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 asymptotics 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.

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

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Nonlinear effects and time-varying effects exist 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 coefficients change 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 motivate us to consider the following time-varying coefficient model,

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

for fitting financial data, where Y_t and ϵ_t are scalar, $X_t = (x_{t_1}, \ldots, x_{t_d})^\top$ is a vector of covariates with dimension d and $\beta(\cdot)$ is a $d \times 1$ vector function.

It is well known that many variables in financial markets are nonstationary. People are interested in how models could be built for 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 an amount of research attentions 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. That is, the cointegrating parameters are constant in the cointegration model introduced by Engle (1987).

A general conclusion of empirical studies is that constant cointegration relationships cannot 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.

A more general set-up for a class of cointegration models is the following model:

$$Y_t = \gamma(z_t)^\top X_t + \epsilon_t, \tag{1.2}$$

where Y_t , z_t and ϵ_t are scalar, $X_t = (x_{t_1}, \ldots, x_{t_d})^\top$ is a vector of covariates with dimension d and $\beta(\cdot)$ is a $d \times 1$ vector function.

Cai and Park (2009) considered model (1.2) for nonstationary time series data. Xiao (2009) also considered (1.2) for nonstationary time series data and focused on inference procedures for 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 log "difference" of 5 year daily Treasury bond yield rate and 6 month daily Treasury bill yield rate (z_t). We estimate $\beta(\cdot)$ functions in model (1.1) by running local linear smoother and show them in the top panels in Figure 1. We estimate $\gamma(\cdot)$ functions in model (1.2) also by running local linear smoother and show them in the bottom panels in Figure 1. Figure 2 displays estimated residuals from the two models. Both residuals are stationary according to the ADF test in Table 1. This indicates that both time-varying effects and nonlinearity effects are found here. Naturally, one would ask: "Which model is better? Which effect is true? Do these two effects exist in the relationship between the predictors and response?" To address these important questions, we propose the following model:

$$y_i = \beta_0(t_i) + \gamma_0(z_i) + \{\beta_1(t_i) + \gamma_1(z_i)\}^\top x_i + \varepsilon_i. \tag{1.3}$$

Since model (1.3) includes models (1.1) and (1.2) as specific examples, it can be used to validate if models (1.1) and (1.2) are appropriate for fitting the above data.

We fit model (1.3) with real data and compare the goodness of fit with models (1.1) and (1.2) in the Chapter 8.

We propose a two-step estimation method to estimate the time-varying and nonlinear coefficients for stationary or nonstationary explanatory variables. We show that our estimators are "oracle" in the sense that the asymptotic distribution of the estimator of one coefficient function is the same as if other coefficient functions are known.

The rest of this dissertation is organized as follows. In Chapter 2 we show the model we consider in this dissertation. In Chapter 3 we give a brief introduction of the two-step estimation procedure. In Chapter 4 we consider the case when x_i is stationary. The asymptotic results for stationary x_i are showed here. In Chapter 5 we consider the case when x_i is nonstationary. The asymptotic results for nonstationary x_i are showed here. In Chapter 6 we run simulation for both stationary x_i and nonstationary

x_i. In Chapter 7 we consider a real example. Concluding remarks are presented in Chapter 8. Proofs are contained in the Appendix.

Table 1: ADF test for estimated residuals in model (1.1) and estimated residuals in model (1.2)

	ADF Test Statistic	P value
Estimated residuals in model (1.1)	-5.6321	< 0.01
Estimated residuals in model (1.2)	-4.311	< 0.01

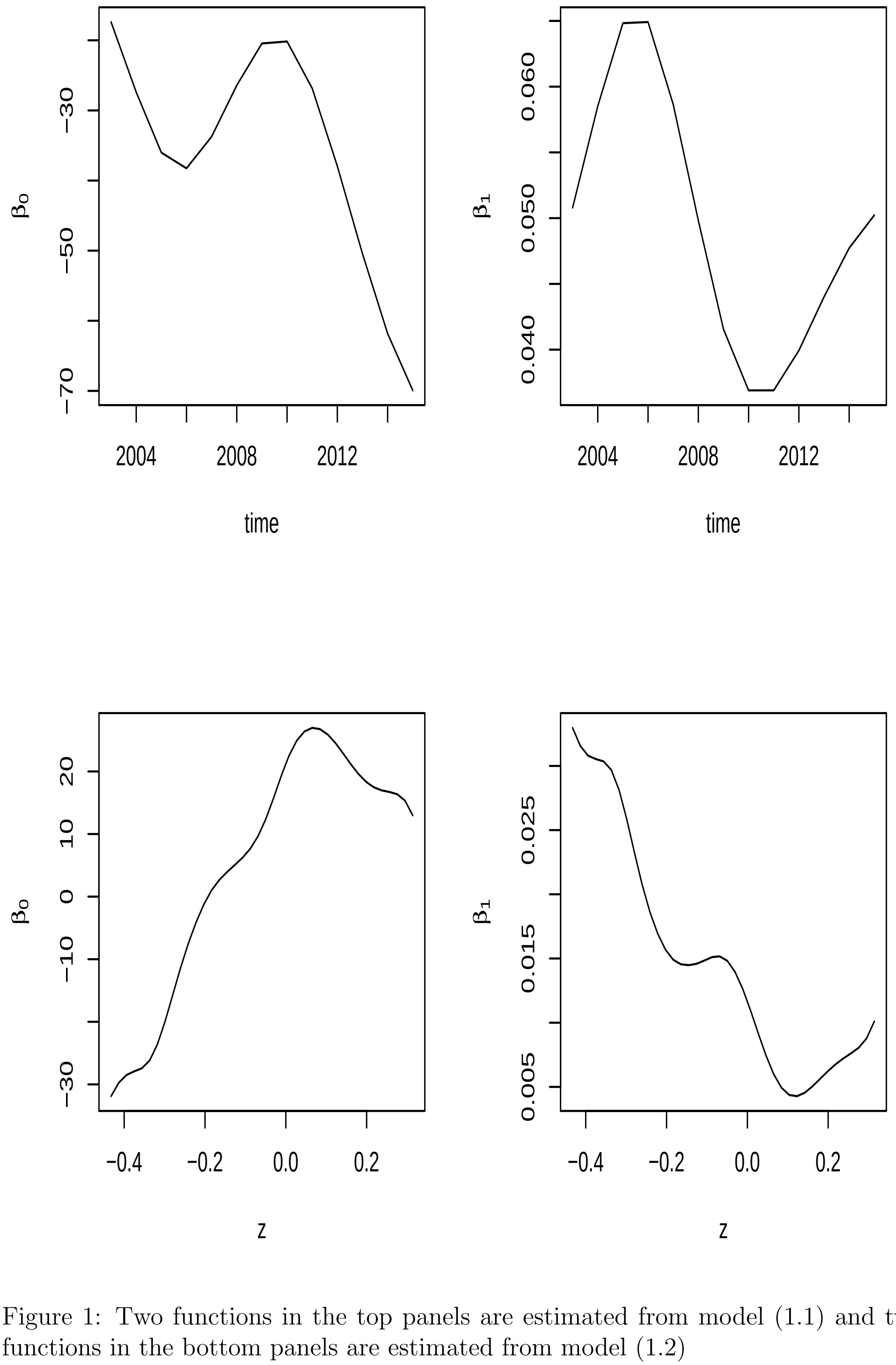
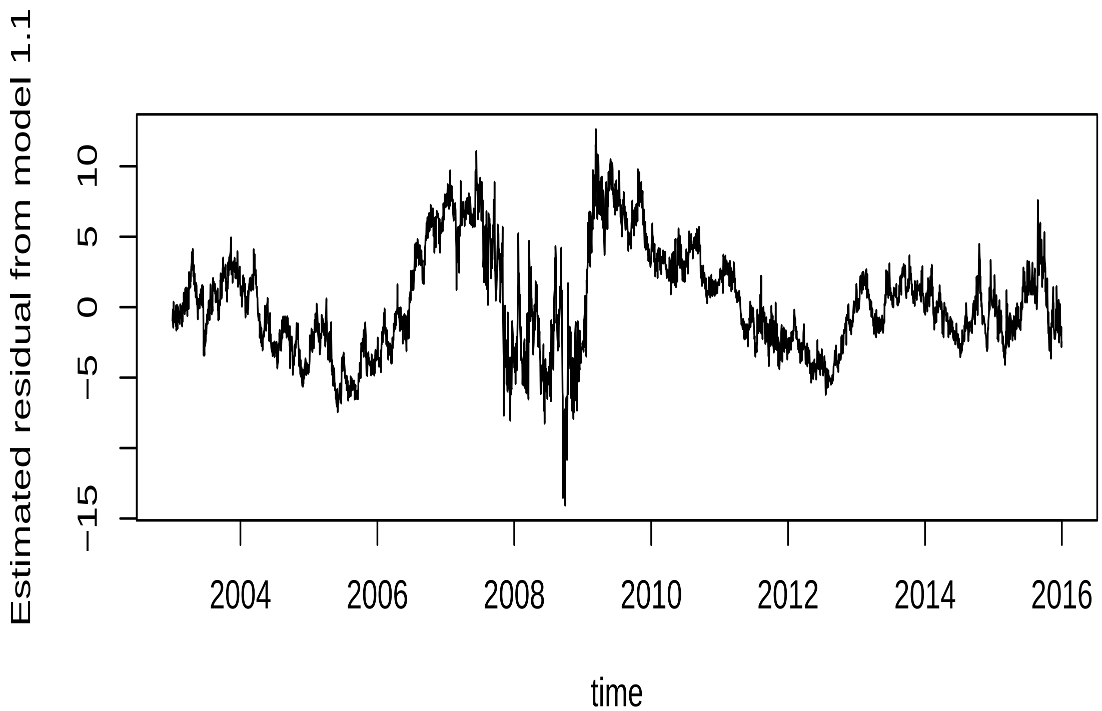


Figure 1: Two functions in the top panels are estimated from model (1.1) and two functions in the bottom panels are estimated from model (1.2)



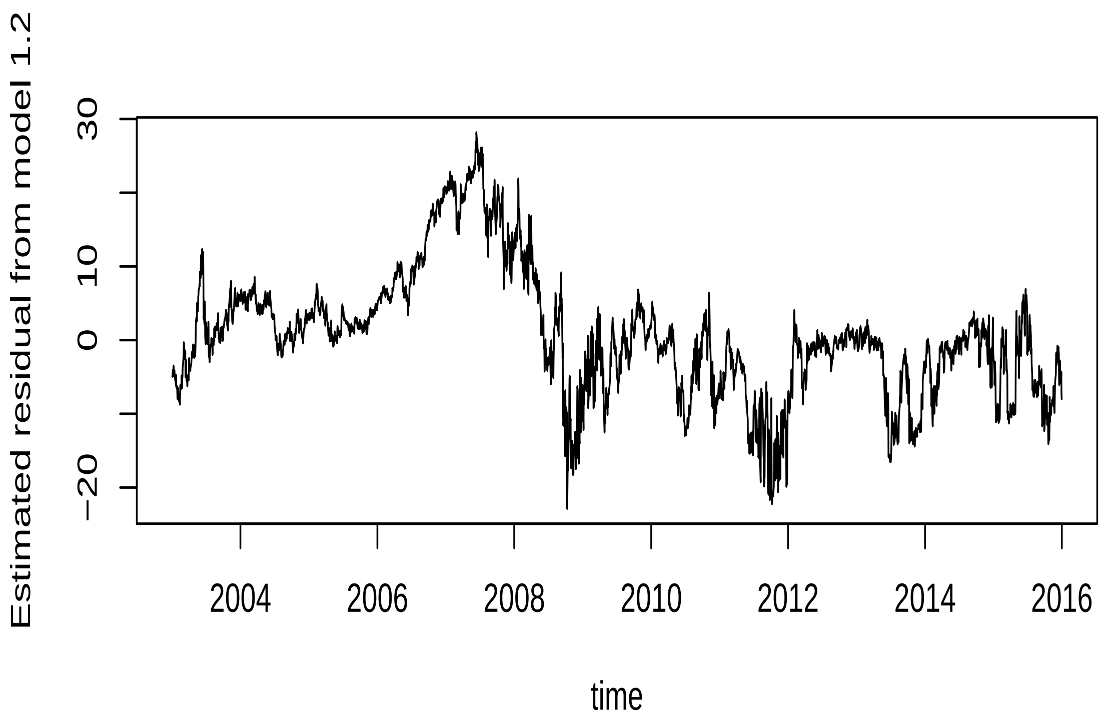


Figure 2: Top panel shows residuals from model (1.1) and bottom panel shows residuals from model (1.2)

CHAPTER 2: MODEL WITH TIME VARYING AND NONLINEAR EFFECTS

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

$$y_i = \beta_0(t_i) + \gamma_0(z_i) + \{\beta_1(t_i) + \gamma_1(z_i)\}^\top x_i + \varepsilon_i. \tag{2.1}$$

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 $I(0)$. $E(\varepsilon_i|x_i, z_i) = 0$. $var(\varepsilon_i|x_i, z_i) = \delta^2$. $\beta_0(t_i)$ and $\gamma_0(z_i)$ both are scalar. $\beta_1(t_i)$ and $\gamma_1(z_i)$ both are $p \times 1$ function vectors. ε_i is a strictly α-mixing stationary process. We assume $E[\gamma_0(z_i)] = E[\gamma_1(z_i)] = 0$ for identifiability.

When x_i and y_i both are nonstationary and ε_i is stationary, we say that x_i and y_i are cointegrated with a varying coefficient cointegration vector $\beta_1(t_i) + \gamma_1(z_i)$, which are function vectors of time t_i and smooth functions of z_i. This setting is more general than the usual assumption that the cointegration vector is constant.

