prove Lemma A.4 show that \( E \left\| \frac{1}{n} Q_k^{-1} \sum_{j=1}^{n} A_j b_{k0} \right\|^2 = O(\kappa^{-4}) \), by Lemma A.8, \( \|C_{n2}\| \leq \left\| \frac{1}{n} \sum_{i=1}^{n} x_i \bar{\mathcal{A}}(t_i) \top K_h(z_i - z) \right\| \left\| \frac{1}{n} Q_k^{-1} \sum_{j=1}^{n} A_j b_{k0} \right\| \leq O_p(\kappa^{-2}) = o_p(h^2) \)

\( \|C_{n3}\| \leq \frac{1}{n} \sum_{i=1}^{n} \|x_i\| K_h(z_i - z) \max \bar{b}_{k0}(i) = O_p(1) O(\kappa^{-2}) = O_p(\kappa^{-2}) = o_p(h^2) \)

\( \|C_{n4}\| \leq \left\| \frac{1}{n} \sum_{i=1}^{n} x_i \bar{\mathcal{A}}(t_i) \top K_h(z_i - z) \right\| \|R_n\| = O_p(\kappa^2/n) = o_p(h^2) \)

**Proof of Theorem 4.1**

To simplify the notation,

Recall \( R(z_i) = \gamma(z_i) - \gamma(z) - \gamma^{(1)}(z)(z_i - z), W_{ih}(z) = \begin{pmatrix} x_i \\ x_i (z_i - z) / h \end{pmatrix}, \)
Define $\Delta_{\beta_0, \beta_1} = \beta_0(t_i) + \beta_1(t_i)^T x_i - \tilde{\beta}_0(t_i) - \tilde{\beta}_1(t_i)^T x_i$, then

$$\hat{\eta}_i = R(z_i) x_i - W_{ih}(z) \Phi + \Delta_{\beta_0, \beta_1},$$

then $W_{ih}(z) \Phi = x_i^T \hat{\gamma}(z) - x_i^T \gamma(z) + x_i^T \hat{\gamma}^{(1)}(z)(z_i - z) - x_i^T \gamma^{(1)}(z)(z_i - z)$

and

$$\hat{\eta}_i = \{x_i^T \gamma(z_i) - x_i^T \gamma(z) - x_i^T \gamma^{(1)}(z)(z_i - z)\} \Phi = \{\beta_0(t_i) + \beta_1(t_i)^T x_i - \tilde{\beta}_0(t_i) - \tilde{\beta}_1(t_i)^T x_i\} \Phi$$

From 3.3, by taking the first derivative, we have

$$\sum_{i=1}^n \{y_i - \tilde{\beta}_0(t_i) - \tilde{\beta}_1(t_i)^T x_i - x_i^T \hat{\gamma}(z) - x_i^T \hat{\gamma}^{(1)}(z)(z_i - z)\} W_{ih}(z) K_h(z_i - z) = 0$$

$$\Rightarrow \sum_{i=1}^n \{\varepsilon_i + \beta_0(t_i) + \beta_1(t_i)^T x_i + \gamma(z_i)^T x_i - \hat{\beta}_0(t_i) - \hat{\beta}_1(t_i)^T x_i - x_i^T \hat{\gamma}(z) - x_i^T \hat{\gamma}^{(1)}(z)(z_i - z)\} W_{ih}(z) K_h(z_i - z) = 0$$

$$\Rightarrow \sum_{i=1}^n (\varepsilon_i + \tilde{\eta}_i) K_h(z_i - z) W_{ih}(z) = 0$$

$$\Rightarrow 0 = \sum_{i=1}^n \varepsilon_i K_h(z_i - z) W_{ih}(z) + \sum_{i=1}^n \tilde{\eta}_i K_h(z_i - z) W_{ih}(z) = I_{n_1} + I_{n_2}$$

