The Pennsylvania State University

The Graduate School

Department of Statistics

VARYING-COEFFICIENT MODELS: NEW MODELS, INFERENCE PROCEDURES AND APPLICATIONS

A Thesis in

Statistics

by

Yang Wang

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2007

The thesis of Yang Wang was read and approved* by the following:

Runze Li
Associate Professor of Statistics
Thesis Adviser
Chair of Committee

Jingzhi Huang
Associate Professor of Finance
Co-Chair of Committee

James L. Rosenberger
Professor of Statistics
Head of the Department of Statistics

Damla Sentürk
Assistant Professor of Statistics

*Signatures are on File in the Graduate School.

Abstract

Varying-coefficient Models: New Models, Inference Procedures and Applications

Motivated by an empirical analysis of a data set collected in the field of ecology, we proposed nonlinear varying-coefficient models, a new class of varying-coefficient models. We further propose an estimation procedure for the nonlinear varying-coefficient models by using local linear regression, study the asymptotic properties of the proposed procedures, and establish the asymptotic normality of the resulting estimate. We also extend generalized likelihood ratio-type test (Fan, Zhang and Zhang, 2001) for the nonlinear varying-coefficient models for testing whether the coefficients really depend on a covariate. To assess the finite sample performance of the proposed procedures, we conduct extensive Monte Carlo simulation studies to assess the finite sample performance of the procedures. By Monte Carlo simulation, we empirically show the Wilks' phenomenon valid for the proposed generalized likelihood ratio test. That is, we empirically show that the asymptotic null distribution has a chi-square distribution with degrees of freedom which do not depend on the unknown parameters presented in the model under the null hypothesis. As new applications of varying coefficient models, we applied some existing procedures for some financial data sets. We demonstrated the varying-coefficient models are superior to an ordinary linear regression model, the commonly used model in finance research. We also apply the proposed estimation and inference procedure on the empirical study in the field of ecology.

Table of Contents

List of Tables

List of Figures

Acknowledgements

During my study at Penn State University, I received support from many wonderful individuals. My deep gratitude goes to everyone that has understood, encouraged, and helped me. I could not have accomplished my PhD study without their generous support.

My foremost thanks unquestionably go to my advisor Dr. Runze Li. I could not have asked for a more perfect mentor. His understanding, patience and encouragement helped me through obstacles in research, his insightful guidance many times let me find the pleasure and great interest among the challenging research work. I especially thank him for leading me into the economic and financial applications of statistics, the field where I have great interest and landed my industrial profession. I often find his great personality an inspiration to my personal development and maturity, which certainly is a life-long asset.

I am also extremely indebted to Dr. Jingzhi Huang, my committee member from Finance Department. Dr. Huang generously outreached his duty and put considerable amount of time in discussion and insightful guidance of the financial application part of my dissertation. I truly find his support encouraging and his introduction of opportunity in financial research an invaluable experience and asset in my future professional life.

My gratitude also goes to two other committee members Dr. James L. Rosenberger and Dr. Demla Sentürk, who took time from their busy schedule to read, react to, and edit this dissertation. As immediate former department head, Dr. Rosenberger continuously provided the kindest assistance of all sorts and made me succeed in the PhD program.

I need to sincerely mention and thank a special faculty member, Dr. Naomi Altman, who provided me the most generous mentoring in academics and profession, also, in some occasions, personally. Her assistance spanned throughout my entire PhD study.

My fellow graduate students are invaluable to my PhD study and dissertation writing such as discussions that lead to understanding of difficult concepts. Of them, I especially thank Mian Huang, also a personal friend, for his generosity of putting substantial time and energy in discussions of the theoretical part of my dissertation, which I found inconceivably inspiring and helpful.

Finally, I want to express my gratitude to my parents, my brother, and Jessie for their unparalleled love.

The research was partially funded by National Science Foundation grants DMS 0348869 and CCF 0430349, and National Institute on Drug Abuse, NIH, P50-DA10075. I want to sincerely thank them for the financial support from these grants.

Chapter 1

Introduction

As the most commonly used statistical technique, regression analysis has been utilized to explore the association between dependent and independent variables and further to identify how independent variables impact the dependent variable. A linear predictor is assumed in both linear regression models and generalized linear regression models; however, the assumption is restrictive and thus may produce biased estimation if the assumption is violated. Many modeling techniques have been proposed in aim to relax the model assumptions and widen the model applicability. These works include Hastie and Tibshirani (1990), Green and Silverman (1994), Wand and Jones (1995) and Fan and Gijbels (1996), and among others. For high dimensional covariates, it becomes very difficult to estimate the mean regression function using fully nonparametric models in an efficient way due to "curse of dimensionality". Additive models (Breiman and Friedman, 1985; Hastie and Tibshirani, 1990) are among the powerful approaches available to avoid the "curse of dimensionality".

Varying-coefficient models are a natural alternative to the additive model and have greatly widened the scope of application by allowing the model coefficients to vary over certain covariates, such as time and temperature. Among various useful nonparametric regression models, varying coefficient models can be used to explore features in high dimensional data. The varying-coefficient models have received much attention recently. Varying-coefficient models and their variations, such as functional linear models, have been used in the analysis of longitudinal data, see, for example, Brumback and Rice (1998), Hoover, Rice, Wu and Yang (1998), Wu, Chiang and Hoover (1998), Fan and Zhang (2000), Chiang, Rice and Wu (2001), Huang, Wu and Zhou (2002) among others. Cai, Fan and Li (2000) proposed an efficient estimation and inference procedure for generalized varying-coefficient models with a known link function. Cai, Fan and Yao (2001) proposed an adaptive estimation procedure for generalized varying-coefficient models with an unknown link function. Time-dependent coefficient models have also been explored for nonlinear time series applications, see, for example, Chen and Tsay (1993), Cai, Fan and Yao (2000) and Cai, Yao and Zhang (2001). In this dissertation, we shall propose a new class of varying coefficient models. Let us introduce some motivation first.

This work was motivated by an empirical study in the field of ecology. This data set was collected within the AmeriFlux network during summer growth seasons (from June 1 to August 31) of years 1993 — 1995 at the Walker Branch Watershed Site in eastern Tennessee (35.96^{o}N, 84.29^{o}W). Detailed analysis of this data set using the proposed procedures will be given in Chapter 5. Here, we give only a brief introduction. It is known that sunlight intensity affects the rate of photosynthesis in an ecosystem. Since leaves absorb carbon dioxide (CO_2) during the course of photosynthesis, the *Net Ecosystem Exchange* of CO_2, denoted by NEE, is used to measure the level of photosynthetic activity in a natural ecosystem. Photosynthetic rate as measured by NEE is dependent on the amount of *Photosynthetically Active Radiation* available to an ecosystem, denoted by PAR.

It is believed based on empirical studies that the relationship between NEE and PAR is nonlinear (Montheith, 1972)and can be characterized by the following model

$$\mathrm{NEE} = R - \frac{\beta_1 \mathrm{PAR}}{\mathrm{PAR} + \beta_2} + \varepsilon, \tag{1.1}$$

where ε is random error with zero mean, and R, β_1 and β_2 are unknown parameters with physical interpretations. Specially, R is the dark respiration rate, β_1 is the light-saturated

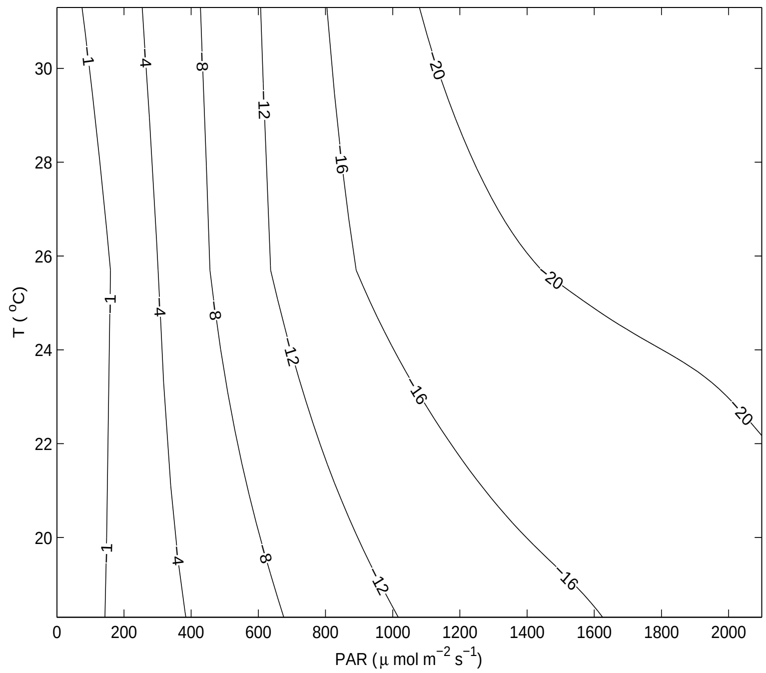

Figure 1.1: *Contour Plot Using 2-dimensional Kernel Regression*

net photosynthetic rate, and β_1/β_2 is the apparent quantum yield. The empirical NEE-PAR relationship in (1.1) has been applied widely to canopy levels to assess net primary productivity (NPP) since remote sensing data became available (see, for instance, Montheith, 1972, and more recent work by Ruimy et al., 1999 and references therein). The model in (1.1) was originally adopted because the data were collected from a laboratory in which climate variables, such as temperature and moisture availability, can be well controlled. However, the temperature for an ecosystem cannot be controlled, and the parameters R, β_1 and β_2 likely depends on the temperature. To illustrate this, 2-dimensional kernel smoothing regression is used to estimate the regression function of the NEE on the PAR and the temperature. Figure 1.1 displays the contour plot of regression curve. From Figure 1.1, the parallel pattern of the contour curves with low PAR values implies that there is little temperature effect on NEE for low PAR, while the non-parallel pattern of the contour curves with high PAR values indicates the strong temperature effects. Figure 1.2 (a) depicts the estimated regression

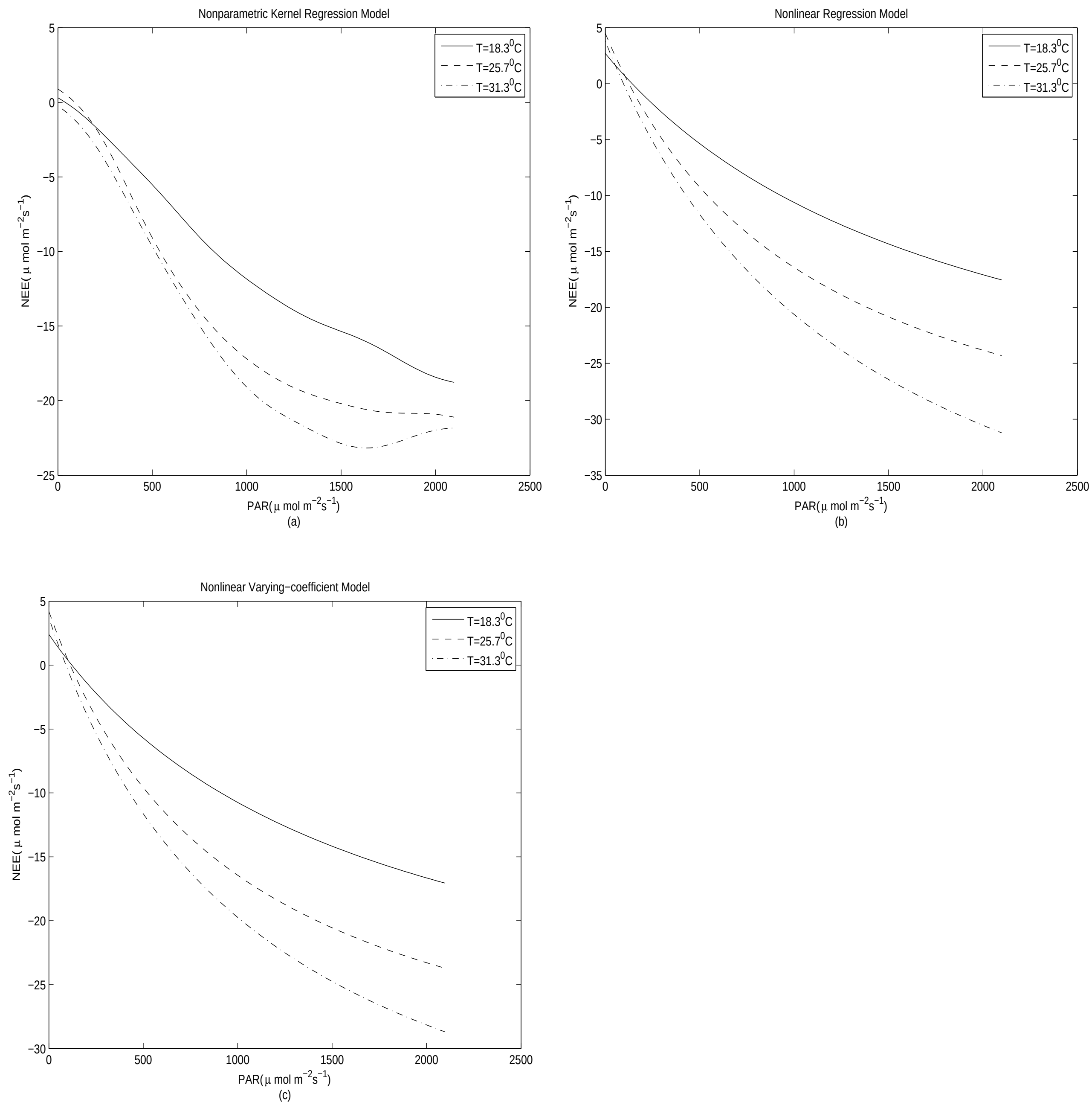

Figure 1.2: *Plot of NEE vs PAR using nonparametric 2-d kernel regression, nonlinear regression, and nonlinear varying-coefficient models for temperatures 18.3°C, 25.7°C and 31.3°C, which are the three sample quartiles of temperatures.*

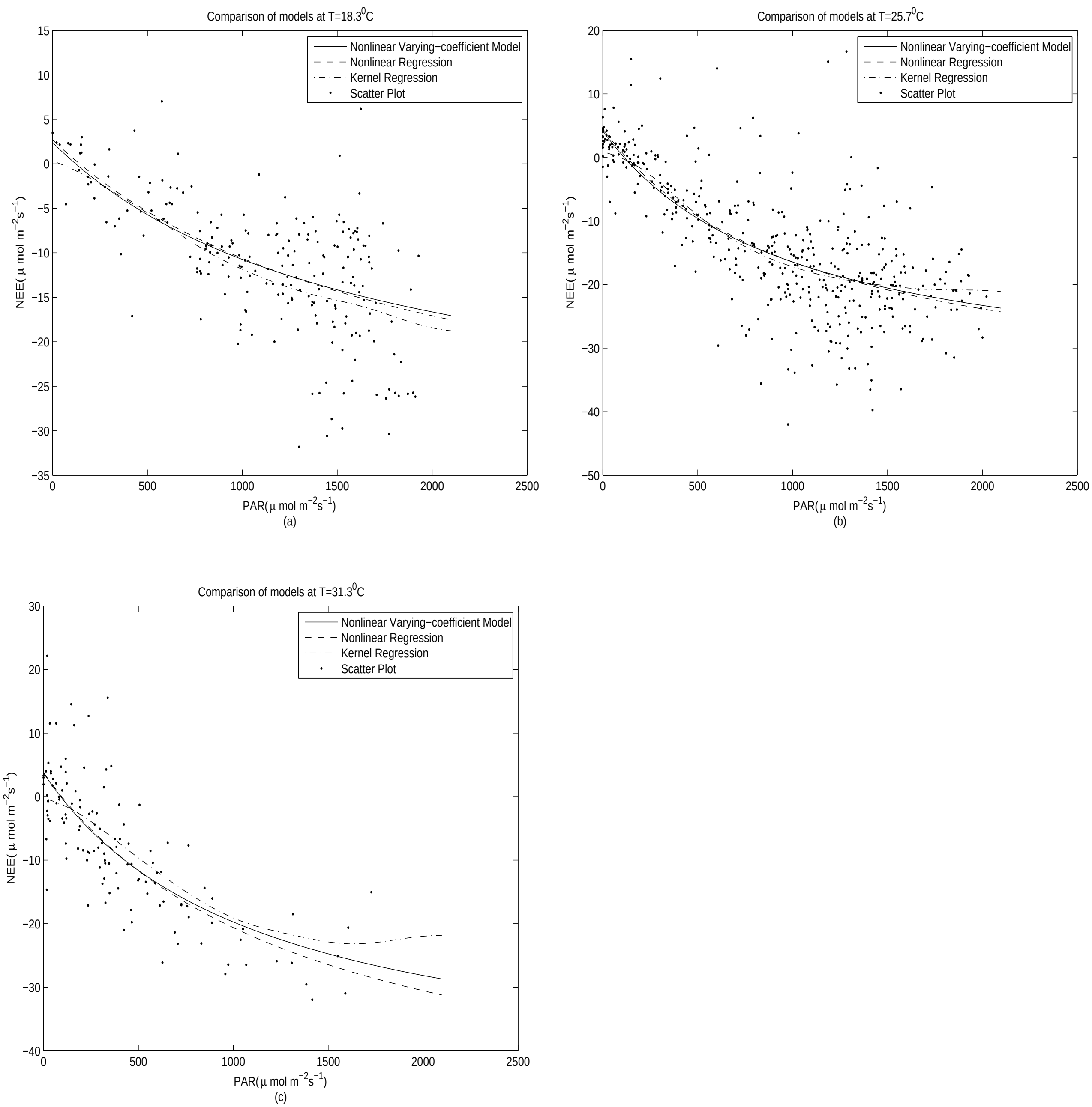

Figure 1.3: *The scatter plots of NEE vs PAR for temperatures 18.3^{o}C, 25.7^{o}C and 31.3^{o}C, which are the three sample quartiles of temperatures. The dash-dotted, solid and dashed lines are plots of NEE vs PAR using nonparametric 2-d kernel regression, nonlinear regression, and nonlinear varying-coefficient models for temperatures 18.3^{o}C, 25.7^{o}C and 31.3^{o}C.*

function of NEE on PAR, given three different values of temperature. The pattern shown in Figure 1.2 (a) implies that the parameters R, β_1, and β_2 depend on temperature. This leads us to consider an alternative model:

$$\text{NEE} = R(T) - \frac{\beta_1(T)\text{PAR}}{\text{PAR} + \beta_2(T)} + \varepsilon, \tag{1.2}$$

where T stands for temperature. Model (1.2) is distinguished from model (1.1) in that the unknown coefficients are allowed to depend on temperature. As a comparison, we further applied model (1.1) for data subsets with temperature 18.3 ± 0.5^oC, 25.7 ± 0.5^oC and 31.3 ± 0.5^oC, respectively, the resulting three regression curves are depicted in Figure 1.2(b). Figure 1.2(c) depicts the resulting regression curves using model (1.2) with the newly proposed estimation procedures in Chapter 3. The patterns in Figure 1.2 (b) and (c) are similar, while the regression curve for $T = 31.3^oC$ shows non-monotone decreasing pattern. To understand why the non-monotone decreasing pattern appears, we rearrange the regression curve plots by temperature and add the scatter plot of data subsets for each temperature in Figure 1.3. From Figure 1.3 (c), we can see that the non-monotone decreasing pattern for 2-dimensional kernel regression curve is due to data sparsity and a few influence data points.