CHAPTER 3: ESTIMATION

Horowitz (2004) considered the nonparametric estimation of an additive model with a link function, they proposed a two-step estimation method to estimate the unknown link function. In the first step, least squares is used to obtain a series approximation to each unknown function. The first-step estimators are inputs to the second stage. But this method was limited on additive model and i.i.d. variables. Cai and Park (2009) showed local linear smoother could be used to estimate the unknown functions even though the independent variables and dependent variable were nonstationary. Their estimators had good properties. Xiao (2009) showed local polynomial could be used to estimate the unknown functions in the same situation. We can estimate all our unknown functions by local linear smooth or local polynomial method at the same time, however, the convergent rate of estimators will be slow. If we know the functions of t_i, we can estimate the unknown functions of z_i by local linear smoother with a fast convergent rate. If we know functions of z_i, we can estimate the unknown functions of t_i by local linear smoother with a fast convergent rate. That is the basic idea of our two-step estimation method. If we have good estimators in the first-step estimation, we should expect estimators from local linear smoother 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 $\mathcal{Z} = [-1, 1]$. We assume $E\gamma_0(z) = E\gamma_1(z) = 0$ so that we could identify $\gamma_0(\cdot)$ and $\gamma_1(\cdot)$. Let $\{p_k(\cdot),\ k = 1, 2, \ldots\}$ be a standard orthogonal basis for smooth functions on $[-1, 1]$ which satisfy $\int_{-1}^{1} p_k(x)dx = 0$ and

$$\int_{-1}^{1} p_k(x)p_j(x)\,dx = \begin{cases} 1 & \text{if } k = j; \\ 0 & \text{otherwise.} \end{cases}$$

One choice of the orthogonal basis is orthogonal spline basis. Let $P_\kappa(t, z) = \big[1, p_1(t), \ldots, p_\kappa(t), p_1(z), \ldots, p_\kappa(z)\big]^\top$ and $\Theta_{\kappa b} = (\theta_{0,b}, \theta_{11,b}, \ldots, \theta_{1\kappa,b}, \theta_{21,b}, \ldots, \theta_{2\kappa,b})^\top$ for $b = 0, 1, \ldots, p$. Then $P_\kappa^\top(t, z)\Theta_{\kappa 0}$ is a series approximation to $\beta_0(t) + \gamma_0(z)$, and $P_\kappa^\top(t, z)\Theta_{\kappa d}$ is a series approximation to the dth component of $\beta_1(t) + \gamma_1(z)$ for $d = 1, \ldots, p$. Our orthogonal series estimator of $\beta_0(t) + \gamma_0(z)$ and the dth component $\beta_{1d}(t) + \gamma_{1d}(z)$ of $\beta_1(t) + \gamma_1(z)$ are respectively defined as

$$\tilde\beta_0(t) + \tilde\gamma_0(z) = P_\kappa^\top(t, z)\hat\Theta_{\kappa 0} \ \text{ and } \ \tilde\beta_{1d}(t) + \tilde\gamma_{1d}(z) = P_\kappa^\top(t, z)\hat\Theta_{\kappa d},$$

where

$$(\hat\Theta_{\kappa 0}, \hat\Theta_{\kappa d}) = \arg\min_{\Theta_{\kappa j}} \sum_{i=1}^{n} \big\{y_i - P_\kappa^\top(t_i, z_i)\Theta_{\kappa 0} - \sum_{d=1}^{p} \Theta_{\kappa d}^\top P_\kappa(t_i, z_i)x_{i,d}\big\}^2, \tag{3.1}$$

where $x_{i,d}$ is the dth component of x_i.

Define following notations:

$B = \big[\Theta_{\kappa 0}^\top, \Theta_{\kappa 1}^\top, \Theta_{\kappa 2}^\top, \ldots, \Theta_{\kappa p}^\top\big]^\top.$

$\hat B = \big[\hat\Theta_{\kappa 0}^\top, \hat\Theta_{\kappa 1}^\top, \hat\Theta_{\kappa 2}^\top, \ldots, \hat\Theta_{\kappa p}^\top\big]^\top.$

$$A_i = \left[P_\kappa(t_i, z_i)^\top, P_\kappa(t_i, z_i)^\top x_{i,1}, P_\kappa(t_i, z_i)^\top x_{i,2}, \ldots, P_\kappa(t_i, z_i)^\top x_{i,p}\right]^\top.$$

Equation 3.1 can be written as

$$\hat{B} = \arg\min_{\Theta_{\kappa j}} \sum_{i=1}^{n} \left\{y_i - A_i^\top B\right\}^2. \tag{3.2}$$

Then $\hat{B}$ can be found. $\hat{B} = \left(\sum_{i=1}^n A_i A_i^\top\right)^{-1}\left(\sum_{i=1}^n y_i A_i\right)$. In order to keep $E\gamma_0(z) = E\gamma_1(z) = 0$, we have to centralize $\tilde{\gamma}_0(z)$ and $\tilde{\gamma}_1(z)$. Denote $\tilde{\gamma}_b^*(z) = \tilde{\gamma}_b(z) - E\tilde{\gamma}_b(z)$, $\tilde{\beta}_b^*(t) = \tilde{\beta}_b(t) + E\tilde{\gamma}_b(z)$. Then $E\tilde{\gamma}_b^*(z) = 0$. $\tilde{\gamma}_b^*(z)$ and $\tilde{\beta}_b^*(t)$ are first-step estimators.

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

$$\beta_k(t_i) \approx \beta_k(t) + \beta_k'(t)(t_i - t) \equiv a_k + b_k(t_i - t), \quad k = 0, 1,$$

and for any z_i in the neighborhood of z, by Taylor's expansion,

$$\gamma_k(z_i) \approx \gamma_k(z) + \gamma_k'(z)(z_i - z) \equiv c_k + d_k(z_i - z) \quad k = 0, 1.$$

Note that a_1, b_1, c_1 and d_1 are unknown vectors for every t and z. a_0, b_0, c_0 and d_0 are two unknown constants for every t and z. In the second step, we minimize

$$\sum_{i=1}^{n} [y_i - \tilde{\beta}_0^*(t_i) - \tilde{\gamma}_0^*(z_i) - \tilde{\beta}_1^*(t_i)^\top x_i - \{c_1 + d_1(z_i - z)\}^\top x_i]^2 K_{h_1}(z_i - z) \tag{3.3}$$

and get the minimizer $\hat{c}_1$ which estimates $\gamma_1(z)$ denoted by $\hat{\gamma}_1(z)$, where $K_{h_1}(\cdot) = \frac{1}{h_1}K(\frac{\cdot}{h_1})$. Similarly, we minimize

$$\sum_{i=1}^{n}[y_i - \tilde{\beta}_0^*(t_i) - \tilde{\gamma}_0^*(z_i) - \tilde{\gamma}_1^*(z_i)^\top x_i - \{a_1 + b_1(t_i - t)\}^\top x_i]^2 K_{h_2}(t_i - t), \tag{3.4}$$

$$\sum_{i=1}^{n}[y_i - \tilde{\gamma}_0^*(z_i) - \tilde{\beta}_1^*(t_i)^\top x_i - \tilde{\gamma}_1^*(z_i)^\top x_i - \{a_0 + b_0(t_i - t)\}]^2 K_{h_0}(t_i - t) \tag{3.5}$$

and

$$\sum_{i=1}^{n}[y_i - \tilde{\beta}_0^*(t_i) - \tilde{\beta}_1^*(t_i)^\top x_i - \tilde{\gamma}_1^*(z_i)^\top x_i - \{c_0 + d_0(z_i - z)\}]^2 K_{h_4}(z_i - z). \tag{3.6}$$

We get minimizer $\hat{a}_0$, $\hat{a}_1$ and $\hat{c}_0$ which estimates $\beta_0(t)$, $\beta_1(t)$ and $\gamma_0(z)$ denoted by $\hat{\beta}_0(t)$, $\hat{\beta}_1(t)$ and $\hat{\gamma}_0(z)$.

We could see from equation (3.3), (3.4), (3.5) and (3.6) that we estimate the unknown functions each time as if we have already known the other unknown functions. We show oracle properties of our estimators in two cases: stationary x_i and nonstationary x_i in Chapter 4 and Chapter 5. We derive close form of our estimators in the following.

Define following notations:

$P_\kappa(t) = \big(1, p_1(t), \ldots, p_\kappa(t)\big)^\top$. $P_\kappa(z) = \big(p_1(z), \ldots, p_\kappa(z)\big)^\top$.

$\Theta_{\kappa bt} = (\theta_{0,b}, \theta_{11,b}, \ldots, \theta_{1\kappa,b})^\top$. $\Theta_{\kappa bz} = (\theta_{21,b}, \ldots, \theta_{2\kappa,b})^\top$, $\quad for\ b = 0, 1, \cdots p.$

$\hat{\Theta}_{\kappa bt} = (\hat{\theta}_{0,b}, \hat{\theta}_{11,b}, \ldots, \hat{\theta}_{1\kappa,b})^\top$. $\hat{\Theta}_{\kappa bz} = (\hat{\theta}_{21,b}, \ldots, \hat{\theta}_{2\kappa,b})^\top$, $\quad for\ b = 0, 1, \cdots p.$

It can be easily check that

$P_\kappa(t, z) = [P_\kappa(t)^\top, P_\kappa(z)^\top]^\top.$

$\Theta_{\kappa b} = [\Theta_{\kappa bt}^\top, \Theta_{\kappa bz}^\top]^\top$.

$\hat{\Theta}_{\kappa b} = [\hat{\Theta}_{\kappa bt}^\top, \hat{\Theta}_{\kappa bz}^\top]^\top$.

$\tilde{\beta}_0(t_i) = P_\kappa^\top(t_i)\hat{\Theta}_{\kappa 0t}$.

$\tilde{\beta}_1(t_i)^\top x_i = \sum_{d=1}^p P_\kappa^\top(t_i)\hat{\Theta}_{\kappa dt} x_{i,d}$.

$\tilde{\gamma}_0(z_i) = P_\kappa^\top(z_i)\hat{\Theta}_{\kappa 0z}$.

$\tilde{\gamma}_1(z_i)^\top x_i = \sum_{d=1}^p P_\kappa^\top(z_i)\hat{\Theta}_{\kappa dz} x_{i,d}$.

$\tilde{\beta}_0^*(t_i) = P_\kappa^\top(t_i)\hat{\Theta}_{\kappa 0t} + E\tilde{\gamma}_0(z_i)$.

$\tilde{\beta}_1^*(t_i)^\top x_i = \sum_{d=1}^p [P_\kappa^\top(t_i)\hat{\Theta}_{\kappa dt} + E\tilde{\gamma}_0(z_i)] x_{i,d}$.

$\tilde{\gamma}_0^*(z_i) = P_\kappa^\top(z_i)\hat{\Theta}_{\kappa 0z} - E\tilde{\gamma}_0(z_i)$.

$\tilde{\gamma}_1^*(z_i)^\top x_i = \sum_{d=1}^p [P_\kappa^\top(z_i)\hat{\Theta}_{\kappa dz} - E\tilde{\gamma}_0(z_i)] x_{i,d}$.

We have the following notations:

$W_{ih_1}(z) = (1, \frac{z_i - z}{h_1}) \otimes x_i^\top$.

$A^* = \sum_{i=1}^n W_{ih_1}(z)^\top W_{ih_1}(z) K_{h_1}(z_i - z)$.

$B^* = \sum_{i=1}^n [y_i - \tilde{\beta}_0^*(t_i) - \tilde{\gamma}_0^*(z_i) - \tilde{\beta}_1^*(t_i)^\top x_i] W_{ih_1}(z)^\top K_{h_1}(z_i - z)$

$= \sum_{i=1}^n [y_i - P_\kappa^\top(t_i)\hat{\Theta}_{\kappa 0t} - P_\kappa^\top(z_i)\hat{\Theta}_{\kappa 0z} - \sum_{d=1}^p [P_\kappa^\top(t_i)\hat{\Theta}_{\kappa dt} - E\tilde{\beta}_0(t_i)] x_{i,d}] W_{ih_1}(z)^\top$

$K_{h_1}(z_i - z)$.

After we take the first derivative of equation (3.3), we will have the following solution for $(c_1, h_1 d_1)$:

$$\begin{bmatrix} \hat{c_1} \\ h_1 \hat{d_1} \end{bmatrix} = [A^*]^{-1} B^*.$$

Similarly $\hat{a}_0$, $\hat{c}_0$ and $\hat{a}_1$ could be easily determined.

CHAPTER 4: MODELS WITH STATIONARITY X_I

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_κ.

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

$Ep_i(t_k)p_j(z_k) = C_{ij}$ for all i, j from 1 to κ and k from 1 to n.

for $j \leqslant 0$, $\mu_j(K) = \int_{-\infty}^{\infty} \upsilon^j K(\upsilon)d\upsilon$ and $\nu_j(K) = \int_{-\infty}^{\infty} \upsilon^j K^2(\upsilon)d\upsilon$.

$S = E(x_i x_i^\top \big| z_i = z)$, $S_0 = E(x_i x_i^\top \big| t_i = t)$.

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

$R(z_i) = \gamma_1(z_i) - \gamma_1(z) - \gamma_1^{(1)}(z)(z_i - z)$.

$\overline{A}(t_i) = [P_\kappa(t_i)^\top, P_\kappa(z_i)^\top, P_\kappa(t_i)^\top x_{i,1}, P_\kappa(z_i)^\top{\cdot}0, P_\kappa(t_i)^\top x_{i,2}, P_\kappa(z_i)^\top{\cdot}0, \cdots, P_\kappa(t_i)^\top x_{i,p}, P_\kappa(z_i)^\top \cdot 0]^\top$.

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

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

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

$b_{\kappa 0}(i) = \beta_0(t_i) + \gamma_0(z_i) + (\beta_1(t_i) + \gamma_1(z_i))^\top x_i - A_i^\top \theta_{\kappa 0}$.

$\overline{b_{\kappa 0}}(i) = \beta_0(t_i) + \gamma_0(z_i) + \beta_1(t_i)^\top x_i - \overline{A_i}^\top \theta_{\kappa 0}$.

$\theta_k = [\Theta_{\kappa 0t}^\top, \Theta_{\kappa 0z}^\top, \Theta_{\kappa 1t}^\top, \Theta_{\kappa 1z}^\top, \Theta_{\kappa 2t}^\top, \Theta_{\kappa 2z}^\top, \cdots, \Theta_{\kappa dt}^\top, \Theta_{\kappa dz}^\top]^\top$.

$\overline{\theta}_k = [\Theta_{\kappa 0t}^\top, \Theta_{\kappa 0z}^\top, \Theta_{\kappa 1t}^\top, \Theta_{\kappa 1z}^\top * 0, \Theta_{\kappa 2t}^\top, \Theta_{\kappa 2z}^\top * 0, \cdots, \Theta_{\kappa dt}^\top, \Theta_{\kappa dz}^\top * 0]^\top$.

$$\Sigma_\nu = \begin{pmatrix} \nu_0(K)S & \nu_1(K)S \\ \nu_1(K)S & \nu_2(K)S \end{pmatrix}.$$

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

(A1) x_i is p-dimensional $I(0)$. Let $x_{i,j}$ is the jth component of x_i, without loss of generality, assume $x_{i,j} = b_j x_{i-1,j} + \delta_{i,j}$, where $1 \leqslant i \leqslant n$, $1 \leqslant j \leqslant p$, $\delta_{i,j}$ is independent with $E\delta_{i,j} = 0$, $Var\delta_{i,j} = \zeta_j^2$ and $E(\delta_{i,j}\delta_{i,k}) = \zeta_{j,k}$ for $1 \leqslant k \neq j \leqslant q$ so that $E(x_{i,j}x_{i,k}) = \frac{\zeta_{j,k}}{1-b_j b_k}$. There are constants $C_{j,k} < \infty$ such that $E(x_{i,j}^2 x_{i,k}^2) = C_{j,k} < \infty$ for any j, k from 1 to p and any i from 1 to n.