where

$$I_{n_2} = \sum_{i=1}^n \tilde{\eta}_i K_h(z_i - z) W_{ih}(z) = \sum_{i=1}^n R(z_i)^T x_i K_h(z_i - z) W_{ih}(z) - \sum_{i=1}^n W_{ih}(z) W_{ih}(z)^T \Phi K_h(z_i - z) W_{ih}(z) = \sum_{i=1}^n \beta_0(t_i) + \beta_1(t_i)^T x_i - \hat{\beta}_0(t_i) - \hat{\beta}_1(t_i)^T x_i \} K_h(z_i - z) W_{ih}(z) = L_{n_1} - L_{n_2} + L_{n_3}$$

$$\Rightarrow -n^{-1} L_{n_2} - n^{-1} L_{n_3} = n^{-1} I_{n_1}$$

From Lemma A.7.b, $n^{-1} L_{n_1} = n^{-1} \sum_{i=1}^n R(z_i)^T x_i K_h(z_i - z) W_{ih}(z) = \frac{h^2}{2} f_z(z) \gamma(2)(z) S \begin{pmatrix} \mu_2(K) \\ \mu_3(K) \end{pmatrix}$

$$\Rightarrow n^{-1} \sum_{i=1}^n R(z_i)^T x_i K_h(z_i - z) x_i = \frac{h^2}{2} f_z(z) \gamma(2)(z) \mu_2(K) S$$

From Lemma A.7.a, by noting that $\mu_0(K) = 1$, and $\mu_1(K) = 0$,

$$n^{-1} L_{n_2} = n^{-1} \sum_{i=1}^n W_{ih}(z) W_{ih}(z)^T K_h(z_i - z) \Phi$$

$$= \left(n^{-1} \sum_{i=1}^n \begin{pmatrix} x_i x_i^T & x_i x_i^T (\tilde{z} - \bar{z}) \\ x_i x_i^T (\tilde{z} - \bar{z}) & x_i x_i^T (\tilde{z} - \bar{z})^2 \end{pmatrix} K_h(z_i - z) \right) \Phi = f_z(z) \begin{pmatrix} S & 0 \\ 0 & \mu_2(K) S \end{pmatrix} \Phi$$
From Lemma A.9, by condition A(8), \( n^{-1}L_{n_3} = o_p(h^2) = o_p(n^{-1}L_{n_1}) \)

From Lemma A.6, \( \sqrt{\frac{h}{n}}I_{n_1} = \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \varepsilon_i K_h(z_i - z)W_{ih}(z) \to N(0, f_z(z)\delta^2\Sigma_\nu) \)

\[ \Rightarrow \sqrt{\frac{h}{n}} \sum_{i=1}^{n} \varepsilon_i K_h(z_i - z)x_i \to N(0, f_z(z)\delta^2\nu_0(K)S) \]

so that \( \sqrt{\frac{h}{n}}L_{n_2} - \sqrt{\frac{h}{n}}L_{n_1} = \sqrt{\frac{h}{n}}I_{n_1} \)

Hence, \( \sqrt{nh}(\hat{\gamma}(z) - \gamma(z) - \frac{h^2}{2} \mu_2(K)\gamma^{(2)}(z)) \to N(0, f_z(z)^{-1}\delta^2 S^{-1}\nu_0(K)) \)

Proof of Theorem 4.2 and Theorem 4.3

Following the same argument as the proof of Theorem 4.1, we have \( \sqrt{nh_2[\hat{\beta}_1(t) - \beta_1(t) - \frac{h^2}{2} \mu_2(K)\beta_1^{(2)}(t)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, \delta^2 S^{-1}\nu_0(K)\} \) and \( \sqrt{nh_0[\hat{\beta}_0(t) - \beta_0(t) - \frac{h^2}{2} \mu_2(K)\beta_0^{(2)}(t)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, \delta^2 \nu_0(K)\} \)
APPENDIX B: Sketch of Proofs

Theorem 5.1

Proof of Theorem 5.1

Lemma A.10. (a) \( \|\hat{Q}_\kappa^* - Q_\kappa^*\|_2^2 = O_p(\kappa^2/n) \)

(b) \( \|\hat{Q}_\kappa^{*-1}\|_2^2 = O_p(\kappa), \|Q_\kappa^{*-1}\|_2^2 = O_p(\kappa) \)

(c) \( \|Q_\kappa^{*-1}(Q_\kappa^* - \hat{Q}_\kappa^*)\|_2^2 = O_p(\kappa^2/n) \)

Proof of above lemma

(a)(b)(c) hold from the same argument as for Lemma A.3.