Let y be a response variable and both u and $\mathbf{x}$ be covariates. In this dissertation, we consider

$$y = f(\mathbf{x}, \boldsymbol{\beta}(u)) + \varepsilon, \tag{1.3}$$

where $\boldsymbol{\beta}(\cdot)$ is the unknown regression coefficient function, ε is a random error with zero mean and finite variance. Model (1.3) includes many useful models as special cases. For instance, model (1.2) can be written in the form of (1.3) with u being T. In the literature, a functional linear model is defined by

$$y = \mathbf{x}^T \boldsymbol{\beta}(u) + \varepsilon,$$

and a generalized varying coefficient model is defined by

$$g^{-1}\{E(y|u, \mathbf{x})\} = \mathbf{x}^T \boldsymbol{\beta}(u),$$

where g is a known link function. In these two models, it is assumed that $\mathbf{x}$ and $\boldsymbol{\beta}(\cdot)$ have the same dimension. Both the functional linear model and the generalized varying-coefficient model are special cases of model (1.3). We refer model (1.3) to as *nonlinear varying-coefficient model* as we focus on the case in which $f(\cdot, \boldsymbol{\beta})$ is nonlinear function of $\boldsymbol{\beta}$. Varying-coefficient models have been existed in the literature long time ago. They become popular in the statistical literature since the systematic introduction in Hastie and Tibshirani (1993).

It is of importance to make a distinction between model (1.3), *nonlinear varying-coefficient model*, and the following nonlinear model

$$y = g(\mathbf{x}, u, \boldsymbol{\beta}) + \varepsilon \tag{1.4}$$

where g is a nonlinear function, y is a response variable and both u and $\mathbf{x}$ are covariates, $\boldsymbol{\beta}$ are the unknown coefficients that take constant values, and ε is a random error with zero mean and finite variance. Clearly, model (1.4) is a nonlinear regression model with u and $\mathbf{x}$ forming the set of covariates. We can claim that model (1.3) is a more relaxed and general form than model (1.4), in the sense that for any specific model from the family of model (1.4), there is a corresponding model in the family of model (1.3) with a more relaxed and general form. The following two examples illustrate this assertion.

For the first example, from the family of model (1.4) we choose the following statistical model

$$y = x\beta_1 + u\beta_2 + \varepsilon \tag{1.5}$$

then model (1.6) from the family of model (1.3)

$$y = x\beta_1 + \beta_2(u) + \varepsilon \tag{1.6}$$

is clearly a more relaxed and general statistical model than (1.5). For another example, the following nonlinear model (1.7) from the family of (1.4)

$$y = \frac{u}{x + \beta} + \varepsilon \tag{1.7}$$

is less relaxed than the nonlinear varying-coefficient model (1.8)

$$y = \frac{\beta_1(u)}{x + \beta_2} + \varepsilon \tag{1.8}$$

This thesis is organized as follows. Chapter 2 will provide a detailed and technical review of varying-coefficient models and their applications. Some existing estimation procedures are also reviewed in Section 2.2. Existing testing procedures will be summarized in Section 2.3. In Chapter 3, we systematically study nonlinear varying-coefficient models. We propose an estimation procedure for the nonlinear varying coefficient models by local linear regression techniques. Unlike the "linear" varying-coefficient models, optimization of nonlinear regression model is more challenging than that for existing varying coefficient models because there is no closed form for the resulting estimate and/or the objective function is typically nonconvex. We discuss computational issues related to the proposed estimation procedures. In practice, it is of interest to test whether some coefficient is invariant over the covariate u. This type of hypothesis testing problem is different from traditional hypothesis testing in that the parameter space under the alternative hypothesis is infinite. This poses many challenges in dealing with such statistical hypotheses. We will extend the generalized likelihood ratio test (Cai, Fan and Li, 2000 and Fan, Zhang and Zhang, 2001) for such hypotheses. We derive the asymptotic normality for nonlinear varying-coefficient models. In Chapter 4, we conduct extensive Monte Carlo simulation studies to assess the finite sample performance of the proposed procedures. We will also study the statistical sampling property of the proposed generalized likelihood ratio test. Chapter 5 applies the proposed nonlinear varying-coefficient model estimation and generalized likelihood ratio testing (GLRT) procedures to an empirical ecological data set. It is of scientific interest to test the hypothesis whether the relationship between NEE and PAR really depends on temperature T. Chapter 6 presents a new application of varying-coefficient models in the field of finance research. In this chapter, we apply existing estimation procedures for analysis of data sets collected from a finance study. We proposed new prediction procedures to varying-coefficient models and demonstrate varying-coefficient model are superior to the ordinary linear regression model in terms of model fitting and model prediction.

Chapter 2

Literature Review

In this chapter, we briefly review the literature on varying-coefficient models, including those proposed varying-coefficient models of various kinds, the fundamental statistical techniques for estimation procedures to varying-coefficient models, the estimation procedures for varying-coefficient models, and hypothesis testing. The aim for the chapter is to outline the framework based upon which we propose the statistical inference procedures for the newly proposed nonlinear varying-coefficient models (Model 1.3) in Chapter 1.

2.1 Varying-coefficient Models

Let Y be a response variable and $X_1, X_2, ..., X_p$ be covariates, the ordinary linear regression model is

$$Y = \mathbf{X}^T \boldsymbol{\beta} + \varepsilon, \tag{2.1}$$

where $\mathbf{X} = (X_1, X_2, X_3, ..., X_p)^T$ with $X_1 \equiv 1$ to include an intercept term, $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)$, a dimensional unknown parameter vector. Furthermore, it is assumed that $E(\varepsilon|\mathbf{X}) = 0$.

In model (2.1), the regression coefficient $\boldsymbol{\beta}$ is assumed to be constant. However, for certain situations, this assumption may be too strong, and limit the application of linear regression models. For example, in longitudinal studies, financial market studies, economics, and some ecological studies, the regression coefficient $\boldsymbol{\beta}$ may possibly change with some underlying covariates; these covariates could be time, temperature, or geographical locations.

With the aim to increase the flexibility of linear regression models, to broaden the application of linear regression, and to reduce modeling bias, varying coefficient models allow the coefficients $\boldsymbol{\beta}$ to be smooth functions of some covariate u:

$$Y = \boldsymbol{\beta}(u)^T\mathbf{X} + \varepsilon, \tag{2.2}$$

where $E(\varepsilon|\mathbf{X}, u) = 0$.

A more general form of varying-coefficient model can be presented as

$$Y = \beta_1(u_1)X_1 + \beta_2(u_2)X_2 + \beta_3(u_3)X_3 + ... + \beta_p(u_p)X_p + \varepsilon \tag{2.3}$$

Model (2.3) further allows β_1, β_2 ,..., β_p to be smooth and unspecified functions of different covariates u_1, u_2,..., u_p, as compared to the case of depending on a common covariate u in model (2.2).

As a multivariate extension, an even more general form of a varying-coefficient model takes the form

$$Y = \beta_1(\mathbf{u}_1)X_1 + \beta_2(\mathbf{u}_2)X_2 + \beta_3(\mathbf{u}_3)X_3 + ... + \beta_p(\mathbf{u}_p)X_p + \varepsilon \tag{2.4}$$

Model (2.4) says β_1, β_2,..., β_p are smooth and unspecified functions of different multivariate covariates $\mathbf{u}_i's$, and $\mathbf{u}_i = (u_{i1}, u_{i2},..., u_{ip})^T$.

For all of the above varying-coefficient models, they imply a special interaction relationship between the covariate u_1, u_2,..., u_p, and the predictors X_1, X_2,..., X_p. Conditioning on u_1, u_2,..., u_p, varying-coefficient models become linear regression models with constant coefficients, and at different values of u_1, u_2,..., u_p, the relationship between the dependent variable Y and predictors X_1 , X_2 ,..., X_p changes accordingly. Therefore, conditioning on the covariates $\mathbf{u}$, the coefficient functions can be interpreted as those in an ordinary linear regression model.

Model (2.2) is the most popular varying coefficient model in the literature and have been well studied. It implies that all the coefficients β_1, β_2,..., β_p depend on a common covariate u. For example, in a time-varying-coefficient model, the covariate u is time; in a temperature varying-coefficient model, the covariate u is temperature. See Hastie and Tibshirani (1993) and Fan and Zhang (2000). For applications, Hoover, *et al.* (1998), Brumback and Rice (1998) and Fan and Zhang (2000) applied varying-coefficient models to longitudinal data and allow the model structure to change over time. Depending on the nature of underlying covariates, model (2.2) can be extended to model (2.3) and (2.4). In model (2.3), the coefficients β_1, β_2,..., β_p may depend on different covariates u_1, u_2,..., u_p. For example, temperature t and radiation r for two predictors X_1 and X_2, respectively. In Model (2.4), the coefficients β_1, β_2,..., β_p are allowed to depend on covariates $\mathbf{u}_i = (u_{i1}$, u_{i2},..., $u_{ip})^T, i = 1, 2, ..., p$, which are multi-dimensions. An example would be the spatial or geographical locations as the covariate for the varying-coefficient model.

In this chapter, we focus on model (2.2) in which covariate is univariate and the same for all β_1, β_2,..., β_p, since the extension to multivariate covariate is straightforward. However, the implementation with multivariate covariate $\mathbf{u}$ may be very difficult due to the so-called "curse of dimensionality'".

2.1.1 Generalized Varying-coefficient Models

Generalized linear models (GLIM) were systematically studied by (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989) and assumes that the regression function $m(\mathbf{x}) = E(Y|\mathbf{X} = \mathbf{x})$ satisfies the following relation:

$$\eta(\mathbf{x}) = g\{m(\mathbf{x})\} = \mathbf{x}^T\boldsymbol{\beta} \tag{2.5}$$

where $g(\cdot)$ is a known link function that transforms the mean response function $m(\mathbf{x})$ into linear predictor $\eta(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}$. Therefore, a generalized linear model (GLIM) is composed of three components: the random component $Y|\mathbf{X} = \mathbf{x}$; the link function $g(\cdot)$; and the systematic component $\eta(\mathbf{x}) = \mathbf{x}^T\boldsymbol{\beta}$. In particular, GLIM allows that the random component $Y|\mathbf{X} = \mathbf{x}$ belongs to exponential family with density

$$f_y(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}, \tag{2.6}$$

where $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ are specific functions, and θ is referred to as canonical parameter and ϕ is referred to as dispersion parameter. Since the mean function $m(\mathbf{x}) \equiv E(Y|\mathbf{X} = \mathbf{x}) = b'(\theta)$ is a function of θ alone, θ is the parameter of interest; ϕ is usually regarded as a nuisance. A canonical link is defined to be $g(\cdot) = b'^{-1}(\cdot)$. With canonical link we have $\theta = \eta = \mathbf{x}^T\boldsymbol{\beta}$.

For linear models, the random component $Y|\mathbf{X} = \mathbf{x}$ is assumed to be $N(\mu, \sigma^2)$ and the link function $g(t) = t$ which is also the canonical link. For the response Y being binomial response (including binary response) or count response, it can be fitted by binomial distribution or Poisson distribution, respectively; the canonical link for logistic model and Poisson log-linear model are $g(\mu) = \log\left(\frac{\mu}{1-\mu}\right)$ and $g(\mu) = \log(\mu)$, respectively.

A generalized varying-coefficient model(Cai, Fan and Li 2000) has the form

$$\eta(u,\mathbf{x}) = g\{m(u,\mathbf{x})\} = \mathbf{x}^T\beta(u), \tag{2.7}$$

where $g(\cdot)$ is a link function, $\mathbf{x} = (x_0,\, x_1\, x_2,\ldots,\, x_{d-1})^T$ is the vector of independent variables, and $m(\mathbf{u},\mathbf{x})$ is the mean regression function of the response variable Y given that the covariates $\mathbf{U} = \mathbf{u}$ and $\mathbf{X} = \mathbf{x}$. The link function $g(\cdot)$ transforms the mean regression function $m(\mathbf{u},\mathbf{x})$ into a linear predictor. Clearly, (2.7) is an extension of model (2.2).

Many existing models can be regarded as a special case of model (2.7). For instance, when all regression coefficient functions are constant, then model (2.7) becomes a GLIM. By allowing intercept function varying over u and other coefficient functions being constant functions of u, model (2.7) reduces to a generalized partially linear model (Carroll, Fan, Gijbels, and Wand 1997). As a special case of generalized partially linear model, a partially linear model is given

$$Y = \alpha(u) + \mathbf{X}^T\beta + \varepsilon \tag{2.8}$$

This model has been studied by Chen (1988), Green and Silverman(1994) and Speckman (1988) and among others.

Thus, the generalized varying coefficient model greatly widens the applicability of linear models, generalized linear models, partially linear models and generalized linear models by relaxing the model assumptions. The generalized varying coefficient models have been systematically studied in Cai, Fan and Li (2000).

2.1.2 Applications of Varying-coefficient Models in Sciences

Varying-coefficient models have been popular in longitudinal data and panel data study. See Hoover (1998), Brumback and Rice (1998), Fan and Zhang (1998), and Fan and Zhang (2000). In a longitudinal data study, suppose there are n subjects, and for the i^{th} subject, data $\{y_{i1}(t),\ x_{i1}(t),\ x_{i2}(t),\ldots,\ x_{i,p}(t)\}$ were collected at $t = t_{ij}$, $j = 1,\ 2,..,\ n_j$. A time-varying coefficient model is defined as

$$Y_i(t) = \beta_1(t)X_{i1} + \beta_2(t)X_{i2} + \beta_3(t)X_{i3} + ... + \beta_p(t)X_{i,p} + \varepsilon_i(t). \tag{2.9}$$

The time-varying coefficient model has been used to explore the possible time-dependent effects. The time-varying coefficient model has been extended to other research fields. Chen and Tsay (1993) explored nonlinear time series applications, and Cai, Fan, and Yao (1998) have provided statistical inferences on the functional-coefficient autoregressive models. Also see Cai (2000) and Xia and Li (1999) for applications in functional-coefficient nonlinear time series. Cai and Tiwari (2000) applied varying-coefficient models in enviromental study. Hong and Lee (1999), Lee and Ullah (1999) studied the applications of varying-coefficient models in finace and econometrics. Cederman and Penubarti (1999) has applications of varying-coefficient model in political sciences. Time-varying coefficient Cox model has been studied by Cai and Sun (2003) and Tian, Zucker and Wei (2005).

2.2 Estimation procedures

2.2.1 Local Polynomial Regression

We first introduce the fundamental ideas of local polynomial regression. Suppose that we have a random sample $(x_1, y_1), ..., (x_n, y_n)$ from a nonparametric regression model

$$Y = m(X) + \varepsilon, \tag{2.10}$$

where $E(\varepsilon|X) = 0$ and $var(\varepsilon|X = x) = \sigma^2(x)$. We call $m(x) = E(Y|X = x)$ a regression function, and $m(x)$ is assumed to be smooth but not have a parametric form in the context of nonparametric regression. Our goal is to find the estimate $\widehat{m}(x)$ of $m(x)$.

In the setting of local polynomial regression, we apply the Taylor expansion of $m(z)$ for z in a neighborhood of x:

$$m(z) \approx \sum_{j=0}^{p} \frac{m^{(j)}(x)}{j!}(z-x)^j \equiv \sum_{j=0}^{p} \beta_j (z-x)^j \tag{2.11}$$

Therefore, for datum x_i in a neighborhood of x, we have

$$m(x_i) \approx \sum_{j=0}^{p} \beta_j (x_i - x)^j \equiv \mathbf{x}_i^T \boldsymbol{\beta} \tag{2.12}$$

where $\mathbf{x}_i = (1,\ (x_i - x), ...,\ (x_i - x)^p)^T$ and $\boldsymbol{\beta} = (\beta_0,\ \beta_1, ...,\ \beta_p)^T$. Intuitively, a datum point closer to x carries more information of $m(x)$; while a datum point remote from x carries less information of the value of m(x). We therefore use a locally weighted polynomial regression

$$\sum_{i=0}^{n} (y_i - \mathbf{x}_i^T \beta)^2 K_h(x_i - x) \tag{2.13}$$

where $K_h(x_i - x) \equiv h^{-1}K(\frac{x_i - x}{h})$ and $K(\cdot)$ is called a kernel function satisfying with $\int K(x) = 1$ and h is a positive number, called a bandwidth or a smoothing parameter. Commonly used kernel functions are Gaussian Kernel:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{x^2}{2}), -\infty < x < \infty \tag{2.14}$$

and symmetric beta family

$$K(x) = \frac{1}{Beta(1/2, \gamma + 1)}(1 - x^2)^{\gamma}_{+}, \gamma = 0, 1, 2, ..., -1 \leq x \leq 1, \tag{2.15}$$

where $+$ denote the positive part that is taken before exponentiation, and $Beta(\cdot, \cdot)$ denotes a Beta function and the support for Beta family type kernel functions have support of $[-1, 1]$. In addition, $\gamma = 0, 1, 2$ and 3 are referred to as the uniform, the Epanechnikov, the biweight and the triweight kernel functions. As discussed in Marron and Nolan (1988), the kernel function is uniquely determined up to a scale factor.