(A2) $t_i = i/n$. ε_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) Assume that $E[\gamma_0(z_i)] = E[\gamma_1(z_i)] = 0$.

(ii) $\gamma_0(z)$ and $\gamma_1(z)$ are twice continuously differentiable in z for all $z \in [-C, C]$, where C is any constant in $\Re$. S is positive-definite and continuous in a neighborhood of z.

(iii) $\beta_0(t)$ and $\beta_1(t)$ are twice continuously differentiable in t for all $t \in [0, 1]$, S_0 is positive-definite and continuous in a neighborhood of t.

(A5) There are constants $C_Q < \infty$ and $c_\lambda > 0$ such that $|Q_{ij}| \leqslant C_Q$ and $\lambda_{\kappa,min} > c_\lambda$ for all κ 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 $h_1 = C_{h_1} n^{-1/5}$, $h_2 = C_{h_2} n^{-1/5}$, $h_0 = C_{h_0} n^{-1/5}$ and $h_4 = C_{h_4} n^{-1/5}$ for some constants C_{h_1} ,C_{h_2}, C_{h_3} and C_{h_4} satisfying $0 < C_{h_1} < \infty$, $0 < C_{h_2} < \infty$, $0 < C_{h_0} < \infty$ and $0 < C_{h_4} < \infty$.

(ii) $\kappa = C_\kappa n^\nu$ for some constant C_κ satisfying $0 < C_\kappa < \infty$ and some ν 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 $[-1, 1]$, $\mu_0(K) = 1$ and $\mu_1(K) = 0$.

We give some comments on the above conditions. We have assumption A1 to make the proof can be done easily. Assumptions A2 and A3 are regularity conditions. Assumption A4 defines the sense in which $\gamma_1(z_i)$, $\gamma_2(z_i)$, $\beta_1(t)$ and $\beta_2(t)$ 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-step estimators. 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 insures that errors in the series approximations to the $\gamma_1(z)$ and $\gamma_2(z)$ converge to zero sufficiently rapidly as $\kappa \to \infty$. Assumption A7 states the rates at which $\kappa \to \infty$, $h_0 \to \infty$, $h_1 \to \infty$, $h_2 \to \infty$ and $h_4 \to \infty$ as $n \to \infty$. The assumed convergent rate of h_0, h_1, h_2 and h_4 is well known to be asymptotically optimal for kernel regression when the conditional mean functions are twice continuously differentiable. The required rate of κ insures that the asymptotic bias and variance of the first-step estimators are sufficiently small to achieve the $n^{-2/5}$

rate of convergence in the second-step, see Horowitz (2004). Assumption A8 helps the second-step estimators to 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 kernel 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 of two-step estimators when x_i is stationary. Detail proof of the following Theorems are provided in Appendix.

Theorem 4.1. *Under conditions (A1) $\sim$ (A9),*

$$\sqrt{nh_1}[\hat{\gamma}_1(z) - \gamma_1(z) - \tfrac{h_1^2}{2}\mu_2(K)\gamma_1^{(2)}(z)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, f_z(z)^{-1}\delta^2 S^{-1}\nu_0(K)\}.$$

Theorem 4.2. *Under conditions (A1) $\sim$ (A9),*

$$\sqrt{nh_2}[\hat{\beta}_1(t) - \beta_1(t) - \tfrac{h_2^2}{2}\mu_2(K)\beta_1^{(2)}(t)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, \delta^2 S_0^{-1}\nu_0(K)\}.$$

Theorem 4.3. *Under conditions (A1) $\sim$ (A9),*

$$\sqrt{nh_4}[\hat{\gamma}_0(z) - \gamma_0(z) - \tfrac{h_4^2}{2}\mu_2(K)\gamma_0^{(2)}(z)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, f_z(z)^{-1}\delta^2\nu_0(K)\}.$$

Theorem 4.4. *Under conditions (A1) $\sim$ (A9),*

$$\sqrt{nh_0}[\hat{\beta}_0(t) - \beta_0(t) - \tfrac{h_0^2}{2}\mu_2(K)\beta_0^{(2)}(t)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, \delta^2\nu_0(K)\}.$$

Above theorems can be extended to that $x_i, z_i, \varepsilon, \zeta$ is a strictly α-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 $\hat{\gamma}_1(z) - \gamma_1(z)$, $\hat{\beta}_1(t) - \beta_1(t)$, $\hat{\gamma}_0(z) - \gamma_0(z)$ and $\hat{\beta}_0(t) - \beta_0(t)$ reach the optimal convergent rate.

CHAPTER 5: MODEL WITH NONSTATIONARY X_I AND STATIONARY Z_I

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 δ_s is an $I(0)$ process with mean zero and variance Ω_δ.

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

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 i.i.d. δ_i and ρ-mixing δ_i, generalizes in an obvious way to the multivariate cases and leads to

$$\frac{x_{[nr]}}{\sqrt{n}} \Longrightarrow W_\delta(r) \quad as \quad n \longrightarrow \infty, \tag{5.2}$$

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

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

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\Gamma(x_{[nr]})}{\sqrt{n}} \xrightarrow{d} \int_0^1 \Gamma(W_\delta(s)) ds \quad as \quad n \longrightarrow \infty, \tag{5.3}$$

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

$$\frac{1}{n} \sum_{i=1}^{n} (\frac{x_{[nr]}}{\sqrt{n}})^{\otimes \ell} \xrightarrow{d} \int_0^1 W_\delta(s)^{\otimes \ell} ds \equiv W_{\delta,\ell} \quad as \quad n \longrightarrow \infty, \tag{5.4}$$

see Theorem 1.2 in Berkes (2006) and Cai and Park (2009).

Define $\hat{Q}_\kappa^* = n^{-2}\sum_{i=1}^n A_i A_i^\top$ and $Q_\kappa^* = E\hat{Q}_\kappa^*$. Let Q_{ij}^* denote the (i,j) element of Q_κ^*.

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

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

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

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

(iv) Assume $h_4 = C_{h_4} n^{-1/5}$ for a constant C_{h_4} satisfying $0 < C_{h_4} < \infty$, $\kappa = C_\kappa n^\nu$ for a constant C_κ satisfying $0 < C_\kappa < \infty$ and a constant ν satisfying $\frac{1}{5} < \nu < \frac{3}{10}$.

(B3) Assume $\sup_x |b_{\kappa 0}(i)| = O(\kappa^{-2})$.

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

(B5) There are constants $C_{Q^*} < \infty$ and $c_{\lambda^*} > 0$ such that $|Q_{ij}^*| \leqslant C_{Q^*}$ and $\lambda_{\kappa,min} > c_{\lambda^*}$ for all κ and all $i, j = 1, ..., d(\kappa)$.

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

(B7) (i) Assume that $E[\gamma_0(z_i)] = 0$, $E[\gamma_1(z_i)] = 0$, $\gamma_0(z)$ and $\gamma_1(z)$ are twice continuously differentiable in z for all $z \in [-C, C]$, where C is any constant in $\Re$. S is positive-definite and continuous in a neighborhood of z.

(ii) Assume that $\beta_0(t)$ and $\beta_1(t)$ are twice continuously differentiable in t for all

$t \in [0,1]$. S_0 is positive-definite and continuous in a neighborhood of t.

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

We give some comments on above conditions. Assumption B1 makes the proof can be done easily. Assumption B2 states $\kappa \to \infty$ and bandwidths converge to 0 as $n \to \infty$. It requires the first-step estimators to be undersmooth. Undersmoothing is needed to insure the sufficiently rapid convergence of the bias of the orthogonal series estimators. We will show the asymptotic of two-step estimators does not depend on the choice of κ if assumption B2 is satisfied. Optimizing the choice of κ 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 errors in the series approximations to $\gamma_0(z)$ and $\gamma_1(z)$ converge to zero sufficiently rapidly as $\kappa \to \infty$, See Horowitz (2004). Assumption B4 helps second-step estimators to avoid the curse of dimensionality. These conditions are satisfied by splines and the Fourier basis. α-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 α-mixing with exponentially decaying coefficients, see discussions and examples in Cai (2002), Carrasco (2002) and Chen (2005). Assumption B5 insures the existence and nonsingularity of the covariance matrix of the asymptotic form of the first-step estimators, see Horowitz (2004). Assumption B6 can be relaxed to allow for conditional heteroscedasticity of the form $var(\varepsilon_i | x_t, z_t) = \delta^2(z_t)$, i.e. the conditional variance is only a function of the stationary z_t. However, it is technically difficult to let it also be a function of the nonstationary x_t; see Cai and Park (2009). Assumption B7 are smoothness condition. 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 of the two-step estimators when x_i is nonstationary. Detail proofs of following Theorems are provided in Appendix.

Theorem 5.1. *Under conditions (B1),(B2)(i),(B3) ∼(B8),*

$$n\sqrt{h_1}[\hat{\gamma}_1(z) - \gamma_1(z) - \tfrac{h_1^2}{2}\mu_2(K)\gamma_1^{(2)}(z)\{1 + o_p(1)\}] \xrightarrow{d} MN\big(\Sigma_\delta(z)\big)$$

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 $\beta_1(t)$.

Theorem 5.2. *Under conditions (B1),(B2)(ii),(B3) ∼(B8),*

$$n\sqrt{h_2}[\hat{\beta}_1(t) - \beta_1(t) - \tfrac{h_2^2}{2}\mu_2(K)\beta_1^{(2)}(t)\{1 + o_p(1)\}] \xrightarrow{d} MN\big(\Sigma_\delta(t)\big)$$

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

Following theorems show the asympototic of $\gamma_0(z)$ and $\beta_0(t)$

Theorem 5.3. *Under conditions (B1),(B2)(iii),(B3) ∼(B8),*

$$\sqrt{nh_0}[\hat{\beta}_0(t) - \beta_0(t) - \tfrac{h_0^2}{2}\mu_2(K)\beta_0^{(2)}(t)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, \delta^2\nu_0(K)\}.$$

Theorem 5.4. *Under conditions (B1),(B2)(iv),(B3) ∼(B8),*

$$\sqrt{nh_4}[\hat{\gamma}_0(z) - \gamma_0(z) - \tfrac{h_4^2}{2}\mu_2(K)\gamma_0^{(2)}(z)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, f_z^{-1}(z)\delta^2\nu_0(K)\}.$$

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, $\hat{\gamma}_1(z)$ and $\hat{\beta}_1(t)$, are "oracle" in the sense that their asymptotic distribution are the same as the case with known $\beta_1(t)$ and $\gamma_1(z)$. The bandwidth of h_1 and h_2 is taken to be of the order $n^{-2/5}$ so that $\hat{\gamma}_1(z) - \gamma_1(z)$ and $\hat{\beta}_1(t) - \beta_1(t)$ reach the optimal convergent rate. The bandwidth of h_0 and h_4 is taken to be of the order $n^{-1/5}$ so that $\hat{\beta}_0(t) - \beta_0(t)$ and $\hat{\gamma}_0(z) - \gamma_0(z)$ reach the optimal convergent rate.

CHAPTER 6: SIMULATIONS

We have simulations to demonstrate that the proposed two-step estimators give an accurate approximation to the unknown functions. Since B-spline, see C. (1978) is efficient in digital computation and functional approximation, we here use the B-spline basis in the first-step estimation. κ is chosen to be 8. Smaller number κ or larger number κ will not have a big effect on the simulation results. We choose standard normal kernel as our kernel function used in the simulation. We consider the following model.

$$\begin{aligned} y_i &= \beta_0(t_i) + \gamma_0(z_i) + (\beta_1(t_i) + \gamma_1(z_i))x_{i,1} + (\beta_2(t_i) + \gamma_2(z_i))x_{i,2} + \varepsilon_i \\ &= e^{3t_i} + 20t_i + (0.5z_i^3 - 1.5z_i^2 - 0.5z_i + 8) + (-40t_i^2 - 20t_i + 1.5z_i^2 - 7.5z_i - 8)x_{i,1} \\ &\qquad + (e^{4t_i} + 2t_i + 3z_i^2 - 6z_i - 16)x_{i,2} + \varepsilon_i, \quad (6.1) \end{aligned}$$

where $\gamma_0(z_i) = 0.5z_i^3 - 1.5z_i^2 - 0.5z_i + 8$, $\gamma_1(z_i) = 1.5z_i^2 - 7.5z_i - 8$ 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 $U_i \sim$ Uniform $(-4, 4)$ in 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 $x_{1,2}$ and the second component of x at time i is denoted by $x_{i,2}$. Note that we choose those $\gamma(\cdot)$ functions so that $E(0.5z^3 - 1.5z^2 - 0.5z + 8) = 0$,

$E(1.5z^2 - 7.5z - 8) = 0$ and $E(3z^2 - 6z - 16) = 0$.

6.1 Stationary

Example 1:

Choose $x_{1,1} = x_{1,2} = 0$, $x_{i,1} = 0.9x_{i-1,1} + \delta_{1i}$, where $\delta_{1i} \sim t(3)$ and $x_{i,2} = 0.6x_{i-1,2} + \delta_{2i}$, where $\delta_{2i} \sim t(7)$. y is generated from above model (6.1). So that x_1, x_2 and y all are stationary. 10 grid points for β functions and 40 grid points for γ functions are chosen with 500 simulations at each grid point.

Simulation results are shown in Figure 3 for $n = 100$.

Simulation results are shown in Figure 4 for $n = 400$.

The solid lines are true lines of β_0, γ_0, β_1, β_2, γ_1 and γ_2 functions in Figure 3 and Figure 4. The middle dash dot lines in Figure 3 and Figure 4 are the median of the estimators. The upper and lower dot lines in Figure 3 and Figure 4 are 2.5% and 97.5% quantile of the estimators.

You could see from Figure 3 and Figure 4 that the estimation is very good. The solid lines almost cover the middle dash dot lines.

6.2 Nonstationary

Example 2:

Choose $x_{1,1} = x_{1,2} = 0$, $x_{i,1} = x_{i-1,1} + \delta_{1i}$, where $\delta_{1i} \sim t(3)$ and $x_{i,2} = x_{i-1,2} + \delta_{2i}$, where $\delta_{2i} \sim t(7)$. y is generated from above model (6.1). So that x_1, x_2 and y all are nonstationary. x_i and y_i are cointegration. 10 grid points for β functions and 40 grid points for γ functions are chosen with 500 simulations at each grid point. So the only difference between example 1 and example 2 is stationarity of x_i.