Lemma A.11. By condition (B1)-(B8)

(a) \( \|n^{-2}\hat{Q}_k^{*-1}\sum_{i=1}^n (A_i b_{k_0}(i))\| = O_p(\kappa^{-2}/n) \)

(b) \( \|n^{-2}\hat{Q}_k^{*-1}\sum_{i=1}^n (A_i \varepsilon_i)\| = O_p(\kappa^{1/2}/n) \)

Proof of lemma A.11

(a) define \( \varrho = [b_{k_0}(1), b_{k_0}(2), \ldots, b_{k_0}(n)]^\top \) and \( \Lambda = [A_1, A_2, \ldots, A_n] \), by condition (B3)

\[
\|n^{-2}\hat{Q}_{k}^{*-1}\sum_{i=1}^{n}\{A_{i}b_{k_0}(i)\}\|^2 = \|\hat{Q}_{k}^{*-1}\Lambda \varrho/n^2\|^2 = n^{-4} \varrho^\top \Lambda \hat{Q}^{*-2}_{k} \Lambda \varrho \leq n^{-4} \lambda_{\max}(\hat{Q}^{*-2}_{k}) \varrho^\top \Lambda \Lambda^\top \Lambda \varrho = n^{-2} \lambda_{\max}(\hat{Q}^{*-2}_{k}) \varrho^\top \hat{Q}^{*-1}_{k} \Lambda \varrho \leq n^{-2} \lambda_{\max}(\hat{Q}^{*-2}_{k})^2 \lambda_{\max}(\hat{Q}_{k}) \varrho^\top \varrho = O(\kappa^{-4}n^{-2})
\]

(b) hold from the same argument as for Lemma A.4.

Lemma A.12. By condition (B1)-(B8)

\[
\hat{B} - \theta_{k_0} = n^{-2}Q_{k}^{*-1}\sum_{i=1}^{n} A_{i}\varepsilon_{i} + n^{-2}Q_{k}^{*-1}\sum_{i=1}^{n} A_{i}b_{k_0} + R_{n} \text{ where } \|R_{n}\| = O_p(\kappa^{3/2}/n^{3/2})
\]

Proof of above lemma
This result holds from the same argument as for Lemma A.5.

\[ \hat{B} - \theta_{k_0} = -Q^{-1}_\kappa \sum_{i=1}^n A_i \xi_i + n^{-2} Q^{-1}_\kappa \sum_{i=1}^n A_i b_{k_0}(i) + n^{-2} (\hat{Q}^{-1}_\kappa - Q^{-1}_\kappa) \sum_{i=1}^n A_i \xi_i + n^{-2} (\hat{Q}^{-1}_\kappa - Q^{-1}_\kappa) \sum_{i=1}^n A_i b_{k_0}(i) = J_{n_1} + J_{n_2} + J_{n_3} + J_{n_4} \]

\[ \| J_{n_3} \| = \| Q^{-1}_\kappa (\hat{Q}^{-1}_\kappa - Q^{-1}_\kappa) \sum_{i=1}^n A_i \xi_i \| < \| Q^{-1}_\kappa (\hat{Q}^{-1}_\kappa - Q^{-1}_\kappa) \| \| n^{-2} \hat{Q}^{-1}_\kappa \sum_{i=1}^n A_i \xi_i \| = O_p(\kappa/n^{1/2})O_p(\kappa^{1/2}/n) = O_p(\kappa^{3/2}/n^{3/2}) \]