The local polynomial estimator $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1, ..., \widehat{\beta}_p)^T$ is

$$\widehat{\beta} = \arg\min_{\beta} \sum_{i=0}^{n} (y_i - \mathbf{x}_i^T \beta)^2 K_h(x_i - x). \tag{2.16}$$

Since, by definition, $m(x) \equiv \beta_0$, the estimated regression function $\widehat{m}(x) = \widehat{\beta}_0(x)$. Note that functional notation $\widehat{\beta}_0(x)$ is used here to emphasize that $\widehat{\beta}_0$ is a function of x. In addition, an estimator for the v-th order derivative of regression function $m(x)$ at x is

$$\widehat{m}^{(v)}(x) = v!\widehat{\beta}_v(x). \tag{2.17}$$

2.2.2 Definition of Kernel Regression and Local Linear Regression

Kernel regression and local linear regression are special cases for local polynomial regression. When $p = 0$, the local polynomial regression is referred to as kernel regression.

Table 2.1: Pointwise asymptotic bias and variance

Method	Bias	Variance
Nadaraya-Watson	$(m''(x) + \frac{2m'(x)f'(x)}{f(x)})b_n$	V_n
Gasser-Műller	$m''(x)b_n$	$1.5V_n$
Local Linear	$m''(x)b_n$	V_n

Here, $b_n = \frac{1}{2}\int_{-\infty}^{\infty} u^2 K(u)du\ h^2$ and $V_n = \frac{\sigma^2(x)}{f(x)nh}\int_{-\infty}^{\infty} K(u)du$.

We can view kernel regression as a local constant fitting. However, kernel regression does not estimate the derivatives of regression function $m(x)$. The kernel regression estimator is

$$\widehat{m}_h(x) = \frac{\sum_{i=1}^n K_h(X_i - x)Y_i}{\sum_{i=1}^n K_h(X_i - x)}. \tag{2.18}$$

This estimator is referred to as Nadaraya-Watson kernel regression estimator. See Nadaraya (1964) and Watson (1964).

Noteworthy is another well-known kernel type of estimator, Gasser-Műller kernel estimator (Gasser and Műller 1979 and 1984).

$$\widehat{m}_h(x) = \sum_{i=1}^n \int_{s_{i-1}}^{s_i} K_h(u - x)duY_i,$$

where $\mathrm{s}_i = (X_i + X_{i+1})/2$ assuming $X_i < X_{i-1}$, and $X_0 = 0$ and $X_{n+1} = +\infty$.

For $p = 1$, the local polynomial regression is referred to as *local linear regression*. *Local linear regression* not only estimates the regression function $m(x)$, but also estimates the first derivative $m'(x)$ of regression function $m(x)$.

Table 2.1, adapted from (Fan, 1992), summarizes the asymptotic behavior of the Nadaraya-Watson estimator, Gasser-Müller estimator, and local linear estimator.

The local linear fitting has several desirable properties, including high statistical efficiency in an asymptotic minimax sense (Fan 1993), design adaption (Fan 1993), and nice boundary behavior (Fan and Gijbels 1996; Ruppert and Wand 1994).

2.2.3 Estimation Procedure for Varying-coefficient Models

Suppose that the coefficients β_1,..., β_p are smooth functions of the covariate u. Let $(u_1, \mathbf{x}_1, y_1)$,..., $(u_n, \mathbf{x}_n, y_n)$ be a random sample from

$$Y = \mathbf{X}^T \boldsymbol{\beta}(u) + \varepsilon \tag{2.19}$$

We apply the local linear regression technique to estimate the coefficient functions $\beta_j(u)$, $j = 1$, 2,..., p, For a given u_0, we locally, linearly approximate the coefficient functions $\beta_j(u)$, $j = 1$, 2,..., p

$$\beta_j(u) \approx \beta_j(u_0) + \beta_j^{(1)}(u_0)(u - u_0) \equiv a_j + b_j(u - u_0), \tag{2.20}$$

for $j = 1, 2, ..., p$ and for u in a neighborhood of u_0. This leads to a local least-squares function, which can be used to find the local least-squares estimators for coefficient function $\beta_j(\cdot)$ at $u = u_0$ for $j = 1$, 2, ..., p:

$$\sum_{i=1}^{n} [Y_i - \mathbf{X}^T(\mathbf{a} + \mathbf{b}(u_i - u_0))]^2 \cdot K_h(u_i - u_0). \tag{2.21}$$

Denote $\mathbf{a} = (a_1{,}a_2{,}a_3{,}...{,}a_p)^T$, and $\mathbf{b} = (b_1{,}b_2{,}b_3{,}...{,}b_p)^T$ For each given u_0, we find the local least-squares estimate

$$\begin{aligned}
(\widehat{\mathbf{a}}, \widehat{\mathbf{b}})^T &\equiv (\widehat{a}_1, \widehat{a}_2, \widehat{a}_3, ..., \widehat{a}_p, \widehat{b}_1, \widehat{b}_2, \widehat{b}_3, ..., \widehat{b}_p)^T \\
&\equiv (\widehat{\beta}_1(u_0), \widehat{\beta}_2(u_0), ..., \widehat{\beta}_p(u_0), \widehat{\beta}_1^{(1)}(u_0), \widehat{\beta}_2^{(1)}(u_0), ..., \widehat{\beta}_p^{(1)}(u_0))^T
\end{aligned}$$

Then the component

$$\widehat{\mathbf{a}}^T \equiv (\widehat{a}_1, \widehat{a}_2, ..., \widehat{a}_p)^T \equiv (\widehat{\beta}_1(u_0), \widehat{\beta}_2(u_0), ..., \widehat{\beta}_p(u_0))^T.$$

is the estimate for the coefficient function $\beta(u_0) \equiv (\beta_1(u_0),\ \beta_2(u_0),\ \beta_2(u_0),\ldots,\ \beta_p(u_0))^T$ at the given covariate point u_0.

In this dissertation, the above procedure is referred to as a one-step estimation procedure, which is simple and useful. See Cleveland, Grosse and Shyu (1991). However, it implicitly assumes that the coefficient functions $\beta_j(u), j = 1,\ ,\ldots,\ p$ possess approximately the same degree of smoothness. This assumption allows the coefficient functions to be estimated equally well in the same interval of covariate u.

In real-world applications, the same degree of smoothness assumption is hardly satisfied, and different coefficient functions usually possess different degrees of smoothness. A two-step estimation procedure is therefore proposed to account for difference in smoothness of different coefficient functions. See Fan and Zhang (1999).

In two-step estimation procedure, without loss of generality, assume that coefficient function $\beta_p(u)$ is smoother than the rest of the coefficient functions, and assume that $\beta_p(u)$ possesses a bounded fourth order derivative so that we can use a cubic function to approximate $\beta_p(u)$

$$\begin{aligned} \beta_p(u) &\approx \beta_p(u_0) + \beta_p^{(1)}(u_0)(u-u_0) + \beta_p^{(2)}(u_0)(u-u_0)^2 + \beta_p^{(3)}(u_0)(u-u_0)^3 \\ &\equiv a_p + b_p(u-u_0) + c_p(u-u_0)^2 + d_p(u-u_0)^3 \end{aligned} \tag{2.22}$$

for u in a neighborhood of u_0. This updates the local least-squares function into:

$$\sum_{i=1}^{n}[Y_i - \sum_{j=1}^{p-1}\{a_j + b_j(u_i-u_0)\}\cdot X_{ij} - \{a_p + b_p(u-u_0) + c_p(u-u_0)^2 + d_p(u-u_0)^3\}X_{i,p}]^2 \cdot K_{h_1}(u_i-u_0)$$

If we denote $\widehat{a}_j,\ \widehat{b}_j,\ j = 1,\ 2,\ 3,\ldots,\ p-1$ and $\widehat{a}_p,\ \widehat{b}_p,\ \widehat{c}_p,\ \widehat{d}_p$ the minimizer of the above local least-squares function, then the resulting estimator of $\widehat{\beta}_{p,OS}(u_0) \equiv \widehat{a}_p$ is called a one-step estimator. The one-step estimator $\widehat{\beta}_{p,OS}(u_0)$ has bias $O(h_1^2)$ and variance of $O((nh_1)^{-1})$. This shows that the optimal variance of $O(n^{-8/9})$ is not achieved by one-step estimator.

In a two-step procedure, we in the first step find initial estimates of coefficient functions $\beta_j(u), j = 1, 2, ..., p$ by minimizing the local least-squares function

$$\sum_{i=1}^{n}[Y_i - \sum_{j=1}^{p}\{a_j + b_j(u_i - u_0)\} \cdot X_{ji}]^2 \cdot K_{h_0}(u_i - u_0) \tag{2.23}$$

for a given initial bandwidth h_0 and kernel K . The estimates through the first step are denoted by $\widehat{\beta}_1(u_0)$, $\widehat{\beta}_2(u_0)$,..., $\widehat{\beta}_p(u_0)$.

In the second step, we substitute $\widehat{\beta}_1(\cdot)$, $\widehat{\beta}_2(\cdot)$,..., $\widehat{\beta}_{p-1}(\cdot)$ and minimize

$$\sum_{i=1}^{n}[Y_i - \sum_{j=1}^{p-1}\widehat{a}_j(u_0) \cdot X_{ij} - \{a_p + b_p(u - u_0) + c_p(u - u_0)^2 + d_p(u - u_0)^3\}X_{i,p}]^2 \cdot K_{h_2}(u_i - u_0)$$

with respect to a_p, b_p, c_p, d_p. In the second step, bandwidth h_2 is used. We can consider the second step is to find the local least-squares cubic fit of $\widehat{\beta}_p(\cdot)$, which leads to the two-step estimator of $\widehat{\beta}_{p,TS}(u_0) \equiv \widehat{a}_p$ of $\beta_p(u_0)$.

Now we summarize the two-step estimation procedure and present the formulas. Define

$$\mathbf{Y} = [\ y_1 \quad y_2 \quad . \quad . \quad . \quad . \quad y_n\ ]^T,$$

and

$$\mathbf{X}_0 = \begin{pmatrix} X_{11} & (U_1 - u_0)X_{11} & X_{12} & (U_1 - u_0)X_{12} & ... & X_{1,p} & (U_1 - u_0)X_{1,p} \\ X_{21} & (U_2 - u_0)X_{21} & X_{22} & (U_2 - u_0)X_{22} & ... & X_{2,p} & (U_2 - u_0)X_{2,p} \\ . & . & . & . & ... & . & . \\ . & . & . & . & ... & . & . \\ Xn_1 & (U_n - u_0)X_{n1} & X_{n2} & (U_n - u_0)X_{n2} & ... & X_{n,p} & (U_n - u_0)X_{n,p} \end{pmatrix},$$

$$\mathbf{W}_0\text{=diag}\{K_{h_0}(u_1 - u_0),\ K_{h_0}(u_2 - u_0), ...,\ K_{h_0}(u_n - u_0)\}.$$

The solution to the local least-squares problem

$$\sum_{i=1}^{n}[Y_i - \sum_{j=0}^{d-1}\{a_j + b_j(u_i - u_0)\} \cdot X_{ji}]^2 \cdot K_{h_0}(u_i - u_0)$$

can be expressed as

$$\widehat{\beta}_{j,0}(u_0) = e^T_{2j-1,2(d-1)}(\mathbf{X}_0^T \cdot \mathbf{W_0} \cdot \mathbf{X}_0)^{-1}\mathbf{X}_0^T \cdot \mathbf{W}_0 \cdot \mathbf{Y} \text{ at } u = u_0,\ j = 0, 1, 2, ..., d-1$$

where $e_{k,m}$ is the notation for the unit vector of length m with 1 at the k^{th} position. The solution to the problem

$$\sum_{i=1}^{n}[Y_i - \sum_{j=0}^{d-2}\{a_j + b_j(u_i - u_0)\} \cdot X_{ij} - \{a_{d-1} + b_{d-1}(u - u_0) + c_{d-1}(u - u_0)^2 + d_{d-1}(u - u_0)^3\}X_{i,d-1}]^2 \cdot K_{h_1}(u_i - u_0)$$

can be expressed as following. Define

$$\mathbf{X}_2 = \begin{pmatrix} X_{1,p} & (U_1 - u_0)X_{1,p} & (U_1 - u_0)^2 X_{1,p} & (U_1 - u_0)^3 X_{1,p} \\ X_{2,p} & (U_2 - u_0)X_{2,p} & (U_2 - u_0)^2 X_{2,p} & (U_2 - u_0)^3 X_{2,p} \\ . & . & . & . \\ . & . & . & . \\ X_{n,p} & (U_n - u_0)X_{n,p} & (U_n - u_0)^2 X_{n,p} & (U_n - u_0)^3 X_{n,p} \end{pmatrix}$$

and

$$\mathbf{X}_3 = \begin{pmatrix} X_{11} & (U_1 - u_0)X_{11} & X_{12} & (U_1 - u_0)X_{12} & ... & X_{1,p-1} & (U_1 - u_0)X_{1,p-1} \\ X_{21} & (U_2 - u_0)X_{21} & X_{22} & (U_2 - u_0)X_{22} & ... & X_{2,p-1} & (U_2 - u_0)X_{2,p-1} \\ . & . & . & . & ... & . & . \\ . & . & . & . & ... & . & . \\ Xn_1 & (U_n - u_0)X_{n1} & X_{n2} & (U_n - u_0)X_{n2} & ... & X_{n,p-1} & (U_n - u_0)X_{n,p-1} \end{pmatrix}$$

$$\mathbf{X}_1 = (\mathbf{X}_2, \mathbf{X}_3), \quad \text{and} \quad \mathbf{W}_1 = \text{diag}\{K_{h_1}(u_1 - u_0),\ K_{h_1}(u_2 - u_0), ...,\ K_{h_1}(u_n - u_0)\}.$$

Then the solution to the least-squares problem is

$$\widehat{\beta}_{d-1,1}(u_0) = e^T_{2d-2,2d}(\mathbf{X}_1^T \cdot \mathbf{W}_1 \cdot \mathbf{X}_1)^{-1}\mathbf{X}_1^T \cdot \mathbf{W}_1 \cdot \mathbf{Y}$$

at $u = u_0$

Based on the above notation, the two-step estimator can be expressed as

$$\widehat{\beta}_{d-1,2}(u_0) = (1,0,0,0)(\mathbf{X}_2^T \cdot \mathbf{W}_2 \cdot \mathbf{X}_2)^{-1} \mathbf{X}_2^T \cdot \mathbf{W}_2 \cdot (\mathbf{Y} - \mathbf{V})$$

at $u = u_0$, where $\mathbf{W}_2$=diag$\{K_{h_2}(u_1 - u_0),\ K_{h_2}(u_2 - u_0),\ K_{h_2}(u_3 - u_0), ...,\ K_{h_2}(u_n - u_0)\}$ and $\mathbf{V} = (V_1, V_2, ..., V_n)^T$ with $V_i = \sum_{j=1}^{d-2} \widehat{\beta}_{j,0}(U_i)\mathbf{X}_{ij}$

Note that the two-step estimator $\widehat{\beta}_{d-1,2}(\cdot)$ is a linear estimator for the given bandwidth h_0 and h_2, since it is still a weighted sum of Y_1,Y_2,...,Y_n.

Local Likelihood Estimation Procedure For generalized varying-coefficient models, we estimate the coefficient functions $\beta_j(\cdot)$, $j = 1$, 2, ..., p through local likelihood approach. If the likelihood function is not available, quasi-likelihood function can be used.

A generalized varying-coefficient model has the form

$$\eta(u, \mathbf{x}) = g\{m(u, \mathbf{x})\} = \mathbf{x}^T \beta(\mathbf{u}) \tag{2.24}$$

where $g(\cdot)$ is a link function, $\mathbf{x} = (x_0,\ x_1,\ x_2,...,\ x_{d-1})^T$ is the vector of independent variables, and $m(\mathbf{u}, \mathbf{x})$ is the mean regression function of the response variable Y given that the co-variates $\mathbf{U} = \mathbf{u}$ and $\mathbf{X} = \mathbf{x}$. The link function $g(\cdot)$ transforms the mean regression function $m(\mathbf{u}, \mathbf{x})$ into a linear predictor.

A local likelihood for generalized varying-coefficient model has the form

$$\ell_n(\mathbf{a}, \mathbf{b}) = \frac{1}{n} \sum_{i=1}^{n} \ell[g^{-1}\{\mathbf{X}_i^T(\mathbf{a} + \mathbf{b}(u_i - u_0)), Y_i]^2 \cdot K_h(U_i - u_0). \tag{2.25}$$

Denote $\mathbf{a} = (a_1$,a_2,...,$a_p)^T$, and $\mathbf{b} = (b_1$,b_2,...,$b_p)^T$. We can write the local likelihood function $l_n(\mathbf{a}, \mathbf{b})$ using matrix notation

$$\ell_n(\mathbf{a}, \mathbf{b}) = \frac{1}{n} \sum_{i=1}^{n} l[g^{-1}\{\mathbf{X}^* \beta^*\}, Y_i]^2 \cdot K_h(U_i - u_0) \tag{2.26}$$

where

$$\mathbf{X}^* = \begin{pmatrix} X_{11} & X_{12} & ... & X_{1,p} & (U_1-u_0)X_{11} & (U_1-u_0)X_{12} & ... & (U_1-u_0)X_{1,p} \\ X_{21} & X_{22} & ... & X_{2,p} & (U_2-u_0)X_{21} & (U_2-u_0)X_{22} & ... & (U_2-u_0)X_{2,p} \\ . & . & ... & .. & . & ... & . & \\ . & . & ... & .. & . & ... & . & \\ Xn_1 & X_{n2} & ... & X_{n,p} & (U_n-u_0)X_{n1} & (U_n-u_0)X_{n2} & ... & (U_n-u_0)X_{n,p} \end{pmatrix}$$

and

$$\begin{aligned} \beta^* &= (\mathbf{a},\mathbf{b})^T \\ &\equiv (\begin{array}{cccccccccc} a_1 & a_2 & a_3 & ... & a_p & b_1 & b_2 & b_3 & ... & b_p \end{array})^T \\ &\equiv (\beta_1(u_0),\ \beta_2(u_0), ...,\ \beta_p(u_0),\ \beta_1^{(1)}(u_0),\ \beta_2^{(1)}(u_0), ...,\ \beta_p^{(1)}(u_0))^T \end{aligned}$$

For each given u_0, we find the local likelihood estimator

$$(\widehat{\mathbf{a}},\widehat{\mathbf{b}})^T \equiv (\begin{array}{cccccccccc} \widehat{a}_1 & \widehat{a}_2 & \widehat{a}_3 & ... & \widehat{a}_p & \widehat{b}_1 & \widehat{b}_2 & \widehat{b}_3 & ... & \widehat{b}_p \end{array})^T$$

$$\equiv (\widehat{\beta}_1(u_0),\ \widehat{\beta}_2(u_0), ...,\ \widehat{\beta}_p(u_0),\ \widehat{\beta}_1^{(1)}(u_0),\ \widehat{\beta}_2^{(1)}(u_0), ...,\ \widehat{\beta}_p^{(1)}(u_0))^T$$

Then the component $\widehat{\mathbf{a}}^T \equiv (\begin{array}{cccc} \widehat{a}_1 & \widehat{a}_2 & ... & \widehat{a}_p \end{array})^T \equiv (\widehat{\beta}_1(u_0), \widehat{\beta}_2(u_0), ..., \widehat{\beta}_p(u_0))^T$ is the estimate for u_0 $\beta(u_0) \equiv (\beta_1(u_0),\ \beta_2(u_0),...,\ \beta_p(u_0))$, the coefficient function at the given covariate point.