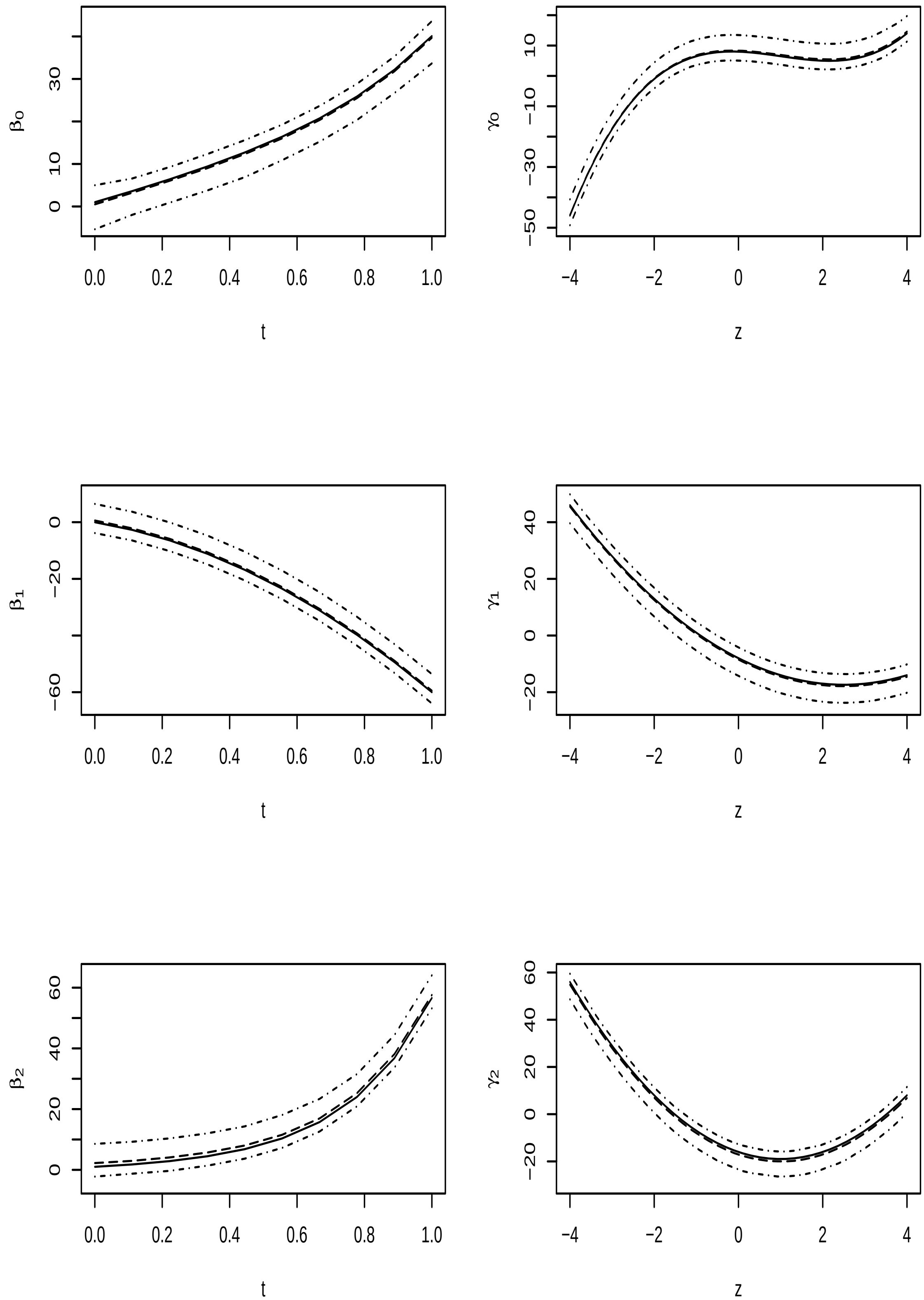


Figure 3: Estimated function β and γ, their mediums and 95% pointwise confidence intervals for Model (6.1) when x is stationary and $n = 100$

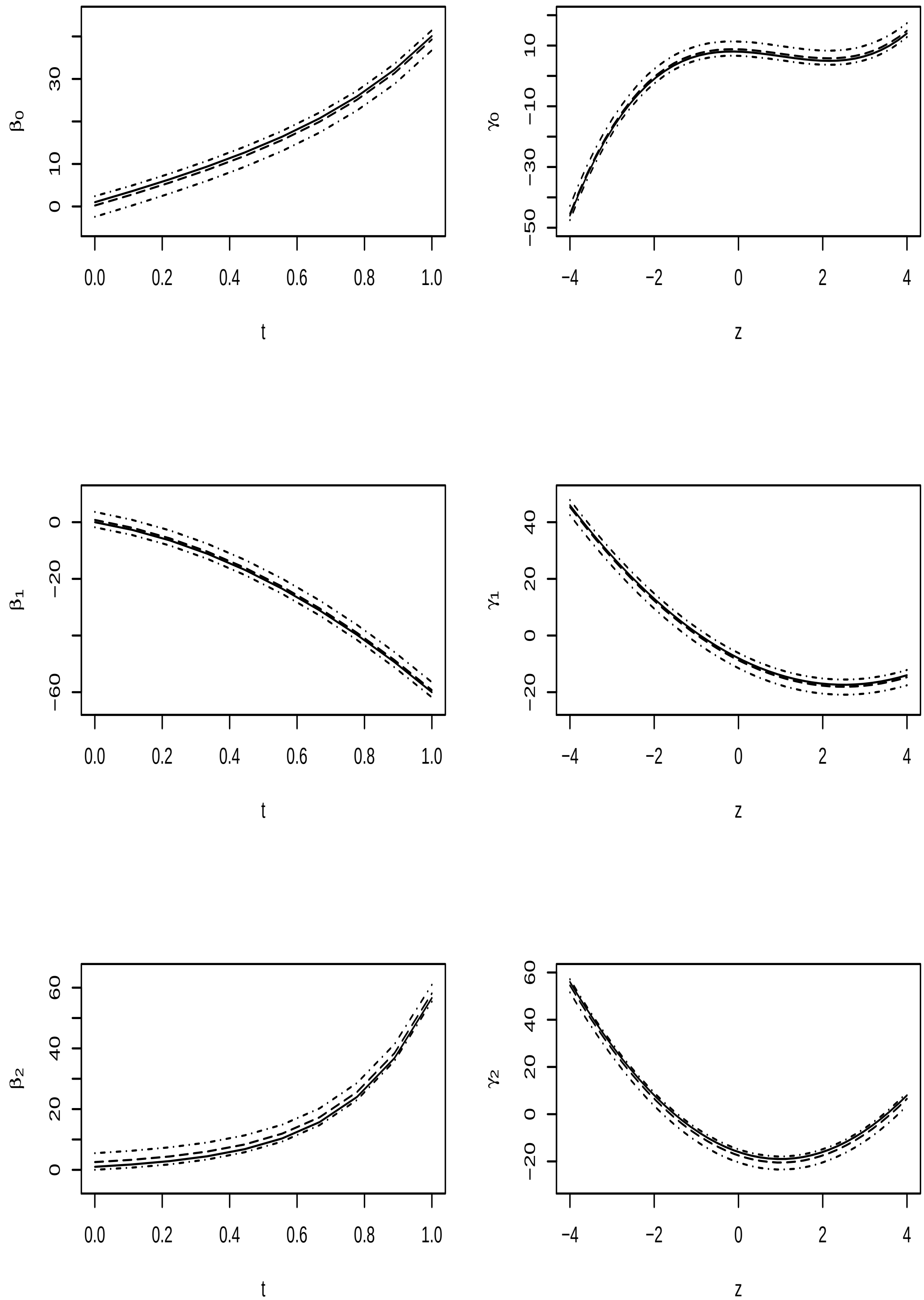


Figure 4: Estimated function β and γ, their mediums and 95% pointwise confidence intervals for Model (6.1) when x is stationary and $n = 400$

Simulation results are shown in Figure 5 for $n = 100$.

Simulation results are shown in Figure 6 for $n = 400$.

The solid lines are true lines of β_0, γ_0, β_1, β_2, γ_1 and γ_2 functions in Figure 5 and Figure 6. The middle dash dot lines in Figure 5 and Figure 6 are the median of the estimators. The upper and lower dot lines in Figure 5 and Figure 6 are 2.5% and 97.5% quantile of the estimators.

You could see from Figure 5 and Figure 6 that the estimation is very good. The solid lines almost cover the middle dash dot lines.

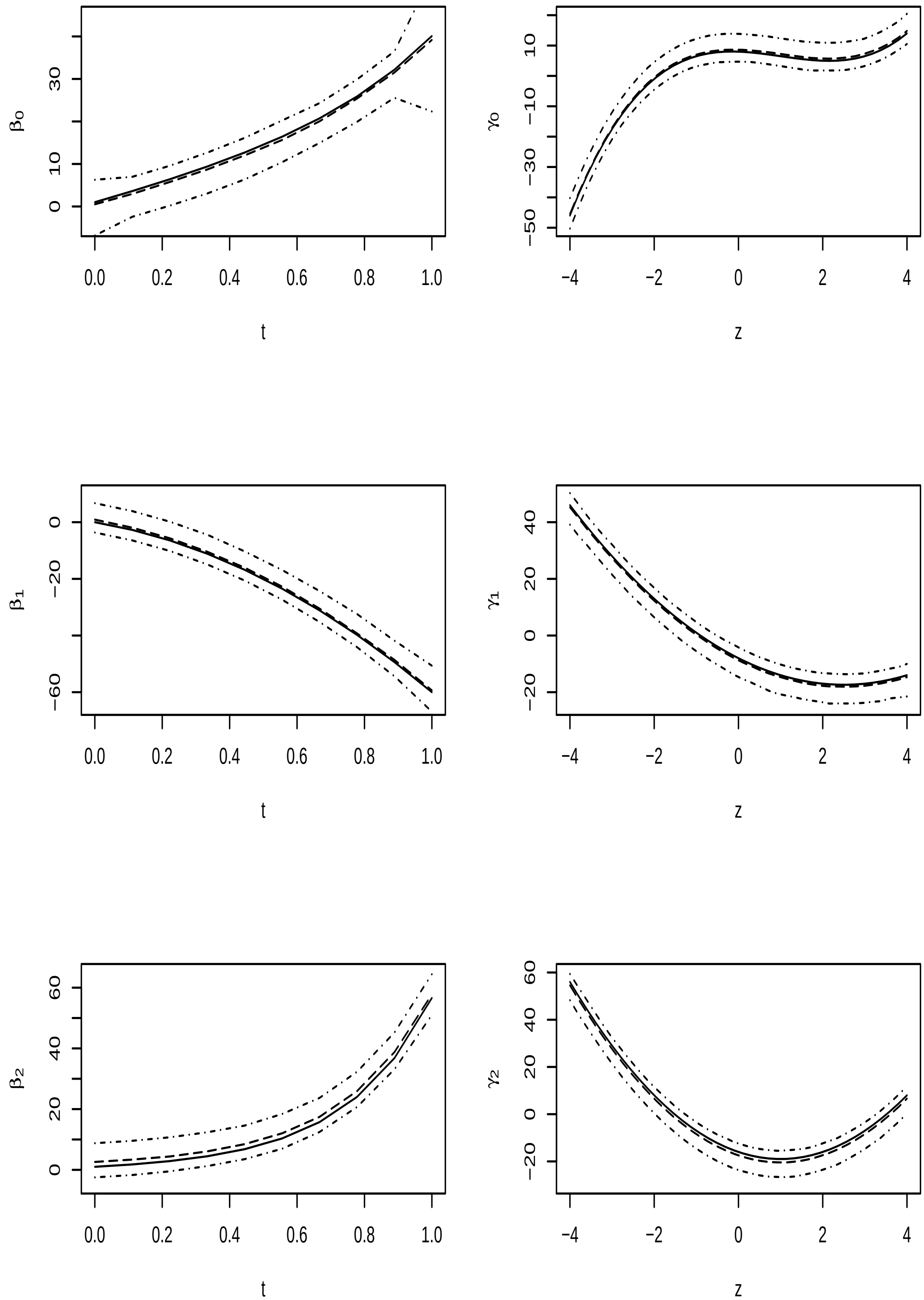


Figure 5: Estimated function β and γ, their mediums and 95% pointwise confidence intervals for Model (6.1) when x is nonstationary and $n = 100$

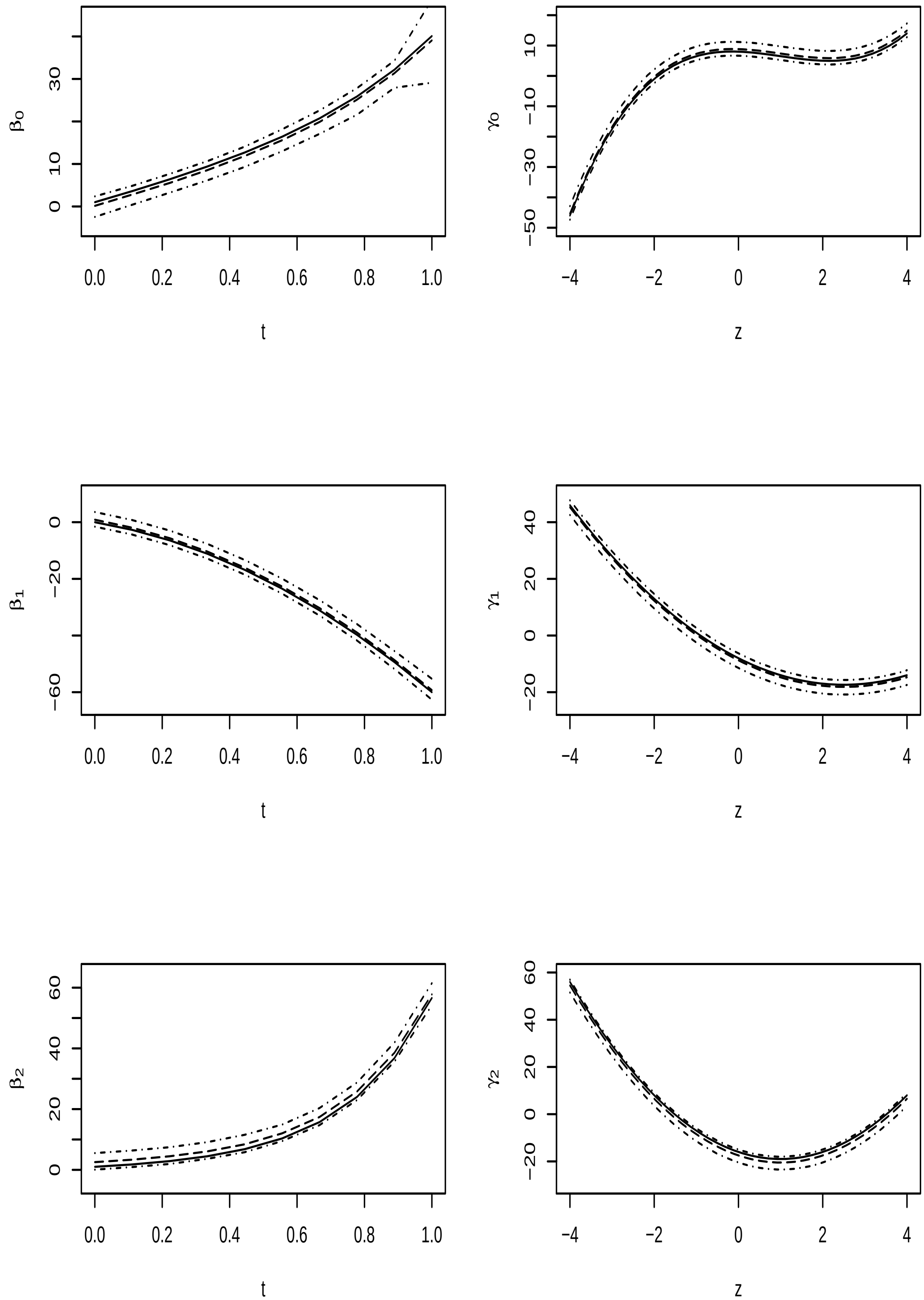


Figure 6: Estimated function β and γ, their mediums and 95% pointwise confidence intervals for Model (6.1) when x is nonstationary and $n = 400$