\[ \| J_{n_4} \| = \| Q^{-1}_\kappa (\hat{Q}^{-1}_\kappa - Q^{-1}_\kappa) \sum_{i=1}^n A_i b_{k_0}(i) \| < \| Q^{-1}_\kappa (\hat{Q}^{-1}_\kappa - Q^{-1}_\kappa) \| \| n^{-2} \hat{Q}^{-1}_\kappa \sum_{i=1}^n A_i b_{k_0}(i) \| = O_p(\kappa/n^{1/2})O_p(\kappa^{2}/n) = O_p(\kappa^{-1}/n^{3/2}) \]

**Lemma A.13.** By condition (B1)-(B8)

(a) \( \frac{1}{n^2} \sum_{i=1}^n x_i x_i^T R(z_i)^T K_{h_1}(z_i - z) = \frac{h_1^2}{2} f_z(z) W_{\delta, \gamma} (z) \mu_2(K) \{ 1 + o_p(1) \} \)

(b) \( \frac{1}{n^2} \sum_{i=1}^n x_i x_i^T K_{h_1}(z_i - z) (\frac{z_i - z}{h_1}) = f_z(z) \mu_j(K) W_{\delta, \gamma} + o_p(1) \)

Proof of lemma A.13

See the proof of Theorem 2.1 in Cai and Park (2009).

**Lemma A.14.** By condition (B2)(i) \( \frac{1}{n^2} \sum_{i=1}^n (\beta_0(t_i) + \beta_1(t_i)^T x_i - \beta_0^* (t_i) - \beta_1^* (t_i)^T x_i) K_{h_1}(z_i - z) x_i = o_p(h_1^2) \)

Proof of lemma A.14 \( \frac{1}{n^2} \sum_{i=1}^n (\beta_0(t_i) + \beta_1(t_i)^T x_i - \beta_0^* (t_i) - \beta_1^* (t_i)^T x_i) K_{h_1}(z_i - z) x_i = \frac{1}{n^2} \sum_{i=1}^n [x_i \bar{A}_i^T K_{h_1}(z_i - z) (\frac{1}{n} Q_{\kappa}^{-1} \sum_{j=1}^n \xi_j A_j)] + \frac{1}{n^2} \sum_{i=1}^n [x_i \bar{A}_i^T K_{h_1}(z_i - z) (\frac{1}{n} Q_{\kappa}^{-1} \sum_{j=1}^n A_j b_{k_0})]- \frac{1}{n^2} \sum_{i=1}^n [x_i \bar{b}_{k_0}(i) K_{h_1}(z_i - z)] + \frac{1}{n^2} \sum_{i=1}^n [x_i \bar{A}_i^T K_{h_1}(z_i - z) R_n] = C_{n_1} + C_{n_2} + C_{n_3} + C_{n_4} \)

By condition (B2)(i), note that \( \frac{1}{n^2} \sum_{i=1}^n x_i x_i^T K_{h_1}(z_i - z) = f_z(z) \mu_0(K) W + o_p(1) = O_p(1) \) from Cai and Park (2009), arguments like those used to prove Lemma 7 in Horowitz (2004) show that, \( \| C_{n_1} \| = o_p(h_1^2) \)

Arguments like those used to prove Lemma A.4 show that \( E \| \frac{1}{n^2} Q_{\kappa}^{-1} \sum_{j=1}^n A_j b_{k_0}(i) \|^2 = O(\kappa^{-4}/n^2) \) \( \| C_{n_2} \| \leq \| \frac{1}{n^2} \sum_{i=1}^n x_i \bar{A}_i^T K_{h_1}(z_i - z) \| \| \frac{1}{n^2} Q_{\kappa}^{-1} \sum_{j=1}^n A_j b_{k_0}(i) \| = O_p(\kappa^{1/2})O_p(\kappa^{-2}/n) = O_p(\kappa^{3/2}/n^{3/2}) \)
Proof of Theorem 5.1