Let $\ell_n^{'}(\beta^*)$ and $\ell_n^{''}(\beta^*)$ denote the gradient and the Hessian matrix of the local log-likelihood $l_n(\beta^*)$. Now given an initial estimator $\widehat{\beta}_0^*(u_0) = (\widehat{\mathbf{a}}_0, \widehat{\mathbf{b}}_0)^T$, we can use iterative algorithm to find the local maximum likelihood estimate (MLE). In practice, we usually use the ordinary least-squares estimate for $\mathbf{a}_0$ as the initial value.

The iterative algorithm involved in estimating coefficient functions of generalized varying-coefficient models rises the problem of high computational expenses. This is because in order to find the estimates for coefficient functions, we usually need to find the functional value $\beta(u_0)$ at hundreds of distinct points; that is, we need to solve hundreds of local maximum likelihood problems. When cross-validation criterion is used to select a bandwidth h, the computational burden becomes even more severe. Note that this computational problem does not arise in varying-coefficient, where local least-squares method is used to find the estimate.

An efficient estimation of varying-coefficient models is proposed using one-step local MLE in Cai, Fan and Li (2000). Given an initial estimator $\widehat{\beta}_0^*(u_0) = (\widehat{\mathbf{a}}_0, \widehat{\mathbf{b}}_0)^T$, one-step of Newton-Raphson algorithm produces the one-step estimator

$$\widehat{\beta}_{OS}^* = \widehat{\beta}_0^* - \{\ell_n^{''}(\widehat{\beta}_0^*)\}^{-1}\ell_n^{'}(\widehat{\beta}_0^*)$$

It can be easily seen that the one-step estimator features the same computational expediency of the local least-squares local polynomial fitting.

Note that $\ell_n^{''}(\widehat{\beta}_0^*)$ can be a nearly singular matrix for some u_0, due to the sparcity of data in some region r when the bandwidth is chosen to be too small. Fan and Chen (1999) and Seifert and Gasser (1996) studied how to deal with such difficulties using ridge regression in a univariate setting.

Bandwidth Selection for Varying-coefficient Models

For nonparametric regression techniques, local polynomial approach requires to use bandwidth h and the choice of bandwidth h is crucial in analysis. In the literature, there are many proposals for selecting bandwidth; these include the cross-validation approach (Bowman 1984; Scott and Terrell 1987; Vieu 1991; Hall and Johnstone 1992; Fan 1996a) and the plug-in approach (Woodroofe 1970; Sheather and Jones 1991; Jones 1996).

In local polynomial fitting of varying-coefficient models, the choice of bandwidth h is crucial in the estimation. A multi-fold cross-validation method is used to select a bandwidth h.

We first partition the data into Q groups, with the j^{th} group consisting datum points with indices

$$d_j = \{Qk + j, k = 1, 2, 3, ...\}, \quad j = 0, 1, 2, ..., Q - 1$$

We fit the model and obtain the j^{th} estimate, using the remaining data after deleting the j^{th} group. Now, we denote by $\widehat{y}_{-d_j}(u_i, \mathbf{x}_i)$ the fitted value using the data with j^{th} group deleted.

For varying-coefficient models, the cross-validation criterion has the form

$$CV(h) = \sum_{j=0}^{Q-1} \sum_{i \in d_j} [y_i - \widehat{y}_{-d_j}(u_i, \mathbf{x}_i)]^2 \tag{2.27}$$

We choose the bandwidth h that minimizes the $CV(h)$; that is, we select the bandwidth h that provide the model that best fits of the data.

For generalized varying-coefficient models, the goodness-of-fit criterion is determined by deviance or the sum of squares of Pearson's residuals. Then, it leads to two cross-validation criteria

$$CV_1(h) = \sum_{j=0}^{Q-1} \sum_{i \in d_j} [y_i \log\{\frac{y_i}{\widehat{y}_{-d_j}(u_i, \mathbf{x}_i)}\} - \{y_i - \widehat{y}_{-d_j}(u_i, \mathbf{x}_i)\}] \tag{2.28}$$

which is based on sum of deviance residuals. Here $0 \log(0) = 0$. and

$$CV_2(h) = \sum_{j=0}^{Q-1} \sum_{i \in d_j} \{\frac{y_i - \widehat{y}_{-d_j}(u_i, \mathbf{x}_i)}{\sqrt{\widehat{y}_{-d_j}(u_i, \mathbf{x}_i)}}\}^2 \tag{2.29}$$

where $\widehat{y}_{-d_j}(u_i, \mathbf{x}_i)$ is the fitted value obtained using data with j^{th} group deleted. In practice, Q is usually chosen to be 20. In general, the cross-validation is not very sensitive to way in which data is partitioned (Cai, Fan and Li, 2000).

In Zhang and Lee (2000), a variable bandwidth selection method is proposed based on local polynomial fitting of varying-coefficient models. It is discussed in the paper that a constant bandwidth is sufficient if the unknown coefficient functions are spatial homogeneous; while for estimating varying-coefficient models with more complex coefficient functions, variable bandwidth is needed.

Hastie and Tibshirani (1993) proposed the use of a smoothing spline to estimate the regression coefficient functions in varying-coefficient models. The smoothing spline approach is to minimize the following penalized least squares function:

$$\sum_{i=1}^{n}[Y_i - \sum_{j=1}^{d}\beta_j(u_0)\cdot X_{ij}]^2 + \sum_{j=1}^{d}\lambda_j\int\{\beta_j^{''}(u)\}^2du,$$

where λ_j, $j = 1$, 1, 2,..., d are smoothing parameters. The estimation of coefficient functions $\beta_j(u_0), j = 1$, 1, 2,..., d through the smoothing spline approach may have some potential problems. The first problem is that the choice of the p smoothing parameters is a difficult task in practice. Second, the computational burden is significant and challenging, as is the iterative procedure proposed in Hastie and Tibshirani (1993). Third, it is difficult to derive the sampling properties.

2.3 Nonparametric Goodness-of-fit Tests

After the estimation procedure for varying-coefficient models, we naturally will be interested to know whether the estimated unknown coefficient functions really depend on covariate u, or whether any of the coefficient functions are significant in the fitted model. We therefore need to consider the following hypothesis testing problems:

$$H_0 : \beta_j(u) \equiv \beta_{j0} \quad \text{versus} \quad H_1 : \beta_j(u) \neq \beta_{j0} \text{ for some } u,\ j = 1, 2, ..., p \tag{2.30}$$

or

$$H_0 : \beta_j(u) \equiv 0 \quad \text{versus} \quad H_1 : \beta_j(u) \neq 0 \text{ for some } u,\ j = 1, 2, ..., p \tag{2.31}$$

where β_{j0} is an unknown constant.

The statistical test can be conducted through a nonparametric likelihood ratio test. Denote

$\ell(H_0)$: the log-likelihood under H_0

$\ell(H_1)$: the log-likelihood under H_1

Define

$$T = \{\ell(H_1) - \ell(H_0)\}.$$

Intuitively, we expect T to be small under H_0; while we expect T to be large under H_1. For generalized linear model (GLIM), the likelihood ratio test statistic follows asymptotically a chi-square distribution with degree of freedom $f - r$, where f and r are dimensions of parameter space under H_1 and H_0, respectively. However, for the nonparametric alternative hypothesis, note that the parameter space is infinite dimensional, although the parameter space under null hypothesis is finite dimensional. Thus, many traditional tests, such as likelihood the ratio test, cannot be applied to the type of hypothesis testing proposed above.

Intuitively, under H_0, there will be little difference between $\ell(H_0)$ and $\ell(H_1)$. However, under the alternative hypothesis, $\ell(H_1)$ should become systematically larger than $\ell(H_0)$, and hence the Generalized Likelihood Ratio(GLR) test statistic $GLRT_0$ will tend to take a large positive value. Hence, a large value of the Generalized Likelihood Ratio(GLR) test statistic $GLRT_0$ indicates that the null hypothesis should be rejected. Cai, Fan and Li (2000) empirically demonstrated that $r_K T$ has an asymptotic chi-square distribution, where r_K is a constant depending on the kernel function. Fan, Zhang and Zhang (2001) provides a general framework for generalized likelihood ratio tests.

Chapter 3

Statistical Inference Procedures for Nonlinear Varying Coefficient Models

Suppose that $\{u_i, \mathbf{x}_i, y_i\}, i = 1, \cdots, n$ is a random sample from the nonlinear varying coefficient model

$$y = f(\mathbf{x}, \boldsymbol{\beta}(u)) + \varepsilon, \tag{3.1}$$

where $\boldsymbol{\beta}(\cdot)$ is the unknown regression coefficient function, ε is a random error with zero mean and finite variance. Denote the dimensions of $\mathbf{x}$ and $\boldsymbol{\beta}(\cdot)$ by p and d, respectively.

3.1 Estimation Procedure

We first propose an estimation procedure for the nonlinear varying coefficient model by using local linear regression techniques. We consider only the case in which u is univariate. Extension to multivariate does not involve extra difficulty, but it may practically be useless because of the curse of dimensionality.

Using Taylor's expansion in a neighborhood of given u_0, for $j = 1, \cdots, d$,

$$\beta_j(u) \approx \beta_j(u_0) + \beta_j(u_0)(u - u_0) \equiv a_j + b_j(u - u_0).$$

Denote $\mathbf{a} = (a_1, \cdots, a_d)^T$ and $\mathbf{b} = (b_1, \cdots, b_d)^T$. Thus, we obtain a local linear regression estimator $(\widehat{\mathbf{a}}^T, \widehat{\mathbf{b}}^T)^T$ by minimizing

$$\ell(\mathbf{a}, \mathbf{b}) = \frac{1}{2}\sum_{i=1}^{n}[y_i - f\{\mathbf{x}_i, \mathbf{a} + \mathbf{b}(u_i - u_0)\}]^2 K_h(u_i - u_0) \tag{3.2}$$

where $K_h(t) = h^{-1}K(t/h)$ and $K(\cdot)$ is a kernel density function. That is.

$$\widehat{\boldsymbol{\beta}}(u_0) = \widehat{\mathbf{a}}$$

We next derive the asymptotic distributions of the local linear estimate $\widehat{\mathbf{a}}$ and $\widehat{\mathbf{b}}$. For simplicity of presentation, denote $\boldsymbol{\theta}(u_0) = (a_1, \cdots, a_p, b_1, \cdots, b_p)^T$, and $\widehat{\boldsymbol{\theta}}(u_0) = (\widehat{\mathbf{a}}(u_0)^T, \widehat{\mathbf{b}}(u_0)^T)^T$. Let $c(u)$ denote the marginal density of U. Define $\mu_k = \int t^k K(t)\, dt$ and $\nu_k = \int t^k K^2(t)\, dt$ and $\mathbf{H} = diag\{1, h\} \otimes I_p$. Further denote

$$\Gamma_1(u_0) = E\{f'\{\mathbf{x}; \boldsymbol{\beta}(u_0)\}[f'\{\mathbf{x}; \boldsymbol{\beta}(u_0)\}]^T | U = u_0\}, \tag{3.3}$$

and

$$\Gamma_2(u_0) = E\{\sigma^2(u_0, \mathbf{x}) f'\{\mathbf{x}; \boldsymbol{\beta}(u_0)\}[f'\{\mathbf{x}; \boldsymbol{\beta}(u_0)\}]^T | U = u_0\}, \tag{3.4}$$

where $f'(\mathbf{x}; \boldsymbol{\beta}) = \partial f(\mathbf{x}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}$.

Theorem 1. *Assume that Conditions A — G in Section 3.4 hold. We have the following asymptotic normality for* $\widehat{\boldsymbol{\theta}}(u_0)$.

$$\begin{aligned}&\sqrt{n\,h}\left[\mathbf{H}\left\{\widehat{\boldsymbol{\theta}}(u_0) - \boldsymbol{\theta}(u_0)\right\} - \frac{h^2}{2(\mu_2 - \mu_1^2)}\begin{pmatrix}(\mu_2^2 - \mu_1\,\mu_3)\,\boldsymbol{\beta}''(u_0) \\ (\mu_3 - \mu_1\,\mu_2)\,\boldsymbol{\beta}''(u_0)\end{pmatrix} + o_p(h^2)\right] \\ &\xrightarrow{\mathcal{D}} \quad N\left(0,\ \Delta^{-1}\,\Lambda\,\Delta^{-1}\right),\end{aligned} \tag{3.5}$$

where

$$\Delta = c(u_0)\begin{pmatrix}1 & \mu_1 \\ \mu_1 & \mu_2\end{pmatrix} \otimes \Gamma_1(u_0) \quad \text{and} \quad \Lambda = c(u_0)\begin{pmatrix}\nu_0 & \nu_1 \\ \nu_1 & \nu_2\end{pmatrix} \otimes \Gamma_2(u_0). \tag{3.6}$$

with $\Gamma_1(u_0)$ and $\Gamma_2(u_0)$ given by (3.3) and (3.4), respectively. Furthermore, if $K(\cdot)$ is symmetric,

$$\sqrt{n\,h}\,\left[\widehat{\mathbf{a}}(u_0) - \boldsymbol{\beta}(u_0) - \frac{h^2\,\mu_2}{2}\,\boldsymbol{\beta}''(u_0) + o_p(h^2)\right] \quad \xrightarrow{\mathcal{D}} \quad N\left(0,\ \Sigma(u_0)\right), \tag{3.7}$$

where

$$\Sigma(u_0) = \nu_0\,\Gamma_1^{-1}(u_0)\Gamma_2(u_0)\Gamma_1^{-1}(u_0)/c(u_0). \tag{3.8}$$

Proof of this theorem is given in Section 3.4.

3.2 Computation Issue

We will address the computation issue associated with the minimization of nonlinear least squares function. For a given initial value $(\mathbf{a}_0^T, \mathbf{b}_0^T)^T$ of $(\mathbf{a}^T, \mathbf{b}^T)^T$, we locally and linearly approximate

$$f\{\mathbf{x}, \mathbf{a}+\mathbf{b}(u-u_0)\} \approx f\{\mathbf{x}, \mathbf{a}_0+\mathbf{b}_0(u-u_0)\}+\{(\mathbf{a}-\mathbf{a}_0)+(\mathbf{b}-\mathbf{b}_0)(u-u_0)\}^T f'\{\mathbf{x}, \mathbf{a}_0+\mathbf{b}_0(u-u_0)\},$$

where $f'(\mathbf{y}, \boldsymbol{\beta}) = \partial f(\mathbf{x}, \boldsymbol{\beta})/\partial\boldsymbol{\beta}$. With this approximation, we can use an iterated least squares algorithm to search the solution of (3.2). Specifically, at the k-th step during the course of iteration, the current value for $\mathbf{a}$ and $\mathbf{b}$ is $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$. Denote

$$y_{i,k} \;=\; y_i - f\{\mathbf{x}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i-u_0)\} + \{\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i-u_0)\}f'\{\mathbf{x}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i-u_0)\},$$

$$F_k \;=\; \begin{pmatrix} f'\{\mathbf{x}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_1-u_0)\}, & \cdots & f'\{\mathbf{x}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_n-u_0)\} \\ (u_1-u_0)f'\{\mathbf{x}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_1-u_0)\}, & \cdots & (u_n-u_0)f'\{\mathbf{x}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_n-u_0)\}^T \end{pmatrix}$$

and $\mathbf{y}_k = (y_{1,k}, \cdots, y_{n,k})^T$.

Then, we update $(\mathbf{a}, \mathbf{b})^T$ by

$$\begin{pmatrix} \mathbf{a}^{(k+1)} \\ \mathbf{b}^{(k+1)} \end{pmatrix} = (F_k^T W F_k)^{-1} F_k^T W \mathbf{y}_k,$$

where $W = \text{diag}\{K_h(u_1 - u_0), \cdots, K_h(u_n - u_0)\}$. When the algorithm converges, the solution is satisfied with

$$\ell'(\mathbf{a}, \mathbf{b}) = 0.$$

Denote the resulting estimate of $(\mathbf{a}, \mathbf{b})^T$ by $(\widehat{\mathbf{a}}, \widehat{\mathbf{b}})^T$. Then

$$\widehat{\boldsymbol{\beta}}(u_0) = \widehat{\mathbf{a}}, \quad \text{and} \quad \widehat{\boldsymbol{\beta}}'(u_0) = \widehat{\mathbf{b}}.$$

Following conventional techniques, the standard error of $(\widehat{\mathbf{a}}, \widehat{\mathbf{b}})^T$ using sandwich formula. In other words, conditioning on $(u_1, \mathbf{x}_1), \cdots, (u_n, \mathbf{x}_n)$,

$$\widehat{\text{cov}}\{\begin{pmatrix} \widehat{\mathbf{a}} \\ \widehat{\mathbf{b}} \end{pmatrix}\} = (F^T W F)^{-1} F^T W \Sigma W F (F^T W F)^{-1},$$

where $F = F_k$ with $\mathbf{a}^{(k)} = \widehat{\mathbf{a}}$ and $\mathbf{b}^{(k)} = \widehat{\mathbf{b}}$, and $\Sigma = \text{diag}\{e_1^2, \cdots, e_n^2\}$, where $e_i = y_i - f\{\mathbf{x}_i, \widehat{\boldsymbol{\beta}}(z_i)\}$.