CHAPTER 7: REAL EXAMPLE

We consider a real application here. We download 5 year daily Treasury bond yield rate, 6 month daily Treasury bill yield rate, stock price of Morgan Stanley and price of S&P500 from websites https://www.treasury.gov/resource-center/data-chart-center/interest-rates/Pages/TextView.aspx?data=yield, https://finance.yahoo.com/quote/MS?p=MS and https://finance.yahoo.com/quote/%5EGSPC/?p=ˆGSPC. All data are from the Jan. 2nd, 2003 to Dec. 31st, 2015. The sample size is 3249. It can be easily seen from Table 2 that stock price of Morgan Stanley and price of S&P500 are nonstationary by ADF test, however, log "difference" of 5 year daily Treasury bond yield rate and 6 month daily Treasury bill yield rate is stationary by ADF test, see AN APPLICATION in Jiang (2014). We build our model based on CAPM model:

$$y_i = \beta_0(t_{i-1}) + \gamma_0(z_{i-1}) + (\beta_1(t_{i-1}) + \gamma_1(z_{i-1}))x_{i-1} + \epsilon_i. \tag{7.1}$$

We choose log "difference" of 5 year daily Treasury bond yield rate and 6 month daily Treasury bill yield rate as z, stock price of Morgan Stanley as y, price of S&P500 as x. It is well known that return of S&P500 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 S&P500 and Morgan Stanley, respectively. That is different with those in the traditional CAPM. The coefficients are constant in traditional CAPM. But they are not constant in our model. We split the sample into two parts: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:

$$\hat{y}_j = \hat{\beta}_0(t_{j-1}) + \hat{\gamma}_1(z_{j-1}) + (\hat{\beta}_1(t_{j-1}) + \hat{\gamma}_1(z_{j-1}))x_{j-1}, \tag{7.2}$$

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 8 shows the cointegration relationship between Stock price of Morgan Stanley and price of S&P500 as well as the estimated β_0 function. The functions in Figure 7 is estimated by the training data. We could see that β_0 and β_1 change with time and γ_0 and γ_1 change with yield rate. Figure 8 shows that the positive relationship between Stock price of Morgan Stanley and price of S&P500 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) and model (1.2) in Table 1. The residual is stationary from ADF test. I calculate the variance of the residual, which is 13.380. That is larger than the variance of the residual in our model (7.1), which is 5.859. We believe our model (7.1) is better than model (1.1). It is easily to see that our model (7.1) is better than model (1.2) from Table 3 too. The variance of the residual in model (1.2) is 68.504. This indicates that our model corrects the error as the error correction model of Engel and Granger.

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. Figure 9 shows $\hat{y}$ from the first 3000 data, which are red dots left of the vertical line, and one-step forecasts, which are red dots right of the vertical line. We can see that the estimation error is small and the forecast is very well.

Table 2: ADF test for stock price of Morgan Stanley, price of S&P500 and log "difference" of 5 year daily Treasury bond yield rate and 6 month daily Treasury bill yield rate

	ADF Test Statistic	P VALUE
Stock price of Morgan Stanley	-0.5896	0.4284
Price of S&P500	1.3095	0.9522
Treasury yield rate from Jiang (2014)	-3.3093	< 0.01

Table 3: Variance of the residual in each model

Variance of the residual	model (1.1)	model (1.2)	model (7.1)
	13.380	68.504	5.859

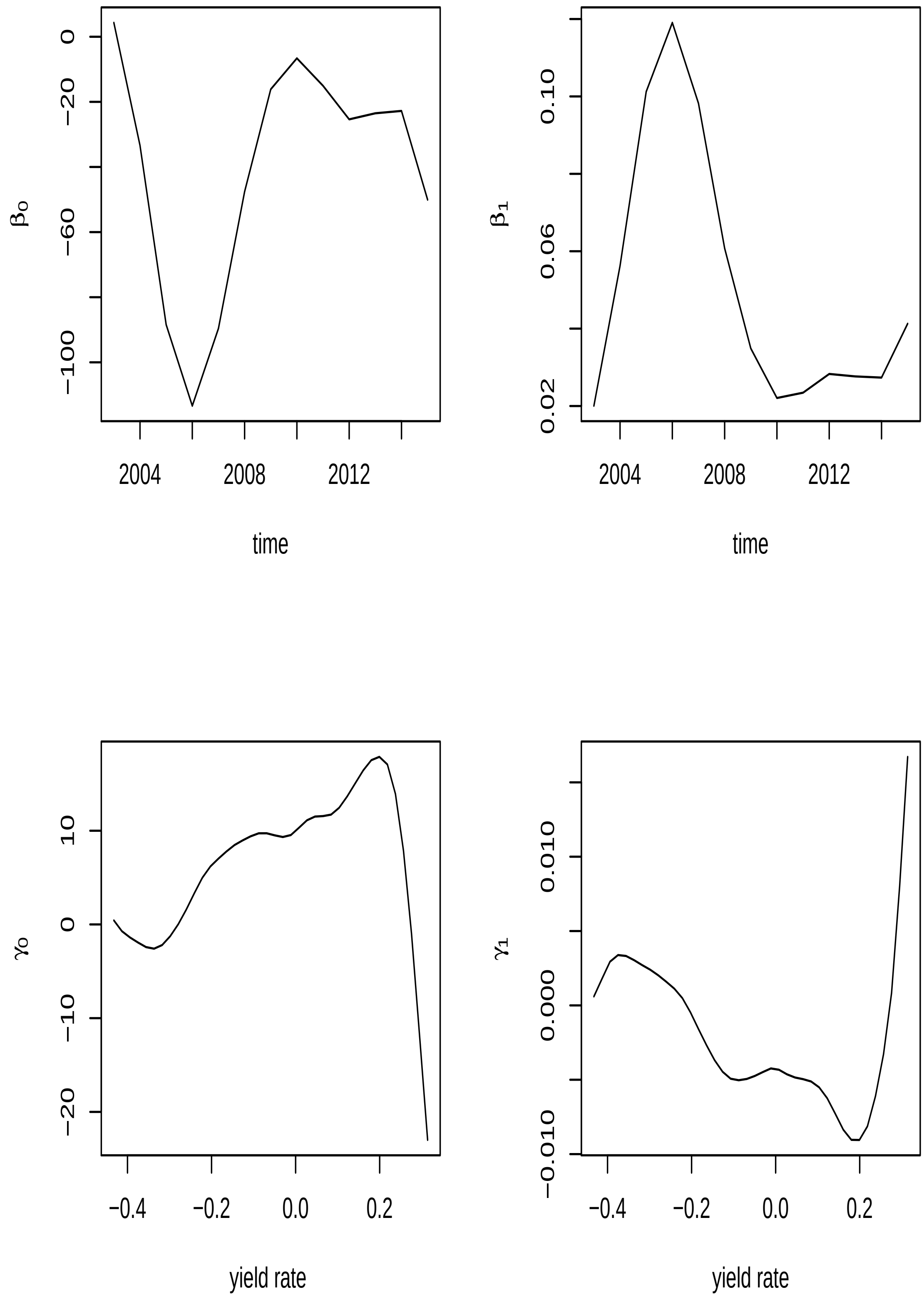
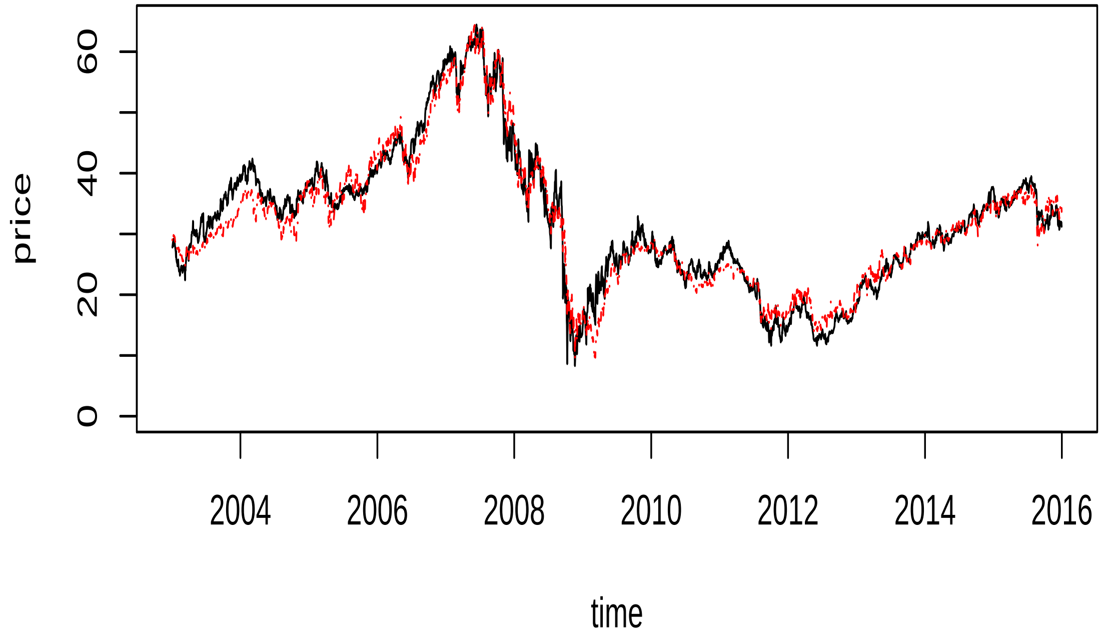


Figure 7: Estimated functions from model (7.1)



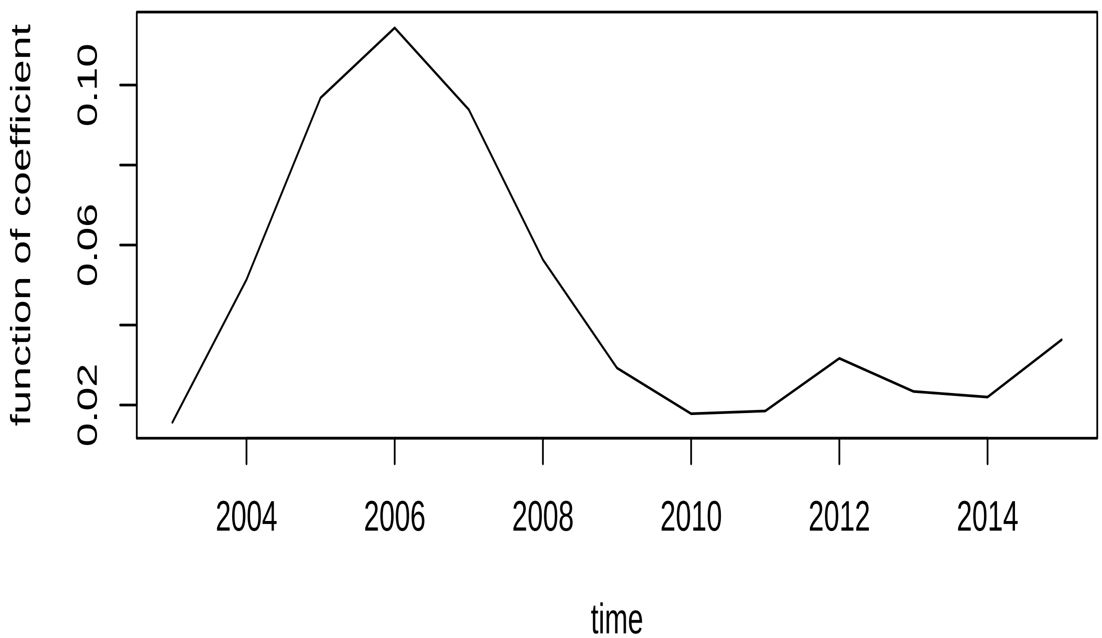


Figure 8: Estimated stock price and function of coefficient from model (7.2)