Following the proof of Theorem 2.1 in Cai and Park (2009), by Lemma A.13 and A.14, we can easily show that

\[ n\sqrt{h_1} [\hat{\gamma}(z) - \gamma(z) - \frac{h_1^2}{2} \mu_2(K) r^{(2)}(z) \{1+o_p(1)\}] = W_{\delta, 2}^{-1} f_z(z)^{-1/2} \sqrt{\nu_0(K)} \int_0^1 W_\delta(r) dW_\epsilon(r) \]

Theorem 5.2
Lemma A.16. $\frac{1}{n^2} \sum_{i=1}^{n} x_i x_i^\top K_{h_2}(t_i - t)(\frac{t_i - t}{h_2})^j = \mu_j(K) W_{\delta,2} + o_p(1)$ for $j = 0, 1, 2$

Proof of lemma A.16

See the proof of Theorem 2.1 in Cai and Park (2009).

Lemma A.17. $\frac{\sqrt{n}}{n^2} \sum_{i=1}^{n} \varepsilon_i x_i K_{h_2}(t_i - t) \xrightarrow{d} \sqrt{\nu_0(K)} \int_0^1 W_{\delta}(r) dW_{\varepsilon}(r)$

Proof of above lemma

See the proof of Theorem 2.1 in Cai and Park (2009).

Lemma A.18. By condition (B2)(i) $\frac{1}{n^2} \sum_{i=1}^{n} (\beta_0(t_i) + \gamma(z_i) x_i - \tilde{\beta}_0^* (t_i) - \tilde{\gamma}^*(z_i) x_i) K_{h_2}(t_i - t)x_i = o_p(h_2^2)$

Proof of above lemma

See the proof of Lemma A.14.

Proof of Theorem 5.2

Theorem 5.2 could be derived by following the same procedure of proof of theorem 5.1.

Theorem 5.3

Lemma A.19. By condition (B1) $\| \frac{1}{n^\frac{1}{2}} \sum_{i=1}^{n} \kappa_1(t_i)x_i K_{h_0}(t_i - t) \| = O_p(1)$

Proof of lemma A.19

By condition,

$$\sup_{0 \leq r \leq 1} \| x_{[n]} / \sqrt{n} - W_{\delta}(r) \| = O(n^{-\theta} \log^\lambda(n)) = o_p(1)$$

, see Theorem 4.1 in Shao (1987) and Einmahl (1987) for details.
By the same argument in Lemma 8, it is easy to show that \( \| \frac{1}{n} \sum_{i=1}^{n} \kappa_1(t_i) K_{h_0}(t_i - t) \| = O_p(1) \), 
\[
\| \frac{1}{n^{3/2}} \sum_{i=1}^{n} \kappa_1(t_i) x_{i,1} K_{h_0}(t_i - t) \| = \| \frac{1}{n} \sum_{i=1}^{n} \kappa_1(t_i) \frac{\tilde{\beta}_1(t_i) - \beta_1^*(t_i)}{\sqrt{n}} K_{h_0}(t_i - t) \| \leq \| \frac{1}{n} \sum_{i=1}^{n} \kappa_1(t_i) W_\delta(t_i) K_{h_0}(t_i - t) \| + \| \frac{1}{n} \sum_{i=1}^{n} \kappa_1(t_i) \{ \tilde{\beta}_1(t_i) - W_\delta(t_i) \} K_{h_0}(t_i - t) \| \leq \| \frac{1}{n} \sum_{i=1}^{n} \kappa_1(t_i) K_{h_0}(t_i - t) \| \sup \| \frac{\tilde{\beta}_1}{\sqrt{n}} - W_\delta(t_i) \| = O_p(1) + o_p(1) = O_p(1)
\]