We will test the accuracy of the proposed standard error formula in the simulation study in next chapter.

3.3 Generalized likelihood ratio tests

Next we deal with nonparametric testing hypothesis problems. Consider the following hypothesis testing problem:

$$H_0 : \beta_j(u) = \beta_{j0}(u, \boldsymbol{\gamma}_j), j = 1 \cdots, p \quad \text{vs} \quad H_1 : \beta_j(u) \neq \beta_{j0}(u, \boldsymbol{\gamma}_j), j = 1 \cdots, p, \tag{3.9}$$

where $\beta_{j0}(u, \boldsymbol{\gamma}_j)$ has a parametric form in which we are interested, and $\boldsymbol{\gamma}_j$ is a vector of unknown parameters. For example, taking $\beta_{j0}(u, \boldsymbol{\gamma}_j) = \gamma_{j0}$, where γ_{j0} is unknown constant, the null hypothesis implies that the $\beta_j(u)$ is a constant.

To gain more insights into the construction of nonparametric likelihood ratio type of tests, assume, tentatively, that the random error ε is $N(0, \sigma^2)$. Then the likelihood function of the data $\{u_i, \mathbf{x}_i, y_i\}, i = 1, \cdots, n$, is proportional to

$$(\sigma^2)^{-n/2} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} \left[y_i - g(\mathbf{x}_i; \boldsymbol{\beta}(u_i))\right]^2 \right\}.$$

Let $\tilde{\boldsymbol{\beta}}(\cdot)$ and $\widehat{\boldsymbol{\beta}}(\cdot)$ be estimates of $\boldsymbol{\beta}(t)$ under H_0 and H_1, respectively. A generalized likelihood ratio (GLR) test statistic is defined as

$$\frac{n}{2} \log(\mathrm{RSS}(H_0)/\mathrm{RSS}(H_1))$$

where

$$\mathrm{RSS}(H_0) = \sum_{i=1}^{n} \{y_i - g(\mathbf{x}_i; \tilde{\boldsymbol{\beta}}(u_i))\}^2$$

and

$$\mathrm{RSS}(H_1) = \sum_{i=1}^{n} \{y_i - g(\mathbf{x}_i; \widehat{\boldsymbol{\beta}}(u_i))\}^2$$

Under H_0, the GLR test statistic is asymptotically equivalent to

$$T_0 = \frac{n}{2} \frac{\mathrm{RSS}(H_0) - \mathrm{RSS}(H_1)}{\mathrm{RSS}(H_1)}.$$

Note that $\widehat{\boldsymbol{\beta}}(\cdot)$ is obtained by using a nonparametric estimator, rather than the maximum likelihood estimator under the alternative model. In fact, as argued in Fan, Zhang and Zhang (2001), the nonparametric MLE usually does not exist, and, even if it does exists, the resulting nonparametric maximum likelihood ratio test is not very powerful. This is the motivation behind the generalized likelihood ratio test.

Intuitively, under H_0, there will be little difference between $\mathrm{RSS}(H_0)$ and $\mathrm{RSS}(H_1)$. However, under the alternative hypothesis, $\mathrm{RSS}(H_0)$ should become systematically larger than $\mathrm{RSS}(H_1)$, and hence the test statistic T_0 will tend to take a large positive value. Hence, a large value of the test statistic T_0 indicates that the null hypothesis should be rejected.

In the nonparametric regression model and varying coefficient models, Fan, Zhang and Zhang (2001) unveiled the following Wilks phenomenon: The asymptotic null distribution of

$$T \equiv r_K T_0 \tag{3.10}$$

is a chi-square distribution, where where $r_K = \{K(0) - 0.5 \int K^2(u)\, du\}/[\int\{K(u) - 0.5K * K(u)\}\, du]$ and $K(\cdot)$ is the kernel function used to estimate the regression coefficients.

As pointed out by Fan, Zhang and Zhang (2001) that, unlike the parametric settings in which, in general, the likelihood ratio test has a χ^2-distribution, it needs to verify that the GLR test has a chi-square limiting distribution for every specific nonparametric model. Here we would conjecture the Wilks type of results to continue to hold for our current setting. We will provide empirical justifications by Monte Carlo simulation. Similar to the proposal of Cai, Fan and Li (2000), the null distribution of T can be estimated by using a bootstrap procedure. This usually provides a better estimate than the asymptotic null distribution, since, in the nonparametric situation, the degrees of freedom tends to infinite and the results in Fan, Zhang and Zhang (2001) give only the main order of the degrees of freedom.

3.4 Proof of Theorem 1

We need the following regularity conditions for the proof of Theorem 1. These conditions are not the weakest conditions, but they are imposed to facilitate the proofs.

Regularity Conditions

A. The sample $\{u_i, \mathbf{x}_i, y_i\}$, $i = 1, \cdots, n$ is independent and identically distributed to the population $(u, \mathbf{x}, y)$. The ε_i are independent and identically distributed with mean zero and finite variance σ^2. The covariate u has a finite support $\mathcal{U} = [L, U]$, where both L and U are finite. Furthermore, The support for $\mathbf{x}$, denoted by $\mathcal{X}$, is closed and bounded of R^p.

B. The coefficient functions $\beta_j(u), j = 1, \cdots, p$ has continuous second order derivative over $\mathcal{U}$.

C. Assume that $f(\mathbf{x}; \boldsymbol{\beta})$ is a continuous function on $\mathcal{X} \times Theta_1$, where $\Theta_1 = \{\beta(u) : L \leq u \leq U\}$. The first and second partial derivative of $f(\mathbf{x}; \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ exist and continuous. Furthermore, $n^{-1} \sum_{i=1}^n f'(\mathbf{x}_i, \boldsymbol{\beta}) f'(\mathbf{x}_i, \boldsymbol{\beta})^T$ converges to some matrix $\Omega(\boldsymbol{\beta})$ uniformly in $\boldsymbol{\beta}$ for $\boldsymbol{\beta} \in \Theta_1$. Also assume that $\Omega(\boldsymbol{\beta}(u_0))$ is finite positive definite.

D. $n^{-1} \sum_{i=1}^n [\partial^2 f(\mathbf{x}_i, \boldsymbol{\beta}) / \partial\beta_r \partial\beta_s]^2$ converges uniformly in $\boldsymbol{\beta}$ for $\boldsymbol{\beta} \in \Theta_1$ $(r, s = 1, \cdots, p)$.

E. The kernel density function $K(\cdot)$ has finite support and satisfies

$$\int K(t) dt = 1, \int |t|^3 K(t)\, dt < \infty, \int t^2 K^2(t) dt < \infty.$$

Without loss of generality, it is assumed that the support $K(\cdot)$ is $[-1, 1]$.

F. The marginal density function $c(u)$ is continuous and positive for $u \in \mathcal{U}$.

G. $h = O(n^{-1/5})$.

Conditions A, B, C, D and G are adopted from the regularity conditions for nonlinear least squares estimator. See, for example, Chapter 12 of Seber and Wild (1989).

Condition B implies that the parameter space for $\boldsymbol{\theta}$, namely, Θ_1 is a closed and bounded (i.e., compact) subset of R^{2d}.

Let us start with the consistency of $\widehat{\boldsymbol{\theta}}$. Denote $\mathbf{a}_0 = \boldsymbol{\beta}(u_0)$ and $\mathbf{b}_0 = \boldsymbol{\beta}'(u_0)$. That is, $(\mathbf{a}_0^T, \mathbf{b}_0^T)$ is the true value of $(\mathbf{a}^T, \mathbf{b}^T)$.

Lemma 1. *Under the regularity conditions A– G, with probability tending to one, there exists a local minimizer* $\{\widehat{\mathbf{a}}, \widehat{\mathbf{b}}\}$ *of* $\ell(\mathbf{a}, \mathbf{b})$ *such that* $\|\widehat{\mathbf{a}} - \mathbf{a}_0\| = O_P(1/\sqrt{nh})$ *and* $\|\widehat{\mathbf{b}} - \mathbf{b}_0\| = O_P(1/(h\sqrt{nh})$*.*

Proof. We want to show that for any given $\eta > 0$, there exists two large constant C_1 and C_2 such that

$$P\{\inf_{\|\mathbf{w}_1\|=C_1, \|\mathbf{w}_2\|=C_2} \ell\{\mathbf{a}_0 + \mathbf{w}_1/\sqrt{nh}, \mathbf{b}_0 + \mathbf{w}_2/(h\sqrt{nh})\} > \ell(\mathbf{a}_0, \mathbf{b}_0)\} \geq 1 - \eta.$$

This implies that with probability at least $1 - \eta$ that there exists a local minimum in the region $\{\mathbf{a}_0 + \mathbf{w}_1/\sqrt{nh} : \|\mathbf{w}_1\| \leq C_1\} \times \{\mathbf{b}_0 + \mathbf{w}_2/h\sqrt{nh} : \|\mathbf{w}_2\| \leq C_2\}$. Hence, there exists a local minimizer such that $\|\widehat{\mathbf{a}} - \mathbf{a}_0\| = O_P(1/\sqrt{nh})$ and $\|\widehat{\mathbf{b}} - \mathbf{b}_0\| = O_P(1/(h\sqrt{nh})$.

Denote

$$D_n(\mathbf{w}_1, \mathbf{w}_2) = \frac{1}{n}\ell(\mathbf{a}_0 + \mathbf{w}_1/\sqrt{nh}, \mathbf{b}_0 + \mathbf{w}_2/h\sqrt{nh}) - \frac{1}{n}\ell(\mathbf{a}_0, \mathbf{b}_0)$$

Decompose $D_n(\mathbf{w}_1, \mathbf{w}_2)$ as

$$D_n(\mathbf{w}_1, \mathbf{w}_2) = I_1 + I_2,$$

where

$$\begin{aligned}
I_1 &= \frac{1}{n}\sum_{i=1}^{n}[y_i - f\{\mathbf{x}_i, \mathbf{a}_0 + \mathbf{b}_0(u_i - u_0)\}] \\
&\quad \times [f\{\mathbf{x}_i, \mathbf{a}_0 + \mathbf{b}_0(u_i - u_0)\} - f\{\mathbf{x}_i, \mathbf{a}_0 + \mathbf{w}_1/\sqrt{nh} + (\mathbf{b}_0 + \mathbf{w}_2/h\sqrt{nh})(u_i - u_0)\}]K_h(u_i - u_0) \\
I_2 &= \frac{1}{2n}\sum_{i=1}^{n}[f\{\mathbf{x}_i, \mathbf{a}_0 + \mathbf{b}_0(u_i - u_0)\} - f\{\mathbf{x}_i, \mathbf{a}_0 + \mathbf{w}_1/\sqrt{nh} + (\mathbf{b}_0 + \mathbf{w}_2/h\sqrt{nh})(u_i - u_0)\}]^2 K_h(u_i - u_0)
\end{aligned}$$

Let us calculate the order of I_1 first. Since

$$\begin{aligned}
E(I_1) &= E[y - f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0(u_i - u_0)\}] \\
&\quad \times [f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0(u - u_0)\} - f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{w}_1/\sqrt{nh} + (\mathbf{b}_0 + \mathbf{w}_2/h\sqrt{nh})(u - u_0)\}]K_h(u - u_0) \\
&= E[f(\mathbf{x}, \boldsymbol{\beta}(u)) - f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0(u_i - u_0)\}] \\
&\quad \times [f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0(u - u_0)\} - f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{w}_1/\sqrt{nh} + (\mathbf{b}_0 + \mathbf{w}_2/h\sqrt{nh})(u - u_0)\}]K_h(u - u_0). \\
&= E\int [f(\mathbf{x}, \boldsymbol{\beta}(u_0 + hv)) - f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0 hv\}] \\
&\quad \times [f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0 hv\} - f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{w}_1/\sqrt{nh} + (\mathbf{b}_0 + \mathbf{w}_2 v/\sqrt{nh})\}]K(v)c(u_0 + hv)\, dv.
\end{aligned}$$

By some straightforward calculation, it follows that

$$f(\mathbf{x}, \boldsymbol{\beta}(u_0 + hv)) - f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0 hv\} = O_p(h^2),$$

and

$$\begin{aligned}
& f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0 hv\} - f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{w}_1/\sqrt{nh} + (\mathbf{b}_0 + \mathbf{w}_2 v/\sqrt{nh}) \\
= & -f'(\mathbf{x}, \mathbf{a}_0)^T(\mathbf{w}_1/\sqrt{nh} + \mathbf{w}_2 v/\sqrt{nh}) + O_P(1/(nh)). \qquad (3.11)
\end{aligned}$$

Thus, the order of leading term of I_1 is $O_P(h^2/\sqrt{nh}) = O_p(n^{-4/5})$ as $h = O(n^{-1/5})$ using Condition G. Now let us calculate the order of I_2.

$$E(I_2) = \frac{1}{2}E[f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0(u - u_0)\} - f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{w}_1/\sqrt{nh} + (\mathbf{b}_0 + \mathbf{w}_2/h\sqrt{nh})(u - u_0)\}]^2 K_h(u - u_0)$$

Using (3.11), it can be shown that the order of the leading term of I_2 is $O_P\{1/(nh)\} = O_P(n^{-4/5})$. Thus, the leading terms of I_1 and I_2 have the same order. Note that I_2 is a quadratic function of $(\mathbf{w}_1, \mathbf{w}_2)$, while I_1 is linear function of $(\mathbf{w}_1, \mathbf{w}_2)$. Thus, I_2 dominates I_1 by taking large enough C_1 and C_2 by Condition C. This completes the proof of Lemma 1.

Proof of Theorem 1. By a Taylor expansion and week consistency $\widehat{\boldsymbol{\theta}}(u_0) \xrightarrow{P} \boldsymbol{\theta}(u_0)$, we have

$$\begin{aligned}
0 &= \ell_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}(u_0)) = \ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0)) + \ell_{\boldsymbol{\theta}\boldsymbol{\theta}^T}(\boldsymbol{\theta}(u_0))\{\widehat{\boldsymbol{\theta}}(u_0) - \boldsymbol{\theta}(u_0)\} + O_P(||\widehat{\boldsymbol{\theta}}(u_0) - \boldsymbol{\theta}(u_0)||^2) \\
&= \ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0)) + \{\ell_{\boldsymbol{\theta}\boldsymbol{\theta}^T}(\boldsymbol{\theta}(u_0)) + o_P(1)\}\{\widehat{\boldsymbol{\theta}}(u_0) - \boldsymbol{\theta}(u_0)\}.
\end{aligned}$$

Then,

$$\widehat{\boldsymbol{\theta}}(u_0) - \boldsymbol{\theta}(u_0) \;=\; -\{\ell_{\boldsymbol{\theta}\boldsymbol{\theta}^T}(\boldsymbol{\theta}(u_0)) + o_P(1)\}^{-1}\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0))$$

Thus,

$$\mathbf{H}\left\{\widehat{\boldsymbol{\theta}}(u_0) - \boldsymbol{\theta}(u_0)\right\} = -\mathbf{H}\{\ell_{\boldsymbol{\theta}\boldsymbol{\theta}^T}(\boldsymbol{\theta}(u_0)) + o_P(1)\}^{-1}\mathbf{H}\mathbf{H}^{-1}\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0)) \tag{3.12}$$

For the expectation of $\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0))$, we do a Taylor expansion around the true function $\boldsymbol{\beta}(u)$ at u_0

$$\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0)) = \begin{pmatrix} \sum_{i=1}^n [y_i - f\{\mathbf{x}_i, \mathbf{a} + \mathbf{b}(u_i - u_0)\}] f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_0)\} K_h(u_i - u_0) \\ \\ \sum_{i=1}^n [y_i - f\{\mathbf{x}_i, \mathbf{a} + \mathbf{b}(u_i - u_0)\}] f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_0)\}(u_i - u_0) K_h(u_i - u_0) \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n [\varepsilon_i + f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_i)\}\frac{1}{2}\boldsymbol{\beta}''(u^*)(u_i - u_0)^2\}] f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_0)\} K_h(u_i - u_0) \\ \\ \sum_{i=1}^n [\varepsilon_i + f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_i)\}\frac{1}{2}\boldsymbol{\beta}''(u^*)(u_i - u_0)^2\}] f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_0)\}(u_i - u_0) K_h(u_i - u_0) \end{pmatrix}$$

where where u^* is between u_i and u_0. Then it can be showed that

$$\mathbf{H}^{-1}\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0)) \xrightarrow{P} c(u_0)\,\frac{1}{2}\boldsymbol{\beta}''(u_0)h^2 \begin{pmatrix} \mu_2 \\ \\ \mu_3 \end{pmatrix} \otimes \Gamma_1(u_0) + o_p(h^2) \tag{3.13}$$

For the variance of $\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0))$

$$E\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0))[\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0))]^T = \begin{pmatrix} E_{11} & E_{11} \\ \\ E_{21} & E_{22} \end{pmatrix} \tag{3.14}$$

where

$$\begin{aligned} E_{11} &= E\sum_{i=1}^{n}[\varepsilon_i + f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_i)\}\frac{1}{2}\boldsymbol{\beta}''(u^*)(u_i - u_0)^2\}]^2 \\ &\quad \times f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_0)\}[f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_0)\}]^T K_h^2(u_i - u_0) \end{aligned}$$

$$\begin{aligned} E_{12} = E_{21} &= E\sum_{i=1}^{n}[\varepsilon_i + f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_i)\}\frac{1}{2}\boldsymbol{\beta}''(u^*)(u_i - u_0)^2\}]^2 \\ &\quad \times f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_0)\}[f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_0)\}]^T (u_i - u_0) K_h^2(u_i - u_0) \end{aligned}$$

$$\begin{aligned} E_{22} &= E\sum_{i=1}^{n}[\varepsilon_i + f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_i)\}\frac{1}{2}\boldsymbol{\beta}''(u^*)(u_i - u_0)^2\}]^2 \\ &\quad \times f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_0)\}[f'\{\mathbf{x}_i, \boldsymbol{\beta}(u_0)\}]^T (u_i - u_0)^2 K_h^2(u_i - u_0) \end{aligned}$$

Then it can be showed that

$$E\mathbf{H}^{-1}\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0))\mathbf{H}^{-1}[\ell_{\boldsymbol{\theta}}(\boldsymbol{\theta}(u_0))]^T = c(u_0) \begin{pmatrix} \nu_0 & \nu_1 \\ \nu_1 & \nu_2 \end{pmatrix} \otimes \Gamma_2(u_0) + O_p(h^2) \tag{3.15}$$

Similarly, we can show

$$\mathbf{H}^{-1}\ell_{\boldsymbol{\theta}\boldsymbol{\theta}^T}(\boldsymbol{\theta}(u_0))\mathbf{H}^{-1} \xrightarrow{P} -c(u_0) \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix} \otimes \Gamma_1(u_0) \tag{3.16}$$

and note that

$$\begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix}^{-1} \times \begin{pmatrix} \mu_2 \\ \mu_3 \end{pmatrix} = \frac{1}{(\mu_2 - \mu_1^2)} \begin{pmatrix} \mu_2^2 - \mu_1\,\mu_3 \\ \mu_3 - \mu_1\,\mu_2 \end{pmatrix}$$

The asymptotic result is proved by combining (3.13), (3.14), (3.15), and (3.16).