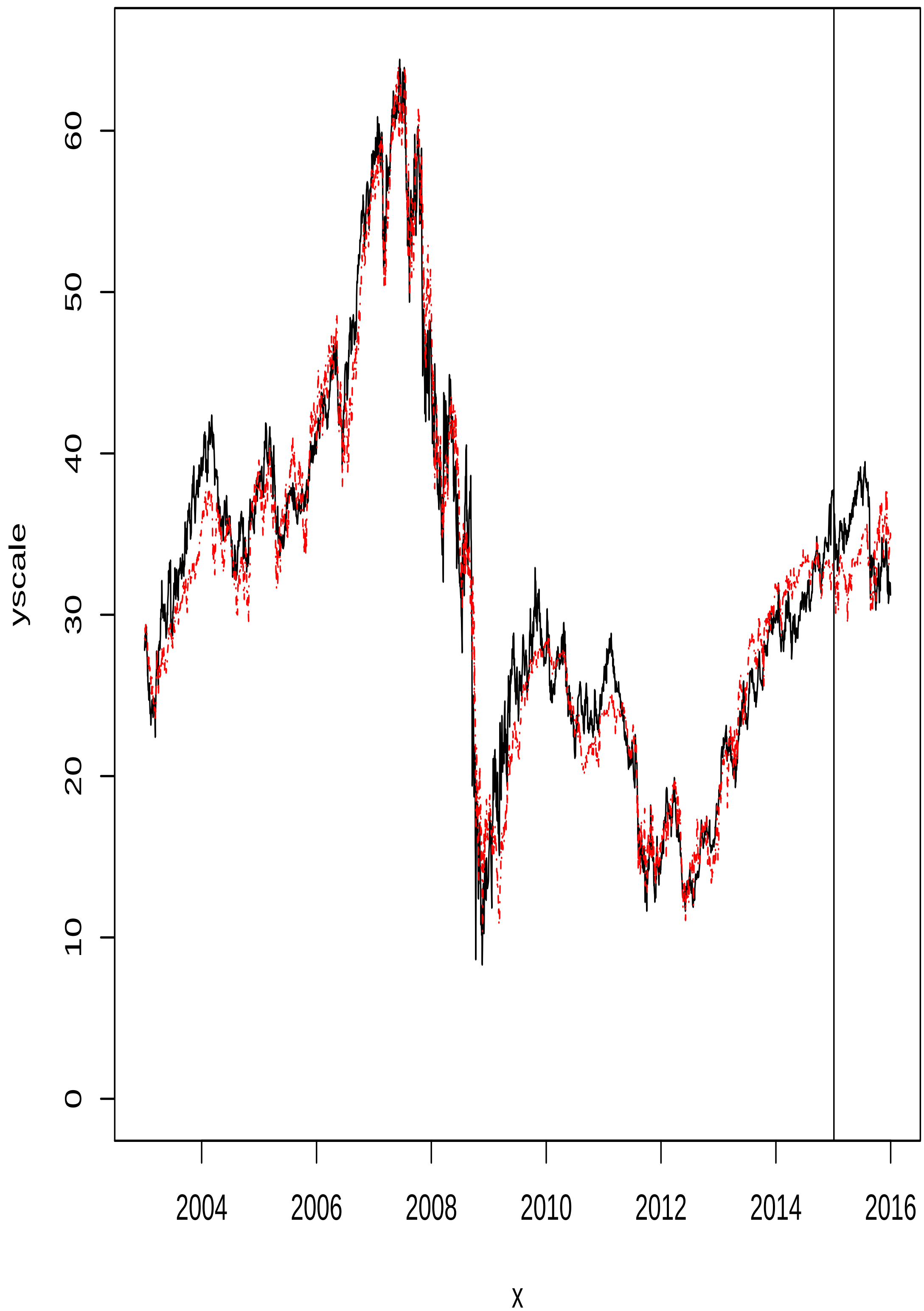


Figure 9: $\hat{y}$ and One step forecast (1.2)

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 extend 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

Berkes, I., H. L. (2006). Convergence of integral functionals of stochastic processes. *Econometric Theory*, 22:304–322.

Billingsley, P. (1999). *Convergence of Probability Measures*. Wiley, New York.

Blume, M. (1975). Betas and their regression tendencies. *Journal of Finance*, 30:785–796.

Blume, M. (1981). Beta stationarity and estimation period: Some analytical results. *Journal of Financial and Quantitative Analysis*, 5:747–757.

C., D. B. (1978). *A Practical Guilde to Splines*. Springer-Verlag New York.

Cai, Z., F. J. Y. Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95:941–956.

Cai, Z., L. Q. and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2):101–113.

Cai, Z. (2002). Regression quantile for time series. *Econometric Theory*, 18:169–192.

Carrasco, M., C. X. (2002). Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18:17–39.

Chang, Y., M.-C. E. (2003). Electricity demand analysis using cointegration and error-correction models with time varying parameters: The mexican case. *Working Paper. Department of Economics, Texas A&M University.*

Chen, S.X., T.-C. (2005). Nonparametric inference of value at risk for dependent financial returns. *Journal of Financial Econometrics*, 3:227–255.

Cochrane, J. H. (2001). *Asset Pricing*. Princeton University Press,Princeton and Oxford.

Einmahl, U. (1987). A useful estimate in the multidimensional invariance principle. *Probability Theory and Related Fields*, 76:81–101.

Engle, R.F., G.-C. (1987). Cointegration and error correction: Representation, estimation, and testing. *Econometrica*, 55:251–276.

Fabozzi, F.J. Francis, J. (1978). Beta as a random coefficient. *Journal of Financial and Quantitative Analysis*, 13:101–116.

Granger, C. (1981). Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, 16(1):121–130.

Horowitz, J. L., M. E. (2004). Nonparametric estimation of an additive model with a link function. *Ann. Statist.*, 32:2412–2443.

Jiang, J. (2014). Multivariate functional-coefficient regression models for nonlinear vector time series data. *Biometrika*, 101:689–702.

Jiang J., L. J. (2008). Two-stage local m-estimation of additive models. *Science in China*, 51:1315–1338.

Kaplanis, E. (1988). Stability and forecasting of the co-movement measures of international stock market returns. *Journal of International Money and Finance*, 7:63–75.

Koch, P.D. Koch, T. (1991). Evolution in dynamic linkages across national stock indexes. *Journal of International Money and Finance*, 10:231–251.

Lee, C.F., N. P. F. J. C. C. (1986). On accounting-based, market-based and composite-based beta predictions: Methods and implications. *The Financial Review*, 21:51–68.

Levy, R. (1971). On the short term stationarity of beta coefficient. *Financial Analyst Journal*, 27:55–62.

McDonald, B. (1985). Making sense out of unstable alphas and betas. *Journal of Portfolio Management*, 11:19–22.

Park, J., H. S. (1999). Cointegrating regressions with time varying coefficients. *Econometric Theory*, 15:664–703.

Phillips, P. (1989). Partially identified econometric models. *Econometric Theory*, 5:181–240.

Phillips, P.C.B., P. J. (1998). Nonstationary density and kernel autoregression. *Unpublished manuscript*.

Rosenberg, B. (1985). Prediction of common stock betas. *Journal of Portfolio Management*, 11:5–14.

Shao, Q., L. C. (1987). Strong approximation for partial sums of weakly dependent random variables. *Scientia Sinica*, 15:576–587.

TSAY, R. S. (2002). *Analysis of Financial Time Series*. A Wiley-Interscience Publication. JOHN WILEY & SONS, INC.

Xiao, Z. (2009). Robust inference in nonstationary time series models. *Journal of Econometrics*, 169:211–223.

Zhou S, Shen X, W. D. A. (1998). Local asymptotics for regression splines and condence regions. *Ann. Statist.*, 26:1760–1782.

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)$\lambda_{min}(A)Tr(B) \leqslant Tr(AB) \leqslant \lambda_{max}(A)Tr(B)$,

(b)$\lambda_{min}(A)\lambda_{max}(B) \leqslant \lambda_{max}(AB) \leqslant \lambda_{max}(A)\lambda_{max}(B)$.

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* κ*.*

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)$E\|\hat{Q}_\kappa - Q_\kappa\|^2 = \sum_{k=1}^{d(\kappa)}\sum_{j=1}^{d(\kappa)} E\big(n^{-1}\sum_{i=1}^n A_{ik}A_{ij} - Q_{kj}\big)^2$

$= \sum_{k=1}^{d(\kappa)}\sum_{j=1}^{d(\kappa)} \big(En^{-2}\sum_{i=1}^n\sum_{\ell=1}^n A_{ik}A_{\ell k}A_{ij}A_{\ell j} - Q_{kj}^2\big).$

Define $M_{k_1,j_1} = En^{-2}\sum_{i=1}^n\sum_{\ell=1}^n A_{ik_1}A_{\ell k_1}A_{ij_1}A_{\ell j_1} - Q^2_{k_1j_1}$, where A_{ik_1} are from $P_\kappa(t_i,z_i)^\top x_{i,m_1}$ and A_{ij_1} are from $P_\kappa(t_i,z_i)^\top x_{i,m_2}$ for any m_1 and m_2 from 1 to p and $m_1 \neq m_2$.

Define $N_{k_2,j_2} = En^{-2}\sum_{i=1}^n\sum_{\ell=1}^n A_{ik_2}A_{\ell k_2}A_{ij_2}A_{\ell j_2} - Q^2_{k_2j_2}$, where A_{ik_2} and A_{ij_2} are from $P_\kappa(t_i,z_i)^\top x_{i,m}$ 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 $N_{k_2,j_2} = O(n^{-1})$.

Without 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 $EA_{ik_2}A_{ij_2} = C_{11}\zeta_1^2/(1-b_1^2)$ and $Q^2_{k_2j_2} = C_{11}^2\zeta_1^4(1-b_1^2)^{-2}$.

$EA_{ik_2}A_{ij_2}A_{i+1k_2}A_{i+1j_2} = E(p_1(t_i)x_{i,1}p_1(z_i)x_{i,1}p_1(t_{i+1})(b_1x_{i,1}+\delta_{i+1,1})p_1(z_{i+1})(b_1x_{i,1}+\delta_{i+1,1})) = C_{11}^2b_1^2Ex_{i,1}^4 + C_{11}^2\zeta_1^4(1-b_1^2)^{-1}$.

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

Similar arguments yield that

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

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

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

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

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

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

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

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

$N_{k_2,j_2} = n^{-2}[C_{11}^2\zeta_1^4(1-b_1^2)^{-2}n^2 + O(n)] - C_{11}^2\zeta_1^4(1-b_1^2)^{-2} = O(n^{-1})$.

Without 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 $Ex_{i,1} =$

$Ex_{i,2} = 0$.

$Q_{k_1 j_1} = \frac{1}{n}\sum_{i=1}^{n} Ep_1(t_i)p_1(z_i)x_{i,1}x_{i,2} = C_{11}\frac{\zeta_{1,2}}{1-b_1b_2}$ so that $Q_{k_1 j_1}^2 = C_{11}^2\frac{\zeta_{1,2}^2}{(1-b_1b_2)^2}$.

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

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

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

$E(p_1(t_i)p_1(z_i)p_1(t_{i+1})p_1(z_{i+1}))Ex_{i,1}x_{i,2}\delta_{i+1,1}\delta_{i+1,2} = b_1b_2C_{11}^2\zeta_1^2(1-b_1^2)^{-1}\zeta_2^2(1-b_2^2)^{-1} +$ $C_{11}^2\frac{\zeta_{1,2}^2}{1-b_1b_2}$.

Following the arguments of N_{k_2,j_2},

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

Therefore, $M_{k_1,j_1} = n^{-2}[C_{11}^2\frac{\zeta_{1,2}^2}{(1-b_1b_2)^2}n^2 + O(n)] - C_{11}^2\frac{\zeta_{1,2}^2}{(1-b_1b_2)^2} = O(n^{-1})$.

(b) follow lemma 3 of Jiang J. (2008).

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

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

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

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

Proof of lemma A.4,

(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 (A6),

$\|n^{-1}\hat{Q}_k^{-1}\sum_{i=1}^{n}\{A_i b_{k_0}(i)\}\|^2 = \|\hat{Q}_\kappa^{-1}\Lambda\varrho/n\|^2 = n^{-2}\varrho^\top\Lambda^\top\hat{Q}_\kappa^{-2}\Lambda\varrho \leqslant n^{-2}\lambda_{\max}(\hat{Q}_\kappa^{-2})\varrho^\top\Lambda^\top \Lambda\varrho = n^{-1}\lambda_{\max}(\hat{Q}_\kappa^{-2})\varrho^\top\hat{Q}_\kappa\varrho \leqslant n^{-1}\lambda_{\max}(\hat{Q}_\kappa^{-1})^2\lambda_{\max}(\hat{Q}_\kappa)\varrho^\top\varrho = O(\kappa^{-4})$.

(b) follow lemma 5 of Horowitz (2004).

Lemma A.5. *By condition (A1)-(A9),*

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

Proof of lemma A.5,

Define $M_i = \beta_0(t_i) + \gamma_0(z_i) + \{\beta_1(t_i) + \gamma_1(z_i)\}^\top x_i$ and $\eta_i = A_i^\top \hat{B} - M_i = A_i^\top(\hat{B} - \theta_{k_0}) - b_{\kappa 0}(i)$, so that $A_i^\top \hat{B} = \eta_i + M_i$,

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

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

$\Rightarrow \hat{B} - \theta_{k_0} = \frac{1}{n}\hat{Q}_\kappa^{-1}\sum_{i=1}^n \varepsilon_i A_i + \frac{1}{n}\hat{Q}_\kappa^{-1}\sum_{i=1}^n A_i b_{\kappa 0}(i)$

$\Rightarrow \hat{B} - \theta_{k_0} = n^{-1}Q_k^{-1}\sum_{i=1}^n A_i\varepsilon_i + n^{-1}Q_k^{-1}\sum_{i=1}^n A_i b_{k_0}(i) + n^{-1}(\hat{Q}_\kappa^{-1} - Q_\kappa^{-1})\sum_{i=1}^n A_i\varepsilon_i + n^{-1}(\hat{Q}_\kappa^{-1} - Q_\kappa^{-1})\sum_{i=1}^n A_i b_{k_0}(i) = J_{n_1} + J_{n_2} + J_{n_3} + J_{n_4}$.

$\|J_{n_3}\| = \|Q_\kappa^{-1}(\hat{Q}_\kappa - Q_\kappa)n^{-1}\hat{Q}_\kappa^{-1}\sum_{i=1}^n A_i\varepsilon_i\| \leqslant \|Q_\kappa^{-1}(\hat{Q}_\kappa - Q_\kappa)\| \cdot \|n^{-1}\hat{Q}_\kappa^{-1}\sum_{i=1}^n A_i\varepsilon_i\|$

$= O_p(\kappa n^{-1/2})O_p(\kappa^{1/2}n^{-1/2}) = O_p(\kappa^{3/2}n^{-1})$.

$\|J_{n_4}\| = \|Q_\kappa^{-1}(\hat{Q}_\kappa - Q_\kappa)n^{-1}\hat{Q}_\kappa^{-1}\sum_{i=1}^n A_i b_{k_0}(i)\| \leqslant \|Q_\kappa^{-1}(\hat{Q}_\kappa - Q_\kappa)\| \cdot \|n^{-1}\hat{Q}_\kappa^{-1}\sum_{i=1}^n A_i$ $b_{k_0}(i)\| = O_p(\kappa n^{-1/2})O_p(\kappa^{-2}) = O_p(\kappa^{-1}n^{-1/2})$.

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

Proof of lemma A.6,

$E(\sum_{i=1}^n \varepsilon_i K_h(z_i - z)x_i)^2 = E\sum_{i=1}^n \varepsilon_i^2 K_h^2(z_i - z)x_i x_i^\top + o_p(1)$

$= nh^{-1}\delta^2\nu_0(K)f_z(z)E(x_i x_i^\top \big| z_i = z)$.