Lemma A.20. By condition (B1) \( \frac{1}{n} \sum_{i=1}^{n} (\beta_1(t_i)^\top x_{i} - \tilde{\beta}_1^*(t_i)^\top x_{i} + \gamma_1(t_i)^\top x_{i} - \tilde{\gamma}_1^*(t_i)^\top x_{i}) K_{h_0}(z_i - z) = o_p(h_0^2) \)

Proof of above lemma

Define \( \overline{A}(t_i, z_i) = [\kappa_0(t_i)^\top \cdot 0, \kappa_1(t_i)^\top x_{i,1}, \kappa_1(z_i)^\top x_{i,1}, \kappa_1(t_i)^\top x_{i,2}, \kappa_1(z_i)^\top x_{i,2}, \cdots, \kappa_1(t_i)^\top x_{i,p}, \kappa_1(z_i)^\top x_{i,p}]^\top \) By the same argument in Lemma A.9, \( \frac{1}{n} \sum_{i=1}^{n} (\beta_1(t_i)^\top x_{i} - \tilde{\beta}_1^*(t_i)^\top x_{i} + \gamma_1(t_i)^\top x_{i} - \tilde{\gamma}_1^*(t_i)^\top x_{i}) K_{h_0}(z_i - z) = \frac{1}{n} \sum_{i=1}^{n} [\overline{A}(t_i, z_i)^\top K_{h_0}(t_i - t) (\frac{1}{n^2} Q_{-1}^{k-1} \sum_{j=1}^{n} A_j b_{k0}(j))] + \frac{1}{n} \sum_{i=1}^{n} [\overline{A}(t_i, z_i)^\top K_{h_0}(t_i - t) N_n] = C_{n1} + C_{n2} + C_{n3} + C_{n4}

Arguments like those used to prove Lemma 7 in Horowitz (2004) show that, \( \| C_{n1} \| = o_p(h_0^2) \)

By Lemma A.11 and A.19, \( \| C_{n2} \| = \| \frac{1}{n} \sum_{i=1}^{n} \overline{A}(t_i, z_i)^\top K_{h_0}(t_i - t) \| \| \frac{1}{n^2} Q_{-1}^{k-1} \sum_{j=1}^{n} A_j b_{k0}(j) \| = O_p(n^{1/2} \kappa^{1/2}) O_p(\kappa^{-2} n^{-1}) = o_p(h_0^2) \)

By the same argument, it can be shown that \( \| C_{n3} \| = O_p(\kappa^{-2}) = o_p(h_0^2) \)

\( \| C_{n4} \| = O_p(n^{1/2} \kappa^{1/2}) O_p(\kappa^{-2} n^{-3/2}) = O_p(\frac{\kappa^2}{n}) = o_p(h_0^2) \)

Proof of Theorem 5.3

It is easily to check that \( \frac{1}{n} \sum_{i=1}^{n} K_{h_0}(t_i - t) \rightarrow 1, \frac{1}{n} \sum_{i=1}^{n} K_{h_0}(t_i - t) \beta_0(t_i) \rightarrow \beta_0(t) + \frac{1}{2} h_0^2 \beta_0^{(2)}(t) + o_p(h_0^2), \) under (B2)(ii) \( \frac{1}{nh_0^2} \sum_{i=1}^{n} (\beta_1(t_i)^\top x_{i} - \tilde{\beta}_1^*(t_i)^\top x_{i} + \gamma(z_i)^\top x_{i} - \tilde{\gamma}_1^*(t_i)^\top x_{i}) K_{h_0}(z_i - z) = o_p(h_0^2) \)

\[
\frac{1}{nh_0^2} \sum_{i=1}^{n} (\beta_1(t_i)^\top x_{i} - \tilde{\beta}_1(t_i)^\top x_{i} + \gamma(z_i)^\top x_{i} - \tilde{\gamma}_1(t_i)^\top x_{i}) K_{h_0}(z_i - z) = o_p(h_0^2)
\]
\[ \tilde{\gamma}(z_i) x_i K_{h_0}(t_i - t) = o_p(1). \]

Following the proof of Theorem 5.1, we can easily show Theorem 5.3.