If $K(u)$ is a symmetric kernel, then $\mu_1 = \mu_3 = 0$. (3.6) and (3.7) follows immediately.

Chapter 4

Numerical Studies

The purpose of this chapter is to study the finite sample performance of the proposed statistical procedures in Chapter 3. To this end, we consider three typical nonlinear regression models in this chapter. For each model, we conduct extensive simulation studies to assess the performance of the proposed estimation procedure, test the accuracy of proposed standard error formula, and evaluate the proposed procedure of testing hypothesis, including the investigation of the type I error rate of the proposed GLR test and calculating the power under some specific alternative hypothesis.

In our simulation, the kernel function is taken to be Epanechnikov kernel, i.e., $K(u) = .75(1-u^2)_+$. The grid points $\{u_k,\ k = 1, ..., n_{grid}\}$, at which $\widehat{\beta}_j(\cdot)'s$ are evaluated, are taken to be evenly distributed over the range of u with $n_{grid} = 200$. The performance of estimator $\widehat{\beta}(\cdot)$ is accessed by the square Root of Average Squared Error (RASE), defined by

$$RASE = \{n_{grid}^{-1} \sum_{j=1}^{p} \sum_{k=1}^{n_{grid}} \{\beta_j(u_k) - \widehat{\beta}_j(u_k)\}^2\}^{1/2} \tag{4.1}$$

For each simulation setting, we consider three bandwidths which represent a widely varying degree of smoothness. Specifically, one bandwidth stands for under-smooth, one for over-smooth and one for about right-smoothness. For each case, we conduct 400 simulations.

4.1 Example 1

In this section, we conduct simulation studies based on the following model:

$$Y = \exp(\mathbf{X}^T \beta(u)) + \varepsilon, \tag{4.2}$$

where $\mathbf{X} = \{1,\ X\}$ and $\beta(u) = \{\beta_1(u), \beta_2(u)\}$ with $\beta_1(u) = \sin(\pi \cdot u)$ and $\beta_2(u) = \sin(4\pi \cdot (u - 1/8)$, X is a standard normal random variable, and u is uniformly distributed on (0,1). The random error ε follows a standard normal distribution $N(0, 1)$. In our simulation, u, X, ε are mutually independent. This model is referred to as *exponential model*. The coefficient functions $\beta_1(u)$ and $\beta_2(u)$ are selected to represent typical functions, namely, $\beta_1(u)$ is a one-mode function, and $\beta_2(u)$ is a two-mode function. See Figure 4.2 for the plots of $\beta_1(u)$ and $\beta_2(u)$.

Before we present the simulation results, let us give some computation details. Using Taylor's expansion in a neighborhood of given u_0, for $j = 1, 2$,

$$\beta_j(u) \approx \beta_j(u_0) + \beta_j(u_0)(u - u_0) \equiv a_j + b_j(u - u_0).$$

Denote $\mathbf{a} = (a_1, a_2)^T$ and $\mathbf{b} = (b_1, b_2)^T$. Thus, we obtain a local linear regression estimator $(\widehat{\mathbf{a}}^T, \widehat{\mathbf{b}}^T)^T$ by minimizing local nonlinear least squares function

$$\ell(\mathbf{a}, \mathbf{b}) \;=\; \frac{1}{2}\sum_{i=1}^{n}[y_i - exp\ \{\mathbf{x_i}^T(\mathbf{a} + \mathbf{b}(u_i - u_0))\}]^2 K_h(u_i - u_0) \tag{4.3}$$

where $\mathbf{x_i} = \{1, x_i\}$ and $\mathbf{a} = (a_1, a_2)^T$ and $\mathbf{b} = (b_1, b_2)^T$.

Minimizing $\ell(\mathbf{a}, \mathbf{b})$ yields an estimate for $\boldsymbol{\beta}(u_0)$. For a given initial value $(\mathbf{a}_0^T, \mathbf{b}_0^T)^T$ of $(\mathbf{a}^T, \mathbf{b}^T)^T$, we locally and linearly approximate

$$\begin{aligned}
& f\{\mathbf{x}, \mathbf{a} + \mathbf{b}(u - u_0)\} \\
\approx\ & f\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0(u - u_0)\} + \{(\mathbf{a} - \mathbf{a}_0) + (\mathbf{b} - \mathbf{b}_0)(u - u_0)\}^T f'\{\mathbf{x}, \mathbf{a}_0 + \mathbf{b}_0(u - u_0)\} \\
=\ & \exp\ \{\mathbf{x}^T(\mathbf{a}_0 + \mathbf{b}_0(u_i - u_0))\} + \\
& \{(\mathbf{a} - \mathbf{a}_0) + (\mathbf{b} - \mathbf{b}_0)(u - u_0)\}^T \{exp\ [\mathbf{x}^T(\mathbf{a}_0 + \mathbf{b}_0(u_i - u_0))]\ \mathbf{x}\}
\end{aligned}$$

where $f'(\mathbf{x}, \boldsymbol{\beta}) = \partial f(\mathbf{x}, \boldsymbol{\beta})/\partial \boldsymbol{\beta}$. With this approximation, we can use an iterated least squares algorithm to search the solution of (4.3). Specifically, at the k-th step during the course of iteration, the current value for $\mathbf{a}$ and $\mathbf{b}$ is $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$. Denote

$$\begin{aligned} y_{i,k} &= y_i - f\{\mathbf{x}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0)\} + \{\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0)\} f'\{\mathbf{x}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0)\} \\ &= y_i - \exp\ \{\mathbf{x}^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\} + \\ &\quad \{(\mathbf{a} - \mathbf{a}^{(k)}) + (\mathbf{b} - \mathbf{b}^{(k)})(u - u_0)\}^T \{\exp\ [\mathbf{x}^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))]\ \mathbf{x}\}, \end{aligned}$$

$\mathbf{y}_k = (y_{1,k}, \cdots, y_{n,k})^T$, and F_k is given by

$$\begin{pmatrix} f'\{\mathbf{x}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_1 - u_0)\}, & \cdots & f'\{\mathbf{x}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_n - u_0)\} \\ (u_1 - u_0) f'\{\mathbf{x}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_1 - u_0)\}, & \cdots & (u_n - u_0) f'\{\mathbf{x}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_n - u_0)\}^T \end{pmatrix}$$

$$= \begin{pmatrix} \exp\ \{\mathbf{x}_1^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\ \mathbf{x}\}, & \cdots & \exp\ \{\mathbf{x}_n^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\ \mathbf{x}\} \\ (u_1 - u_0)\exp\ \{\mathbf{x}_1^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\ \mathbf{x}\}, & \cdots & (u_n - u_0)\exp\ \{\mathbf{x}_n^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\ \mathbf{x}\} \end{pmatrix}$$

Then, we update $(\mathbf{a}, \mathbf{b})^T$ by

$$\begin{pmatrix} \mathbf{a}^{(k+1)} \\ \mathbf{b}^{(k+1)} \end{pmatrix} = (F_k^T W F_k)^{-1} F_k^T W \mathbf{y}_k,$$

where $W = \text{diag}\{K_h(u_1 - u_0), \cdots, K_h(u_n - u_0)\}$ When the algorithm converges, the solution is satisfied with

$$\ell'(\mathbf{a}, \mathbf{b}) = 0.$$

Denote the resulting estimate of $(\mathbf{a}, \mathbf{b})^T$ by $(\widehat{\mathbf{a}}, \widehat{\mathbf{b}})^T$. Then

$$\widehat{\boldsymbol{\beta}}(u_0) = \widehat{\mathbf{a}}, \quad \text{and} \quad \widehat{\boldsymbol{\beta}}'(u_0) = \widehat{\mathbf{b}}.$$

In this example, we take sample size n=250, 500, and 1000. The value of bandwidth h is essential to local modeling technique, so we select bandwidth three different bandwidths for each sample size n, in order to compare the performance of proposed estimation procedure at a wide range of bandwidth. For $n = 250$, the bandwidth is taken to be 0.05, 0.10 and 0.20; for $n = 500$, $h = 0.0375, 0.075$ and 0.15; and $h = 0.03, 0.06$ and 0.12 for $n = 1000$.

Figure 4.1 depicts the boxplots for RASE values over 400 simulations for all sample sizes and bandwidths. Figures 4.2, 4.3 and 4.4 depict the estimate of the coefficient functions from a typical sample, whose RASE value corresponds the median of the RASE values over 400 simulations. Figures 4.2, 4.3 and 4.4 show that the proposed estimation procedure achieves quite favorable results for a broad range of bandwidth choices.

The simulation results are summarized in Table 4.1. In Table 4.1, μ and σ denote the mean and standard deviation of the RASE values in 400 simulations. For any fixed sample size level, the RASE values generally achieve its minimal value at certain bandwidth h, the so-called optimal bandwidth. The optimal bandwidth value is not shown here. The RASE values increase as the bandwidth h increases or decreases from the optimal bandwidth. This is because a larger bandwidth value implies a wider neighborhood and thus more local data, which leads to a smoother estimated curve but at a loss of greater bias; a smaller bandwidth implies a shorter neighborhood and thus less local points, which leads to a more fluctuated estimated curve but at a gain of less bias. Either a smoother estimated curve or a more fluctuated estimated curve, the RASE values will increase due to either more variance or more bias. This can also be verified from the graphical comparisons of estimated coefficient function at different bandwidth values at a fixed sample size level. In the plots, we observe a pattern that a larger bandwidth h leads to a greater estimation bias.

However, the computation in estimation becomes difficult when bandwidth h is too small. This is because when bandwidth h is too small, it implies not sufficient number of local data involved in estimation. As shown in the estimation plots, a larger sample size allows an even smaller bandwidth h for the density of data are higher and thus the number of local data available is larger.

Table 4.1: Summary of Simulation Output for Example 1

n	h	$\mu_{(RASE)}$	$\sigma_{(RASE)}$
	0.05	0.3097	0.1303
250	0.10	0.2159	0.0645
	0.20	0.3649	0.1204
	0.0375	0.2009	0.0625
500	0.075	0.1461	0.0339
	0.150	0.2237	0.0491
	0.03	0.1353	0.0334
1000	0.06	0.1028	0.0321
	0.12	0.1512	0.0334

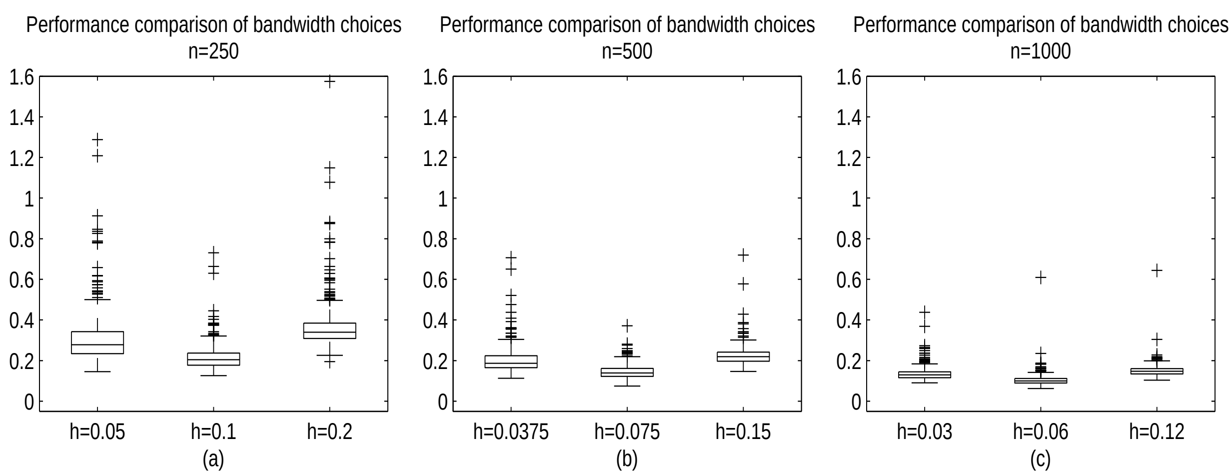

Figure 4.1: Boxplots of RASE values for Example 1. (a), (b) and (c) are boxplots of RASE values for $n = 250, 500, 1000$, respectively. For $n = 250$, bandwidth $h = 0.05, 0.10$ and 0.20; for $n = 500$, bandwidth h =0.0375, 0.075, and 0.150; for n=1000, bandwidth h =0.03, 0.06, and 0.12.

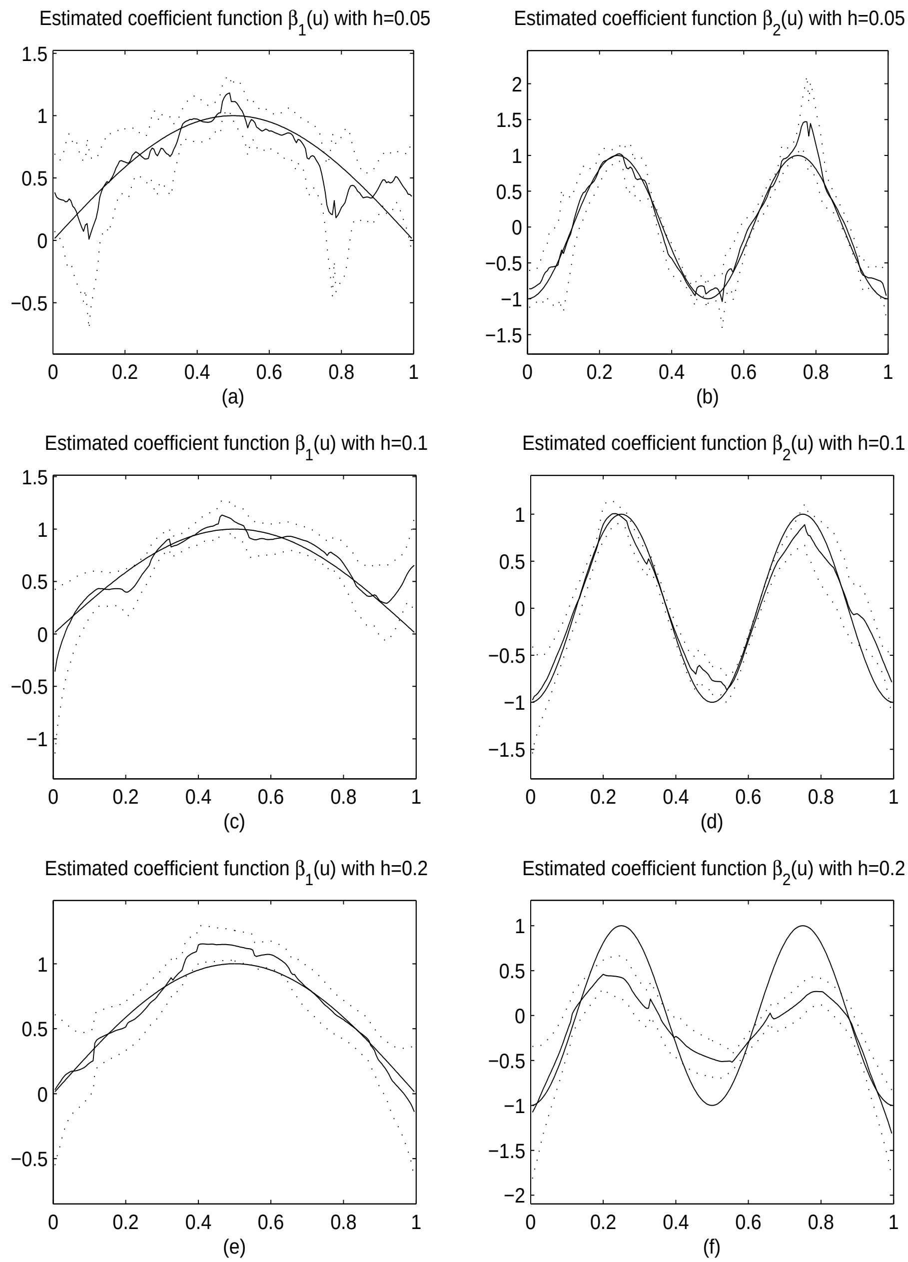

Figure 4.2: Estimated coefficients based on a typical example with $n = 250$. (a)-(b) are plots for $\beta_1(u)$ and $\beta_2(u)$ using bandwidth $h = 0.05$; (c)-(d) are plots for $h = 0.10$ and (e)-(f) are plots for $h = 0.20$. The fluctuated solid curves are estimated coefficient functions and the dotted curves are pointwise 95% confidence intervals obtained using standard error formulas; the one-mode and two-mode solid curves are the true coefficient functions.