Define $\mathcal{F}_t = \sigma(x_i, z_i, \varepsilon_{i-1}, i \leqslant t)$. By martingale central limit theorem,

$\sqrt{\frac{h}{n}}\sum_{i=1}^n \varepsilon_i K_h(z_i - z)x_i$ 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) $n^{-1}\sum_{i=1}^{n} R(z_i)^\top x_i K_h(z_i - z)(\frac{z_i-z}{h})^\ell x_i^\top = \frac{1}{2}h^2 S\gamma_1^{(2)}(z) f_z(z)\mu_{2+\ell}$.

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 $R^{(2)}(z_i|z_i = z) = \gamma_1^{(2)}(z)$, (b) could be easily proof by change-of-variable, the kernel theory and an application of Taylor's expansion.

Lemma A.8. $\|\frac{1}{n}\sum_{i=1}^{n} K_h(z_i - z)x_i \overline{A}(t_i)^\top\| = O_p(1)$

Proof of lemma A.8,

$E\|\frac{1}{n}\sum_{i=1}^{n} K_h(z_i - z)x_i\|^2 = O_p((nh)^{-1})$

$E\|\frac{1}{n}\sum_{i=1}^{n} K_h(z_i - z)x_i\|^2 = \frac{1}{n^2}\{\sum_{i=1}^{n} E x_i^2 K_h^2(z_i - z) + 2\sum_{i=1}^{n-1} E x_i x_{i-1} K_h^2(z_i - z) + \ldots\} = \frac{1}{n^2}\{\sum_{i=1}^{n} O(h^{-1}) + 2\sum_{i=1}^{n-1} b_1 O(h^{-1}) + \ldots\} = O(h^{-1})\frac{1}{n^2}\{n + 2(n-1)b_1 + \ldots\} = O((nh)^{-1})$.

The result holds from the same argument as for Lemma A.7. Define $G(z) = E\{x_i x_{i,1}|z_i = z\}f(z)$, $\xi_i = K_h(z_i - z)x_i P_\kappa^\top(t)x_{i,1}$, $C(z) = \int E\{x_i x_{i,1}|z_i = z\}P_\kappa^\top(t)f(z)dt$, $r_{n1} = \frac{1}{n}\sum_{i=1}^{n}\{\xi_i - E\xi_i\}$ and $r_{n2} = E\xi_1 - C(z)$.

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 allκ, such that $C^\top(z)C(z) \leqslant M$.

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

By the definitions of $C(z)$ and ξ_i,

$r_{n2} = EK_h(z_i - z)x_iP_\kappa^\top(t)x_{i,1} - \int G(z)P_\kappa^\top(t)dt = \int[\int\{G(z+\mu h)K(\mu) - G(z)K(\mu)\} d\mu]P_\kappa^\top(t)dt = \int[\int\{\frac{\partial G(z+\Delta)}{\partial z}\mu hK(\mu)\}d\mu]P_\kappa^\top(t)dt$ (Dominated convergence theorem) $= h\int \frac{\partial G(z)}{\partial z}P_\kappa^\top(t)dt(1+o_p(1))$ where Δ is between 0 and μh. Therefore, we obtain that $\|r_{n2}\|^2 = O(\kappa h^2) = O_P(1)$.

So that $\frac{1}{n}\sum_{i=1}^n K_h(z_i - z)x_iP_\kappa^\top(t)x_{i,1} = C(z) + r_{n1} + r_{n2} = O_P(1)$.

Lemma A.9. $\frac{1}{n}\sum_{i=1}^n\{\tilde{\beta}_0^*(t_i) + \tilde{\gamma_0}^*(z_i) + \tilde{\beta}_1^*(t_i)^\top x_i - \beta_0(t_i) - \gamma_0(z_i) - \beta_1(t_i)^\top x_i\}K_h(z_i - z)x_i = o_p(h^2)$.

Proof of lemma A.9,

By Lemma A.5, $\frac{1}{n}\sum_{i=1}^n\{\tilde{\beta}_0^*(t_i) + \tilde{\gamma_0}^*(z_i) + \tilde{\beta}_1^*(t_i)^\top x_i - \beta_0(t_i) - \gamma_0(z_i) - \beta_1(t_i)^\top x_i\}$ $K_h(z_i - z)x_i = \frac{1}{n}\sum_{i=1}^n\{\overline{A}(t_i)^\top\hat{B} - \beta_0(t_i) - \gamma_0(z_i) - \beta_1(t_i)^\top x_i\}K_h(z_i - z)x_i$

$= \frac{1}{n}\sum_{i=1}^n\{\overline{A}(t_i)^\top(\hat{B} - \theta_{\kappa 0}) - \overline{b_{\kappa 0}}(i)\}K_h(z_i - z)x_i$

$= \frac{1}{n}\sum_{i=1}^n[x_i\overline{A}(t_i)^\top K_h(z_i - z)(\frac{1}{n}Q_k^{-1}\sum_{j=1}^n \varepsilon_j A_j)] + \frac{1}{n}\sum_{i=1}^n[x_i\overline{A}(t_i)^\top K_h(z_i - z)\frac{1}{n}Q_k^{-1}$ $\sum_{j=1}^n A_j b_{\kappa 0}(j)\}] - \frac{1}{n}\sum_{i=1}^n[x_i\overline{b_{\kappa 0}}(i)K_h(z_i - z)] + \frac{1}{n}\sum_{i=1}^n[x_i\overline{A}(t_i)^\top K_h(z_i - z)R_n] = C_{n_1} + C_{n_2} + C_{n_3} + C_{n_4}$.

Arguments like those used to prove Lemma 7 in Horowitz (2004) show that, $\|C_{n_1}\| = o_p(h^2)$.

Arguments like those used to prove Lemma A.4 show that $E\|\frac{1}{n}Q_\kappa^{-1}\sum_{j=1}^n A_j b_{\kappa 0}\|^2 = O(\kappa^{-4})$, by Lemma A.8, $\|C_{n_2}\| \leqslant \|\frac{1}{n}\sum_{i=1}^n x_i\overline{A}(t_i)^\top K_h(z_i - z)\|\|\frac{1}{n}Q_k^{-1}\sum_{j=1}^n A_j b_{\kappa 0}\| \leqslant O_p(\kappa^{-2}) = o_p(h^2)$.

$\|C_{n_3}\| \leqslant \frac{1}{n}\sum_{i=1}^n \|x_i\|K_h(z_i - z)\max\overline{b_{\kappa 0}}(i) = O_p(1)O(\kappa^{-2}) = O_p(\kappa^{-2}) = o_p(h^2)$.

$\|C_{n_4} \leqslant \|\frac{1}{n}\sum_{i=1}^n x_i\overline{A}(t_i)^\top K_h(z_i - z)\|\|R_n\| = O_p(\kappa^2/n) = o_p(h^2)$.

Proof of Theorem 4.1,

To simplify the notation,

Recall $R(z_i) = \gamma_1(z_i) - \gamma_1(z) - \gamma_1^{(1)}(z)(z_i - z)$, $W_{ih}(z) = \begin{pmatrix} x_i \\ x_i \frac{(z_i - z)}{h} \end{pmatrix}$.

Define $\Delta_{\beta_0,\beta_1} = \beta_0(t_i) + \gamma_0(z_i) + \beta_1(t_i)^\top x_i - \tilde{\beta}_0^*(t_i) - \tilde{\gamma_0}^*(z_i) - \tilde{\beta}_1^*(t_i)^\top x_i$, $\Phi = \begin{pmatrix} \hat{\gamma}_1(z) - \gamma_1(z) \\ h\hat{\gamma}_1^{(1)}(z) - h\gamma_1^{(1)}(z) \end{pmatrix}$ and $\hat{\eta}_i = R(z_i)^\top x_i - W_{ih}(z)^\top \Phi + \Delta_{\beta_0,\beta_1}$,

then $W_{ih}(z)^\top \Phi = x_i^\top \hat{\gamma}_1(z) - x_i^\top \gamma_1(z) + x_i^\top \hat{\gamma}_1^{(1)}(z)(z_i - z) - x_i^\top \gamma_1^{(1)}(z)(z_i - z)$ and

$\hat{\eta}_i = \{x_i^\top \gamma_1(z_i) - x_i^\top \gamma_1(z) - x_i^\top \gamma_1^{(1)}(z)(z_i - z)\} - \{x_i^\top \hat{\gamma}_1(z) - x_i^\top \gamma_1(z) + x_i^\top \hat{\gamma}_1^{(1)}(z)(z_i - z) - x_i^\top \gamma_1^{(1)}(z)(z_i - z)\} + \{\beta_0(t_i) + \gamma_0(z_i) + \beta_1(t_i)^\top x_i - \tilde{\beta}_0^*(t_i) - \tilde{\gamma_0}^*(z_i) - \tilde{\beta}_1^*(t_i)^\top x_i\}$

$= x_i^\top \gamma_1(z_i) - \{x_i^\top \hat{\gamma}_1(z) + x_i^\top \hat{\gamma}_1^{(1)}(z)(z_i - z)\} + \{\beta_0(t_i) + \gamma_0(z_i) + \beta_1(t_i)^\top x_i - \tilde{\beta}_0^*(t_i) - \tilde{\gamma_0}^*(z_i) - \tilde{\beta}_1^*(t_i)^\top x_i\}$.

From equation 3.3, by taking the first derivative, we have

$\sum_{i=1}^n \{y_i - \tilde{\beta}_0^*(t_i) - \tilde{\gamma_0}^*(z_i) - \tilde{\beta}_1^*(t_i)^\top x_i - x_i^\top \hat{\gamma}_1(z) - x_i^\top \hat{\gamma}_1^{(1)}(z)(z_i - z)\} W_{ih}(z) K_{h_1}(z_i - z) = 0$

$\Rightarrow \sum_{i=1}^n \{\varepsilon_i + \beta_0(t_i) + \gamma_0(z_i) + \beta_1(t_i)^\top x_i + \gamma_1(z_i)^\top x_i - \tilde{\beta}_0^*(t_i) - \tilde{\gamma_0}^*(z_i) - \tilde{\beta}_1^*(t_i)^\top x_i - x_i^\top \hat{\gamma}_1(z) - x_i^\top \hat{\gamma}_1^{(1)}(z)(z_i - z)\} W_{ih}(z) K_{h_1}(z_i - z) = 0$

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

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

$I_{n_2} = \sum_{i=1}^n \hat{\eta}_i K_h(z_i - z) W_{ih}(z) = \sum_{i=1}^n R(z_i)^\top x_i K_h(z_i - z) W_{ih}(z) - \sum_{i=1}^n W_{ih}(z) W_{ih}(z)^\top \Phi K_h(z_i - z) + \sum_{i=1}^n [\beta_0(t_i) + \gamma_0(z_i) + \beta_1(t_i)^\top x_i - \tilde{\beta}_0^*(t_i) - \tilde{\gamma_0}^*(z_i) - \tilde{\beta}_1^*(t_i)^\top x_i] K_h(z_i - z) W_{ih}(z) = L_{n_1} - L_{n_2} + L_{n_3}$

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

From Lemma A.7.b, $n^{-1} L_{n_1} = n^{-1} \sum_{i=1}^n R(z_i)^\top x_i K_h(z_i - z) W_{ih}(z) = \frac{h^2}{2} f_z(z) \gamma_1^{(2)}(z)$

$$S \begin{pmatrix} \mu_2(K) \\ \\ \mu_3(K) \end{pmatrix} \Rightarrow n^{-1} \sum_{i=1}^n R(z_i)^\top x_i K_h(z_i - z) x_i = \frac{h^2}{2} f_z(z) \gamma_1^{(2)}(z) \mu_2(K) S.$$

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

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

$$= \{n^{-1} \sum_{i=1}^n \begin{pmatrix} x_i x_i^\top & x_i x_i^\top (\frac{z_i - z}{h}) \\ \\ x_i x_i^\top (\frac{z_i - z}{h}) & x_i x_i^\top (\frac{z_i - z}{h})^2 \end{pmatrix} K_h(z_i - z)\} \Phi = f_z(z) \begin{pmatrix} S & 0 \\ \\ 0 & \mu_2(K) S \end{pmatrix} \Phi.$$

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, $\sqrt{nh}(\hat{\gamma}_1(z) - \gamma_1(z) - \frac{h^2}{2} \mu_2(K) \gamma_1^{(2)}(z)) \to N(0, f_z(z)^{-1} \delta^2 S^{-1} \nu_0(K))$.

Proof of Theorem 4.2, Theorem 4.3 and Theorem 4.4,

Following the same argument as the proof of Theorem 4.1, we have $\sqrt{nh_2}[\hat{\beta}_1(t) - \beta_1(t) - \frac{h_2^2}{2} \mu_2(K) \beta_1^{(2)}(t)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, \delta^2 S_0^{-1} \nu_0(K)\}$, $\sqrt{nh_4}[\hat{\gamma}_0(z) - \gamma_0(z) - \frac{h_4^2}{2} \mu_2(K) \gamma_0^{(2)}(z)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, f_z(z)^{-1} \delta^2 \nu_0(K)\}$ and $\sqrt{nh_0}[\hat{\beta}_0(t) - \beta_0(t) - \frac{h_0^2}{2} \mu_2(K) \beta_0^{(2)}(t)\{1 + o_p(1)\}] \xrightarrow{d} N\{0, \delta^2 \nu_0(K)\}$.

APPENDIX B: SKETCH OF PROOFS

Theorem 5.1

Proof of Theorem 5.1,

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

(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 above lemma,

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

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

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

(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),

$\|n^{-2}\hat{Q}_k^{*-1}\sum_{i=1}^n\{A_i b_{k_0}(i)\}\|^2 = \|\hat{Q}_\kappa^{*-1}\Lambda\varrho/n^2\|^2 = n^{-4}\varrho^\top\Lambda^\top\hat{Q}_\kappa^{*-2}\Lambda\varrho \leqslant n^{-4}\lambda_{\max}(\hat{Q}_\kappa^{*-2})$
$\varrho^\top\Lambda^\top\Lambda\varrho = n^{-2}\lambda_{\max}(\hat{Q}_\kappa^{*-2})\varrho^\top\hat{Q}_\kappa^*\varrho \leqslant n^{-2}\lambda_{\max}(\hat{Q}_\kappa^{*-1})^2\lambda_{\max}(\hat{Q}_\kappa^*)\varrho^\top\varrho = O(\kappa^{-4}n^{-2})$.

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

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

$\hat{B} - \theta_{k_0} = n^{-2}Q_\kappa^{*-1}\sum_{i=1}^n A_i\varepsilon_i + n^{-2}Q_\kappa^{*-1}\sum_{i=1}^n A_i b_{k_0} + R_n$ *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.