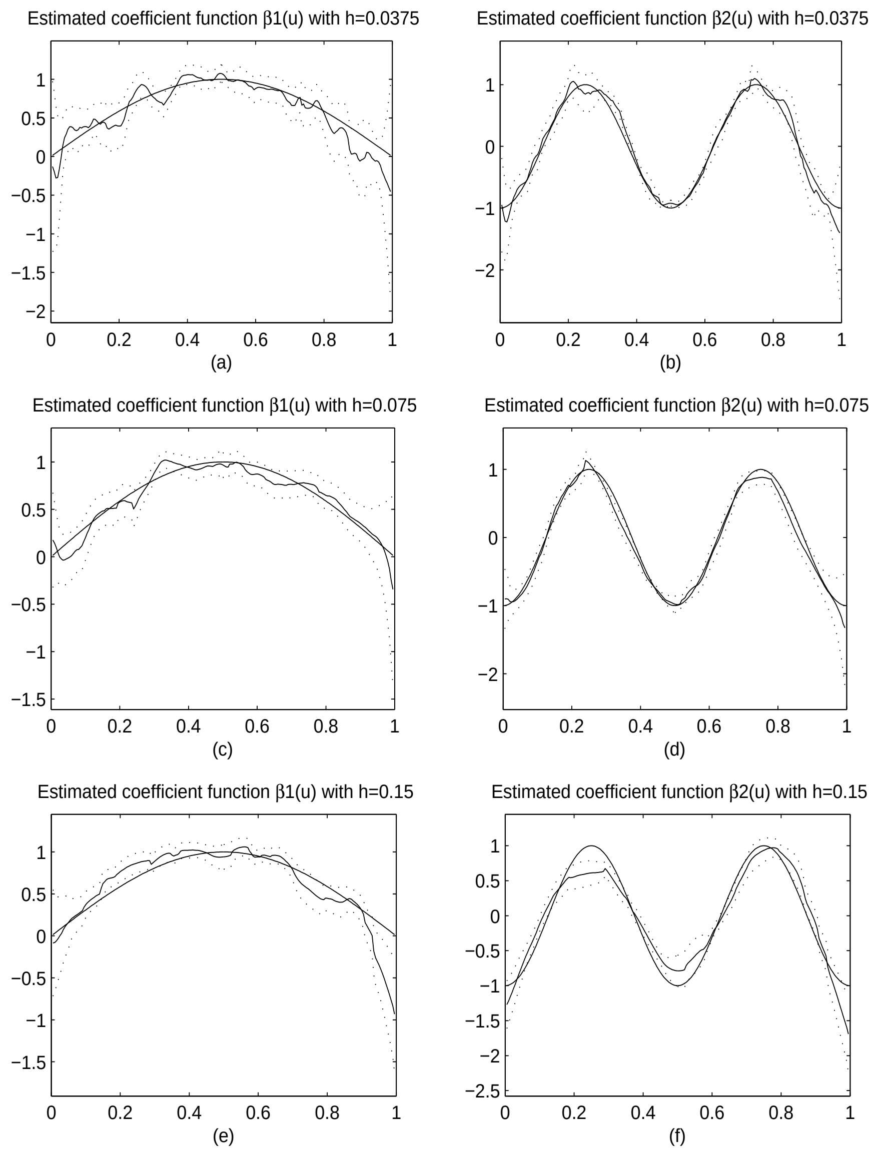

Figure 4.3: Estimated coefficients based on a typical example with $n = 500$. Caption is similar to Figure 4.2.

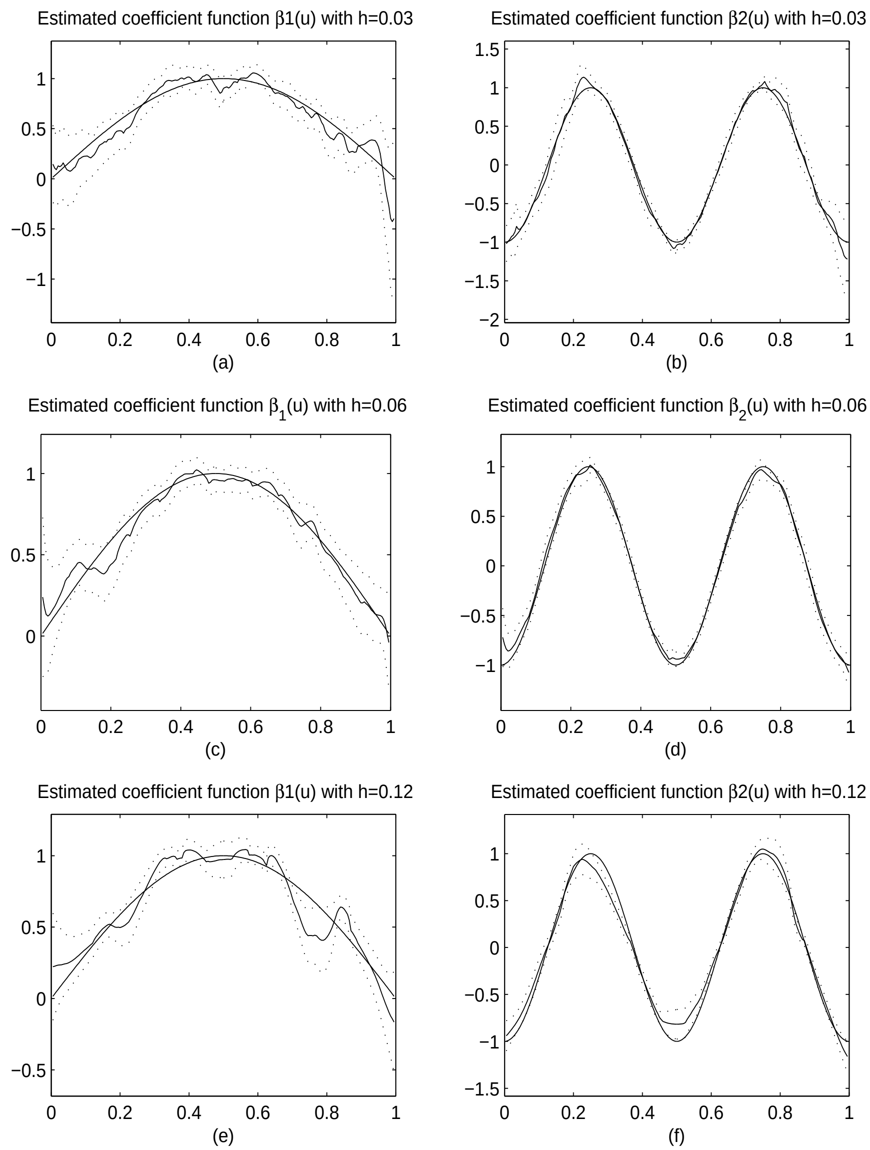

Figure 4.4: Estimated coefficients based on a typical example with $n = 1000$. Caption is similar to Figure 4.2.

We next test the accuracy of our proposed standard error formula, derived from a sandwich formula: conditioning on $(u_1, \mathbf{x}_1), \cdots, (u_n, \mathbf{x}_n)$,

$$\widehat{\text{cov}}\{\begin{pmatrix} \widehat{\mathbf{a}} \\ \widehat{\mathbf{b}} \end{pmatrix}\} = (F^T W F)^{-1} F^T W \Sigma W F (F^T W F)^{-1},$$

where $F = F_k$ with $\mathbf{a}^{(k)} = \widehat{\mathbf{a}}$ and $\mathbf{b}^{(k)} = \widehat{\mathbf{b}}$, and $\Sigma = \text{diag}\{e_1^2, \cdots, e_n^2\}$, where $e_i = y_i - f\{\mathbf{x}_i, \widehat{\boldsymbol{\beta}}(u_i)\}$.

In Table 4.2, the standard deviation of the 400 estimated $\widehat{\beta}_j(u_0)$ for $u_0 = 0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9$, based on 400 simulations, is denoted by SD. We can regard SDs as the true standard errors. We use SD_a to denote the average of 400 estimated standard errors; and use SD_std to denote the standard deviation of 400 estimated standard errors. Thus, SD_a and SD_std summarize the performance of the standard error formulas. The results summarized in Table 4.2 suggest that our standard error formula somewhat underestimates the true standard error.

Table 4.2 suggests that the proposed standard error formula works well for most cases since the difference in the standard error is still within two standard deviations of the Monte Carlo errors. The bias becomes smaller for a larger value of nh_n, which is consistent with our asymptotic theory. For large bandwidth, such as $h = 0.20$ for $n = 250$, $h = 0.15$ for $n = 500$ and $h = 0.12$ for $n = 1000$, some SD values are significantly greater than SD_a. We have further checked the individual SD values and found that there are some outliers of SD values, which may caused due to the divergence of the proposed algorithm. We calculate the robust estimate of standard deviation by using median of absolute deviation dividing by a factor 0.6745. The robust estimate for the standard deviation is very close to the SD_a values. For example, when $n = 250$ and $h = 0.20$, the robust estimate for SD is 0.1381, 0.1232 and 0.1312 for u =0.3, 0.5 and 0.7, respectively, compared with its sample standard deviation 0.2195, 0.2574 and 0.1919 in Tabl 4.2.

We now examine the performance of the proposed generalized likelihood ratio test.

Table 4.2: Summary of Standard Deviations and Standard Errors for Example 1

			$\hat{\beta}_1(u)$		$\hat{\beta}_2(u)$	
n	h	u	SD	$SD_a(SD_{std})$	SD	$SD_a(SD_{std})$
		0.1	0.2604	0.1974 (0.1045)	0.2920	0.1787 (0.1180)
		0.3	0.1295	0.1121 (0.0355)	0.1364	0.0937 (0.0516)
	0.05	0.5	0.1145	0.0950 (0.0307)	0.1257	0.0795 (0.0458)
		0.7	0.1328	0.1149 (0.0399)	0.1304	0.0934 (0.0488)
		0.9	0.2350	0.1945 (0.0933)	0.2377	0.1748 (0.0942)
		0.1	0.1368	0.1278 (0.0311)	0.1292	0.1047 (0.0404)
		0.3	0.1031	0.0862 (0.0214)	0.1030	0.0706 (0.0263)
250	0.10	0.5	0.1329	0.0844 (0.0367)	0.1352	0.0717 (0.0288)
		0.7	0.0994	0.0848 (0.0174)	0.1006	0.0690 (0.0237)
		0.9	0.1459	0.1276 (0.0304)	0.1273	0.1011 (0.0357)
		0.1	0.1211	0.1000 (0.0226)	0.1478	0.0903 (0.0349)
		0.3	0.2195	0.1043 (0.0478)	0.2470	0.1111 (0.0430)
	0.20	0.5	0.2574	0.0985 (0.0648)	0.2501	0.1115 (0.0424)
		0.7	0.1919	0.0970 (0.0379)	0.2394	0.1063 (0.0379)
		0.9	0.1163	0.0991 (0.0191)	0.1338	0.0844 (0.0297)
		0.1	0.1675	0.1464 (0.0412)	0.1545	0.1272 (0.0506)
		0.3	0.0992	0.0903 (0.0228)	0.0827	0.0671 (0.0293)
	0.0375	0.5	0.0769	0.0721 (0.0167)	0.0643	0.0533 (0.0241)
		0.7	0.0997	0.0896 (0.0217)	0.0867	0.0672 (0.0294)
		0.9	0.1718	0.1510 (0.0505)	0.1649	0.1290 (0.0560)
		0.1	0.1107	0.1022 (0.0184)	0.0939	0.0823 (0.0249)
		0.3	0.0747	0.0648 (0.0106)	0.0655	0.0489 (0.0159)
500	0.075	0.5	0.0835	0.0585 (0.0111)	0.0844	0.0452 (0.0167)
		0.7	0.0732	0.0642 (0.0099)	0.0698	0.0483 (0.0151)
		0.9	0.1051	0.1018 (0.0185)	0.0961	0.0818 (0.0236)
		0.1	0.0748	0.0745 (0.0098)	0.0714	0.0588 (0.0159)
		0.3	0.1246	0.0682 (0.0223)	0.1425	0.0719 (0.0252)
	0.15	0.5	0.1730	0.0790 (0.0436)	0.1730	0.0840 (0.0333)
		0.7	0.1237	0.0695 (0.0247)	0.1395	0.0740 (0.0288)
		0.9	0.0776	0.0747 (0.0098)	0.0689	0.0587 (0.0153)
		0.1	0.1268	0.1136 (0.0243)	0.1146	0.0956 (0.0273)
		0.3	0.0776	0.0711 (0.0145)	0.0577	0.0481 (0.0176)
	0.03	0.5	0.0550	0.0550 (0.0107)	0.0434	0.0357 (0.0150)
		0.7	0.0727	0.0690 (0.0131)	0.0537	0.0463 (0.0178)
		0.9	0.1110	0.1111 (0.0219)	0.1048	0.0921 (0.0257)
		0.1	0.0821	0.0795 (0.0113)	0.0712	0.0638 (0.0156)
		0.3	0.0559	0.0500 (0.0064)	0.0453	0.0346 (0.0104)
1000	0.06	0.5	0.0607	0.0443 (0.0069)	0.0544	0.0316 (0.0112)
		0.7	0.0538	0.0490 (0.0064)	0.0435	0.0337 (0.0104)
		0.9	0.0776	0.0779 (0.0110)	0.0714	0.0618 (0.0130)
		0.1	0.0621	0.0562 (0.0057)	0.0520	0.0432 (0.0097)
		0.3	0.0810	0.0504 (0.0151)	0.0867	0.0520 (0.0184)
	0.12	0.5	0.1327	0.0639 (0.0278)	0.1207	0.0631 (0.0229)
		0.7	0.0828	0.0498 (0.0142)	0.0883	0.0508 (0.0176)
		0.9	0.0578	0.0557 (0.0058)	0.0548	0.0425 (0.0092)

After fitting a nonlinear model with varying-coefficients, a natural question is whether the coefficients are really varying; and if any particular coefficient is constant, whether or not it is significant in the model. As illustrations, we consider

$$H_0 : \beta_j(u) = \gamma_{j0}, \quad j = 1 \cdots, p \qquad \text{versus} \qquad H_1 : \beta_j(u) \neq \gamma_{j0}, \qquad \text{for at least one j}$$

where γ_{j0} is unknown constant.

Here we will empirically demonstrate that the GLRT has a chi-square limiting distribution for every specific nonparametric model, and the Wilks type of results continue to hold for our setting. To this end, we conduct simulation study to evaluate whether the asymptotic null distribution of the test statistic T depends on the values of $\{\gamma_{j0}\}$ under H_0, and whether the limiting conditional null distributions are dependent on the covariate values.

For this purpose, we estimate the distribution of test statistic T under H_0 using $n = 1000$ via 1000 Monte Carlo simulations. We provide empirical justifications by estimating the null distribution of T for the five sets of covariate values by using a bootstrap procedure.

To obtain the null distribution, we consider both parametric bootstrap procedure and nonparametric bootstrap procedure. However, we should give special attention to the nonparametric bootstrap procedure, since it is the practical procedure for empirical setting when we handle real data sets.

Parametric bootstrap procedure. For parametric bootstrap procedure, we generate sample of size $n = 1000$, based on model

$$y_i = exp\ (\mathbf{X}^T \beta(u)) + \varepsilon_i,$$

where $\mathbf{X} = \{1,\ X_i\}$ and ε_i is standard normal. We then compute the test statistic

$$T_0 = \frac{n}{2} \frac{\text{RSS}(H_0) - \text{RSS}(H_1)}{\text{RSS}(H_1)}.$$

We conduct 1000 simulations to generate 1000 T_0 values, and then estimate the distribution of the test statistic T_0 based on the T_0's from 1000 simulations.

Nonparametric bootstrap procedure. For nonparametric bootstrap procedure, we rely on only one data set to follow the practical situation when real data is available. As if we have a real data set, we generate a sample of size $n = 1000$, based on model

$$\begin{aligned} y_i &= f(\mathbf{x}_i; \boldsymbol{\beta}(u_i)) + \varepsilon_i && (4.4) \\ &= exp\ (\mathbf{X}^T \beta(u)) + \varepsilon_i && (4.5) \end{aligned}$$

where $\mathbf{X} = \{1,\ X_i\}$ and ε_i is standard normal. However, as if in the empirical setting when only one data set is available, we estimate the model coefficients under null hypothesis and alternative hypothesis based on the sample of size $n = 1000$. We now use $\tilde{\boldsymbol{\beta}}(u_i)$ to denote the estimate of model coefficient under null hypothesis and $\widehat{\boldsymbol{\beta}}(u_i)$ to denote the estimate of model coefficient under alternative hypothesis. We then obtain residuals e_i through

$$e_i = y_i - f(\mathbf{x}_i; \tilde{\boldsymbol{\beta}}(u_i)) \tag{4.6}$$

We now generate data $(y_i^*,\ x_i,\ u_i)$ of size $n = 1000$ using the model

$$y_i^* = f(\mathbf{x}_i; \tilde{\boldsymbol{\beta}}(u_i)) + e_i^* \tag{4.7}$$

where e_i^* represents the residuals resampled from model (4.6). We then compute the test statistic

$$T_0 = \frac{n}{2} \frac{\mathrm{RSS}(H_0) - \mathrm{RSS}(H_1)}{\mathrm{RSS}(H_1)}.$$

We conduct 1000 simulations to generate 1000 T_0 values, and then estimate the distribution of the test statistic T_0 based on the T_0's from 1000 simulations.

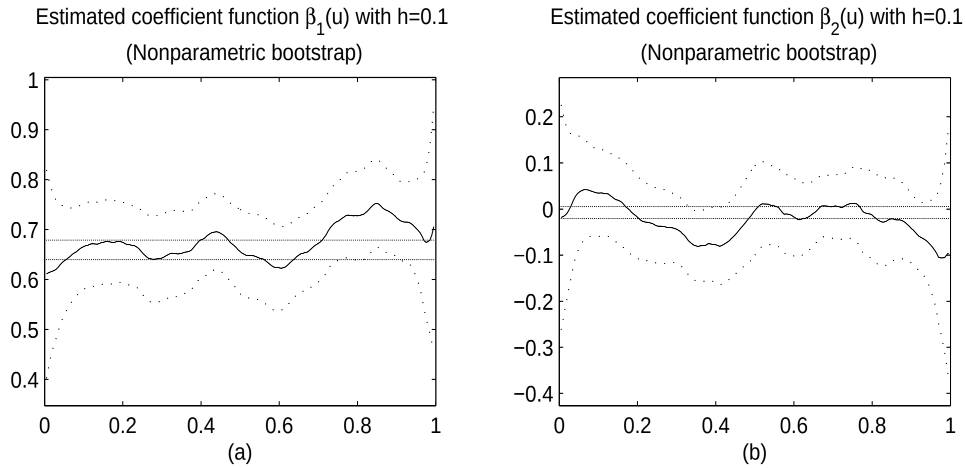

Figure 4.5: The estimated coefficient functions under null hypothesis. The fluctuated solid curves are estimated coefficient functions for the exponential model; the dotted curves are pointwise 95% confidence intervals obtained using standard error formulas; the solid lines are true coefficient functions; the dashed lines are the true function coefficients.