$\hat{B}-\theta_{k_0} = n^{-2}Q_\kappa^{*-1}\sum_{i=1}^n A_i\varepsilon_i + n^{-2}Q_\kappa^{*-1}\sum_{i=1}^n A_i b_{k_0}(i) + n^{-2}(\hat{Q}_\kappa^{*-1} - Q_\kappa^{*-1})\sum_{i=1}^n$ $A_i\varepsilon_i + n^{-2}(\hat{Q}_\kappa^{*-1} - Q_\kappa^{*-1})\sum_{i=1}^n A_i b_{k_0}(i) = J_{n_1} + J_{n_2} + J_{n_3} + J_{n_4}$.

$\|J_{n_3}\| = \|Q_\kappa^{*-1}(\hat{Q}_\kappa^* - Q_\kappa^*)n^{-2}\hat{Q}_\kappa^{*-1}\sum_{i=1}^n A_i\varepsilon_i\| \leqslant \|Q_\kappa^{*-1}(\hat{Q}_\kappa^* - Q_\kappa^*)\| \cdot \|n^{-2}\hat{Q}_\kappa^{*-1}\sum_{i=1}^n$ $A_i\varepsilon_i\| = O_p(\kappa/n^{1/2})O_p(\kappa^{1/2}/n) = O_p(\kappa^{3/2}/n^{3/2})$.

$\|J_{n_4}\| = \|Q_\kappa^{*-1}(\hat{Q}_\kappa^* - Q_\kappa^*)n^{-2}\hat{Q}_\kappa^{*-1}\sum_{i=1}^n A_i b_{k_0}(i)\| \leqslant \|Q_\kappa^{*-1}(\hat{Q}_\kappa^* - Q_\kappa^*)\| \cdot \|n^{-2}\hat{Q}_\kappa^{*-1}$ $\sum_{i=1}^n A_i b_{k_0}(i)\| = O_p(\kappa/n^{1/2})O_p(\kappa^{-2}/n) = O_p(\kappa^{-1}/n^{3/2})$.

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

(a)$\frac{1}{n^2}\sum_{i=1}^n x_i x_i^\top R(z_i)^\top K_{h_1}(z_i - z) = \frac{h^2}{2} f_z(z) W_{\delta,2}\gamma_1^{(2)}(z)\mu_2(K)\{1 + o_p(1)\}$,

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

Proof of lemma A.13,

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

Lemma A.14. *By condition (B2)(i),* $\frac{1}{n^2}\sum_{i=1}^n(\beta_0(t_i) + \gamma_0(z_i) + \beta_1(t_i)^\top x_i - \tilde{\beta}_0{}^*(t_i) - \tilde{\gamma}_0{}^*(z_i) - \tilde{\beta}_1{}^*(t_i)^\top x_i)K_{h_1}(z_i - z)x_i = o_p(h_1^2)$.

Proof of lemma A.14,

$\frac{1}{n^2}\sum_{i=1}^n(\beta_0(t_i) + \gamma_0(z_i) + \beta_1(t_i)^\top x_i - \tilde{\beta}_0{}^*(t_i) - \tilde{\gamma}_0{}^*(z_i) - \tilde{\beta}_1{}^*(t_i)^\top x_i)K_{h_1}(z_i - z)x_i$ $= \frac{1}{n^2}\sum_{i=1}^n[x_i\overline{A_i}^\top K_{h_1}(z_i - z)(\frac{1}{n}Q_\kappa^{*-1}\sum_{j=1}^n \varepsilon_j A_j)] + \frac{1}{n^2}\sum_{i=1}^n[x_i\overline{A_i}^\top K_{h_1}(z_i - z)(\frac{1}{n}Q_\kappa^{*-1}\sum_{j=1}^n$ $A_j b_{\kappa 0})] - \frac{1}{n^2}\sum_{i=1}^n[x_i\overline{b_{\kappa 0}}(i)K_{h_1}(z_i - z)] + \frac{1}{n^2}\sum_{i=1}^n[x_i\overline{A_i}^\top K_{h_1}(z_i - z)R_n] = C_{n_1} + C_{n_2} +$ $C_{n_3} + C_{n_4}$.

By condition (B2)(i), note that $\frac{1}{n^2}\sum_{i=1}^n x_i x_i^\top K_{h_1}(z_i - z) = f_z(z)\mu_0(K)W + o_p(1) = O_p(1)$ 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 $E\|\frac{1}{n^2}Q_\kappa^{*-1}\sum_{j=1}^n A_j b_{\kappa 0}(i)\|^2$

$= O(\kappa^{-4}/n^2)$. $\|C_{n_2}\| \leqslant \|\frac{1}{n^2}\sum_{i=1}^n x_i \bar{A}_i^\top K_{h_1}(z_i - z)\|\|\frac{1}{n^2}Q_\kappa^{*-1}\sum_{j=1}^n A_j b_{\kappa 0}(i)\|$

$= O_p(\kappa^{1/2})O_p(\kappa^{-2}/n) = O_p(\frac{1}{\kappa^{3/2}n}) = o_p({h_1}^2)$.

$\|C_{n_3}\| \leqslant \|\frac{1}{n^2}\sum_{i=1}^n x_i K_{h_1}(z_i - z)\| \sup \|\bar{b}_{\kappa 0}(i)\| = O_p(n^{-1/2})O_p(\kappa^{-2}) = O_p(\frac{1}{\kappa^2 n^{1/2}}) = o_p({h_1}^2)$.

$\|C_{n_4}\| \leqslant \|\frac{1}{n^2}\sum_{i=1}^n x_i \bar{A}_i^\top K_{h_1}(z_i - z)\|\|R_n\| = O_p(\kappa^{1/2})O_p(\kappa^{3/2}n^{-3/2}) = O_p(\frac{\kappa^2}{n^{3/2}}) = o_p({h_1}^2)$.

Lemma A.15. $\frac{\sqrt{h_1}}{n}\sum_{i=1}^n \varepsilon_i x_i K_{h_1}(z_i - z) \xrightarrow{d} \sqrt{\nu_0(K) f_z(z)} \int_0^1 W_\delta(r) dW_\varepsilon(r)$.

Proof of above lemma,

Define $W_\varepsilon(r)$ is a Brownian motion on [0,1] with variance δ^2.

Note that $\sqrt{h_1/n}\sum_{i=1}^n K_{h_1}(z_i - z)\varepsilon_i \xrightarrow{d} N(0, \delta^2\nu_0(K) f_z(z)) = \sqrt{\nu_0(K) f_z(z)} W_\varepsilon(1)$.

This result holds from the same argument as for Theorem 2.1 in Cai and Park (2009).

Proof of Theorem 5.1,

$$\begin{bmatrix} \hat{\gamma_1}(z) \\ h_1 \hat{\gamma_1}^{(1)}(z) \end{bmatrix} = [A^*]^{-1}B^* =$$

$[\frac{1}{n^2}\sum_{i=1}^n W_{ih}(z)^\top W_{ih}(z) K_{h_1}(z_i - z)]^{-1}\big[\frac{1}{n^2}\sum_{i=1}^n (y_i - \tilde{\beta_0}^*(t_i) - \tilde{\gamma_0}^*(z_i) - \tilde{\beta_1}^*(t_i)^\top x_i)$

$W_{ih}(z) K_{h_1}(z_i - z)\big]$.

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

$\hat{\gamma_1}(z) = [f_z(z) W_{\delta,2}]^{-1}[\frac{1}{n^2}\sum_{i=1}^n (\beta_0(t_i) + \gamma_0(z_i) + \beta_1(t_i)^\top x_i - \tilde{\beta_0}^*(t_i) - \tilde{\gamma_0}^*(z_i) - \tilde{\beta_1}^*(t_i)^\top x_i)$

$K_h(z_i - z) x_i + \frac{1}{n^2}\sum_{i=1}^n \gamma_1^\top(z_i) x_i x_i^\top K_{h_1}(z_i - z) + \frac{1}{n^2}\sum_{i=1}^n \varepsilon_i x_i K_{h_1}(z_i - z)]$.

So that

$n\sqrt{h_1}[\hat{\gamma}_1(z) - \gamma_1(z) - \frac{h_1^2}{2}\mu_2(K)r^{(2)}(z)\{1 + o_p(1)\}]$

$= W_{\delta,2}^{-1} f_z(z)^{-1/2}\sqrt{\nu_0(K)}\int_0^1 W_\delta(r)dW_\varepsilon(r).$

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

Proof of above lemma,

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

Lemma A.18. *By condition (B2)(i),* $\frac{1}{n^2}\sum_{i=1}^n (\beta_0(t_i) + \gamma_0(z_i) + \gamma_1(z_i)^\top x_i - \tilde{\beta}_0{}^*(t_i) - \tilde{\gamma_0}^*(z_i) - \tilde{\gamma_1}^*(z_i)^\top x_i)K_{h_2}(t_i - t)x_i = o_p(h_2^2).$

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)* $\|\frac{1}{n^{3/2}}\sum_{i=1}^n P_\kappa(t_i)x_{i,1}K_{h_0}(t_i - t)\| = O_p(1).$

Proof of lemma A.19,

By condition, $\sup_{0\leqslant r\leqslant 1}\|x_{[nr]}/\sqrt{n}-W_\delta(r)\| = O(n^{-\theta_*}log^{\lambda_*}(n)) = o_p(1)$, 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 $\|\frac{1}{n}\sum_{i=1}^{n} P_\kappa(t_i) K_{h_0}(t_i-t)\| = O_p(1)$,

$\|\frac{1}{n^{3/2}}\sum_{i=1}^{n} P_\kappa(t_i)x_{i,1}K_{h_0}(t_i-t)\| = \|\frac{1}{n}\sum_{i=1}^{n} P_\kappa(t_i)\frac{x_{i,1}}{\sqrt{n}}K_{h_0}(t_i-t)\| \leqslant \|\frac{1}{n}\sum_{i=1}^{n} P_\kappa(t_i) W_\delta(t_i)K_{h_0}(t_i-t)\| + \|\frac{1}{n}\sum_{i=1}^{n} P_\kappa(t_i)\{\frac{x_{i,1}}{\sqrt{n}}-W_\delta(t_i)\}K_{h_0}(t_i-t)\| \leqslant \|\frac{1}{n}\sum_{i=1}^{n} P_\kappa(t_i)K_{h_0}(t_i-t)\|\sup\|W_\delta(t_i)\| + \|\frac{1}{n}\sum_{i=1}^{n} P_\kappa(t_i)K_{h_0}(t_i-t)\|\sup\|\frac{x_{i,1}}{\sqrt{n}} - W_\delta(t_i)\| = O_p(1)+o_p(1) = O_p(1)$.

Lemma A.20. *By condition (B1),* $\frac{1}{n}\sum_{i=1}^{n}(\gamma_0(z_i)^\top - \tilde{\gamma_0}^*(z_i)^\top + \beta_1(t_i)^\top x_i - \tilde{\beta_1}^*(t_i)^\top x_i + \gamma_1(t_i)^\top x_i - \tilde{\gamma_1}^*(t_i)^\top x_i)K_{h_0}(z_i-z) = o_p(h_0^2)$.

Proof of above lemma,

Define $\overline{A}(t_i,z_i) = [P_\kappa(t_i)^\top\cdot 0, P_\kappa(z_i)^\top, P_\kappa(t_i)^\top x_{i,1}, P_\kappa(z_i)^\top x_{i,1}, P_\kappa(t_i)^\top x_{i,2}, P_\kappa(z_i)^\top x_{i,2}, \cdots, P_\kappa(t_i)^\top x_{i,p}, P_\kappa(z_i)^\top x_{i,p}]^\top$, by the same argument in Lemma A.9, $\frac{1}{n}\sum_{i=1}^{n}(\gamma_0(z_i)^\top - \tilde{\gamma_0}^*(z_i)^\top + \beta_1(t_i)^\top x_i - \tilde{\beta_1}^*(t_i)^\top x_i + \gamma_1(t_i)^\top x_i - \tilde{\gamma_1}^*(t_i)^\top x_i)K_{h_0}(z_i-z) = \frac{1}{n}\sum_{i=1}^{n}[\overline{A}(t_i,z_i)^\top K_{h_0}(t_i-t)(\frac{1}{n^2}Q_k^{*-1}\sum_{j=1}^{n}\varepsilon_j A_j)] + \frac{1}{n}\sum_{i=1}^{n}[\overline{A}(t_i,z_i)^\top K_{h_0}(t_i-t)\frac{1}{n^2}Q_k^{*-1}\sum_{j=1}^{n} A_j b_{\kappa 0}(j)] - \frac{1}{n}\sum_{i=1}^{n}[\overline{b_{\kappa 0}}(i)K_{h_0}(t_i-t)] + \frac{1}{n}\sum_{i=1}^{n}[\overline{A}(t_i,z_i)^\top K_{h_0}(t_i-t)R_n] = C_{n_1}+C_{n_2}+C_{n_3}+C_{n_4}$.

Arguments like those used to prove Lemma 7 in Horowitz (2004) show that, $\|C_{n_1}\| = o_p(h_0^2)$.

By Lemma A.11 and A.19, $\|C_{n_2}\| = \|\frac{1}{n}\sum_{i=1}^{n}\overline{A}(t_i,z_i)^\top K_{h_0}(t_i-t)\|\|\frac{1}{n^2}Q_k^{*-1}\sum_{j=1}^{n} A_j b_{\kappa 0}(j)\| = O_p(n^{1/2}\kappa^{1/2})O_p(\kappa^{-2}n^{-1}) = o_p(h_0^2)$.

By the same argument, it can be shown that $\|C_{n_3}\| = O_p(\kappa^{-2}) = o_p(h_0^2)$.

$\|C_{n_4}\| = O_p(n^{1/2}\kappa^{1/2})O_p(\kappa^{3/2}n^{-3/2}) = O_p(\frac{\kappa^2}{n}) = o_p({h_0}^2)$.

Proof of Theorem 5.3,

It is easily to check that $\frac{1}{n}\sum_{i=1}^{n} K_{h_0}(t_i - t) \to 1$, $\frac{1}{n}\sum_{i=1}^{n} K_{h_0}(t_i - t)\beta_0(t_i) \to \beta_0(t) + \frac{1}{2}h_2^2\beta_0^{(2)}(t) + o_p(h_0^2)$, under (B2)(ii) $\frac{1}{nh_0^2}\sum_{i=1}^{n}(\gamma_0(z_i)^\top - \tilde{\gamma_0}^*(z_i)^\top + \beta_1(t_i)^\top x_i - \tilde{\beta}_1^*(t_i)^\top x_i + \gamma_1(z_i)^\top x_i - \tilde{\tilde{\gamma}}_1(z_i)^\top x_i)K_{h_0}(t_i - t) = o_p(1)$.

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

Proof of Theorem 5.4,

Following the same argument as the proof of Theorem 5.3, Theorem 5.4 can be proved.