Figure 4.5 illustrates the nonparametric estimated coefficient functions under null hypothesis. The dashed lines are the true function coefficients. The solid lines are the estimated function coefficients within the null space of constant coefficients; while the solid curves are the nonparametric estimated coefficient functions within the alternative space of varying-coefficients, with pointwise 95% confidence interval bands in dashed curves.

Figure 4.6 depicts the estimated null distribution of test statistic $T \equiv r_K T_0$ compared to a Chi-squared distribution with degree of freedom by Wilks phenomenon. The null distribution of test statistic T are estimated through both parametric bootstrap and nonparametric bootstrap described above. From the figure it shows that under null hypothesis, test statistic T has a Chi-squared distribution with degree of freedom 20.65.

We next further demonstrate that the null distribution of the GLRT does not depend on the specific values of unknown regression coefficients. To address this, we use five different

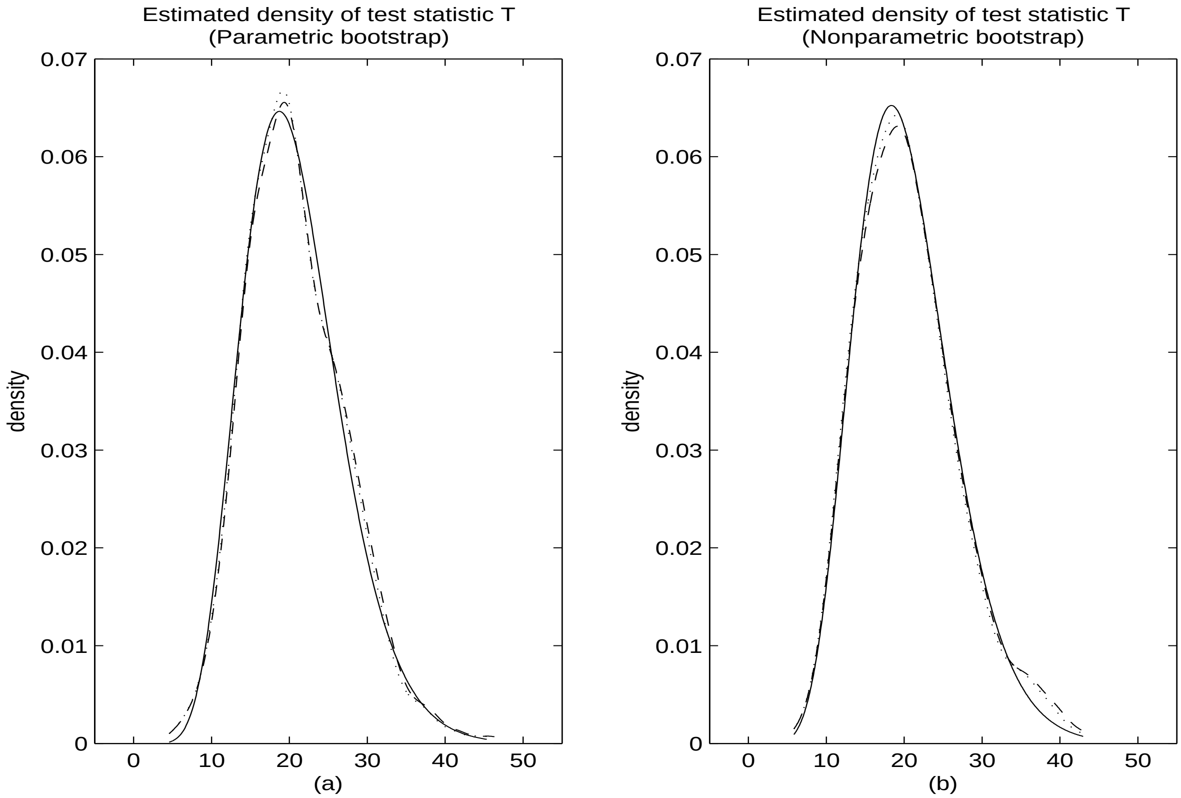

Figure 4.6: The estimated density of test statistic T by Monte Carlo Simulation. The dotted curve is the estimated density of generalized likelihood ratio test statistic T; the dash-dotted curve is the estimated density of generalized F-test statistic T_0. The solid curve is the density of Chi-squared distribution with degrees of freedom 20.65.

sets of values $\{\gamma_{j0}\}$. The five sets of values are quite far apart. Namely,

$$\begin{aligned}\{\gamma_{j0}\} \;\; &= \;\; \{\ (\beta_1,\ \beta_2),\ (\beta_1 - 2\cdot\mathrm{std}(\widehat{\beta}_1),\ \beta_2),\ (\beta_1 + 2\cdot\mathrm{std}(\widehat{\beta}_1),\ \beta_2),\\ &\qquad (\beta_1,\ \beta_2 - 2\cdot\mathrm{std}(\widehat{\beta}_2)\ ),\ (\beta_1,\ \beta_2 + 2\cdot\mathrm{std}(\widehat{\beta}_2)\ )\ \}\end{aligned}$$

where $\beta_j = E\{\beta_j(u)\}$ for $j = 1, 2$, and $\mathrm{std}(\widehat{\beta}_j)$ is the standard deviation of $\widehat{\beta}_j$ for $j = 1, 2$.

Figure 4.7 depicts the estimated null distribution of test statistic T for five different sets of covariate values to a Chi-squared distribution with degree of freedom 20.65. The null distribution of test statistic T are estimated through both parametric bootstrap and nonparametric bootstrap described above. This figure empirically justifies Wilks type of

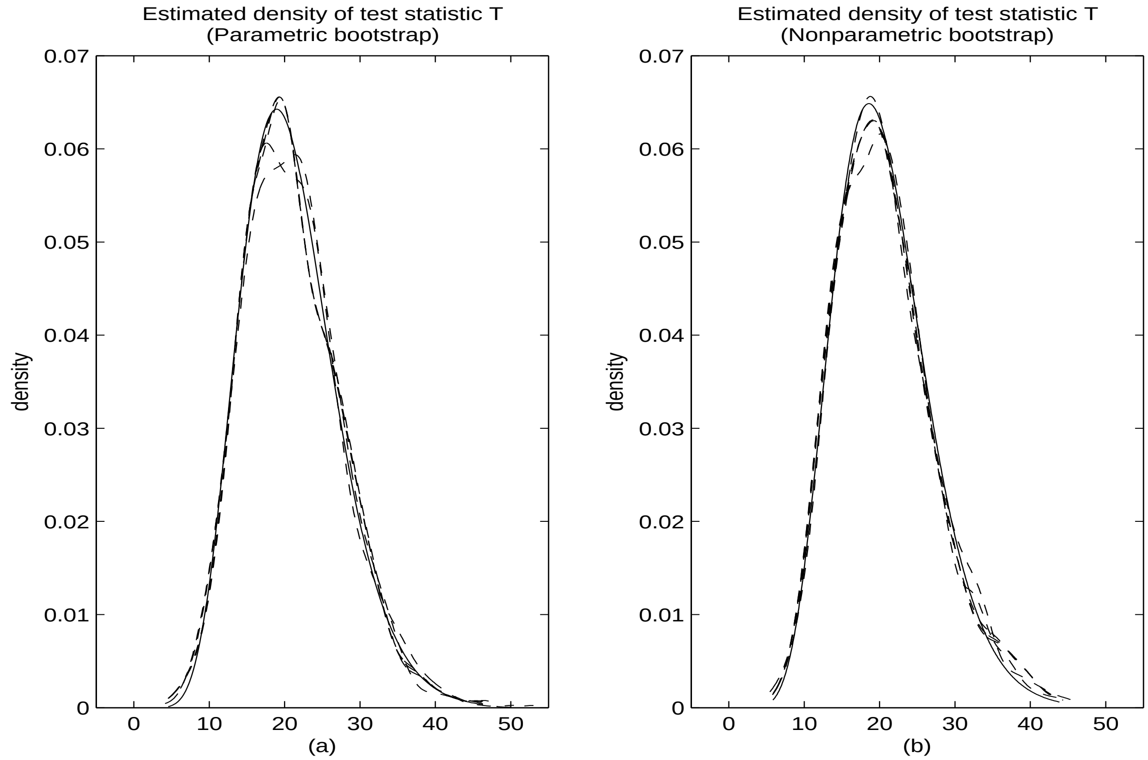

Figure 4.7: The estimated densities of test statistic T by Monte Carlo Simulation, using 5 different coefficient values. The five dashed curves are estimated densities. The solid curve is Chi-squared density with degree of freedom 20.65

results that the null distribution of test statistic T follows Chi-squared distribution and is independent of the covariate values.

In what follows, we examine the power of the proposed hypothesis test procedure. We will consider the null hypothesis

$$H_0 : \beta_j(u) = \gamma_{j0}, \qquad j = 1, 2$$

versus

$$H_1 : \beta_j(u) \neq \gamma_{j0}, \qquad \text{for at least one } \ j$$

Where γ_{j0} are constants. We then evaluate the power functions of the proposed test under a sequence of alternative models that are indexed by δ,

$$H_1 : \beta_j(u, \delta) = (1 - \delta)\gamma_{j0} + \delta\beta_j^0(u), \qquad j = 1, 2, (0 \leq \delta \leq 0.8)$$

where $\beta_1^0(u) = sin(\pi{\cdot}u)$ and $\beta_2^0(u) = sin(4\pi{\cdot}(u{-}1/8))$ as defined previously in the simulation section for model estimation, and $\gamma_{j0} = E\{\beta_j^0(u)\}$.

The power functions are depicted in Figure 4.8. In Figure 4.8, we plot five power functions at five different significance levels: 0.50, 0.25, 0.10, 0.05, and 0.01, based on 1000 simulations for sample size n=1000. The alternative hypothesis is chosen such that when $\delta = 0$, the alternative collapses into the null hypothesis. In particular, the powers at $\delta = 0$ for the five significance levels are 0.505, 0.248, 0.105, 0.054, and 0.011. This shows that the nonparametric bootstrap method provides the correct levels of test. Also, as $\delta = 0$ value increases, the signal of varying-coefficient carried in alternative hypothesis is amplified. Therefore, the power should increase as the value of δ increases. The results shown in figure 4.8 exhibit that the powers increase very rapidly as δ increases, which justifies that our proposed testing procedures works well.

4.2 Example 2

In this section, we generate data from the following model

$$Y = \frac{c \cdot \exp\ (\mathbf{x}^T\beta(u))}{1 + \exp\ (\mathbf{x}^T\beta(u))} + \varepsilon. \tag{4.8}$$

with $c = 10$. The coefficient $\boldsymbol{\beta}$ and the distribution of $\{\mathbf{x}, u, \varepsilon\}$ are the same as those in Section 4.1. This model is referred to as *logistic model*

Let us begin with some computation details. The local least squares function is

$$\ell(\mathbf{a}, \mathbf{b}) = \frac{1}{2}\sum_{i=1}^{n}[\ y_i - \frac{c \cdot \exp\ \{\mathbf{x_i}^T(\mathbf{a} + \mathbf{b}(u_i - u_0))\}}{\exp\ \{\mathbf{x_i}^T(\mathbf{a} + \mathbf{b}(u_i - u_0))\} + 1}\ ]^2 K_h(u_i - u_0). \tag{4.9}$$

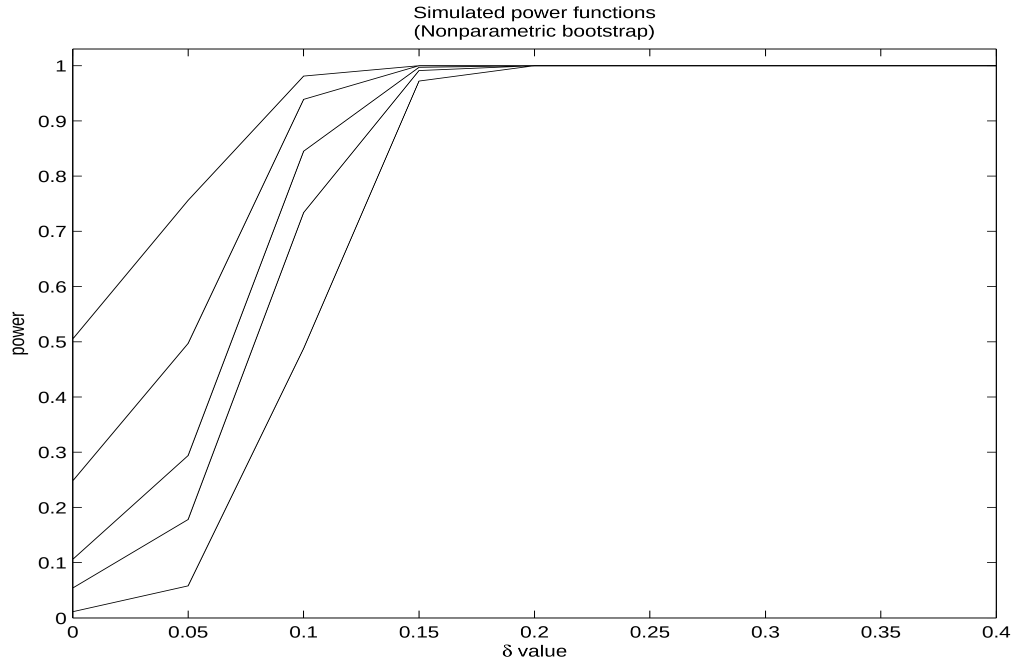

Figure 4.8: The simulated power functions at five different significance levels: 0.5, 0.25, 0.10, 0.05, and 0.01.

We will use an iterated least squares algorithm to search the solution of (4.9). Specifically, at the k-th step during the course of iteration, the current value for $\mathbf{a}$ and $\mathbf{b}$ is $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$. Denote

$$\begin{aligned} y_{i,k} &= y_i - \frac{c \cdot exp\,\{\mathbf{x}^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\}}{exp\,\{\mathbf{x}^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\} + 1} + \\ & \quad \{(\mathbf{a} - \mathbf{a}^{(k)}) + (\mathbf{b} - \mathbf{b}^{(k)})(u - u_0)\}^T \frac{c \cdot exp\,\{\mathbf{x}^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\}\,\mathbf{x}}{\{exp\,\{\mathbf{x}^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\} + 1\}^2} \end{aligned}$$

and $\mathbf{y}_k = (y_{1,k}, \cdots, y_{n,k})^T$. Furthermore, denote

$$F_k = \begin{pmatrix} f'\{\mathbf{x}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_1 - u_0)\}, & \cdots & f'\{\mathbf{x}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_n - u_0)\} \\ (u_1 - u_0)f'\{\mathbf{x}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_1 - u_0)\}, & \cdots & (u_n - u_0)f'\{\mathbf{x}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_n - u_0)\}^T \end{pmatrix}$$

with

$$f'\{\mathbf{x}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0)\} \quad = \quad \frac{c \cdot \exp\ \{\mathbf{x}_i^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\}\ \mathbf{x}_i}{\{\exp\ \{\mathbf{x}_i^T(\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0))\} + 1\}^2}$$

In the iterated least squares algorithm, we update $(\mathbf{a}, \mathbf{b})^T$ by

$$\begin{pmatrix} \mathbf{a}^{(k+1)} \\ \mathbf{b}^{(k+1)} \end{pmatrix} = (F_k^T W F_k)^{-1} F_k^T W \mathbf{y}_k.$$

When the algorithm converges, the solution is satisfied with

$$\ell'(\mathbf{a}, \mathbf{b}) = 0.$$

Denote the resulting estimate of $(\mathbf{a}, \mathbf{b})^T$ by $(\widehat{\mathbf{a}}, \widehat{\mathbf{b}})^T$. Then

$$\widehat{\boldsymbol{\beta}}(u_0) = \widehat{\mathbf{a}}, \quad \text{and} \quad \widehat{\boldsymbol{\beta}}'(u_0) = \widehat{\mathbf{b}}.$$

In this example, we take n =250, 500, and 1000, and for each sample size, three bandwidths are chosen, corresponding under-smooth, about right-smooth and over-smooth. Figure 4.9 depicts the boxplots, and Figures 4.10, 4.11 and 4.12 depict estimated coefficients based on a typical sample, whose RASE value is the median of the RASE values over 400 simulated samples. The pattern in Figure 4.9 to 4.12 is similar to those for Example 1.

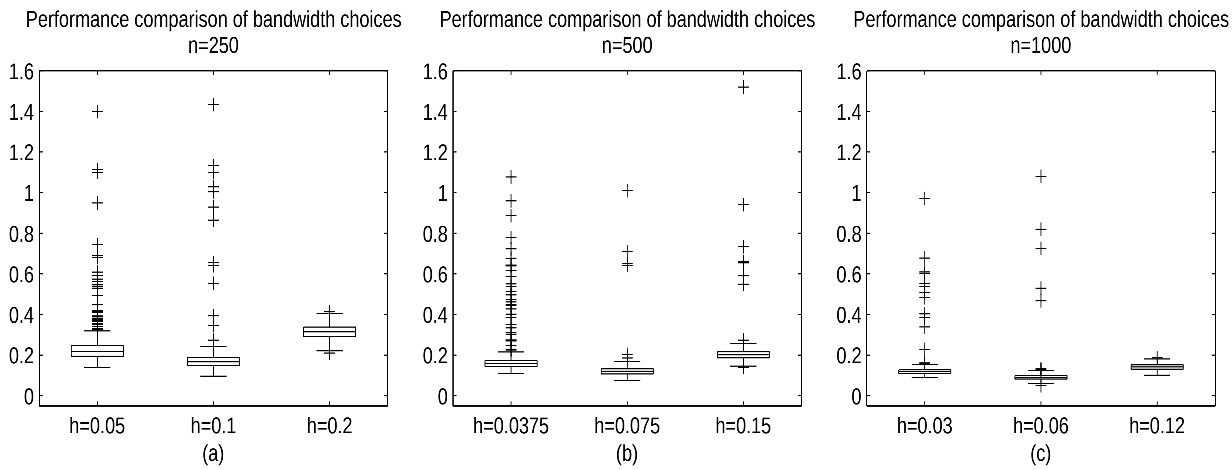

Figure 4.9: Boxplots of RASE values of Example 2. (a), (b) and (c) are boxplots of RASE values of the logistic model using sample size n=250, 500, 1000, respectively. For n=250, bandwidth h=0.05, h=0.10, and h=0.20 are used and compared; for n=500, bandwidth h=0.0375, h=0.075, and h=0.150 are used and compared; for n=1000, bandwidth h=0.03, h=0.06, and h=0.12 are used and compared.

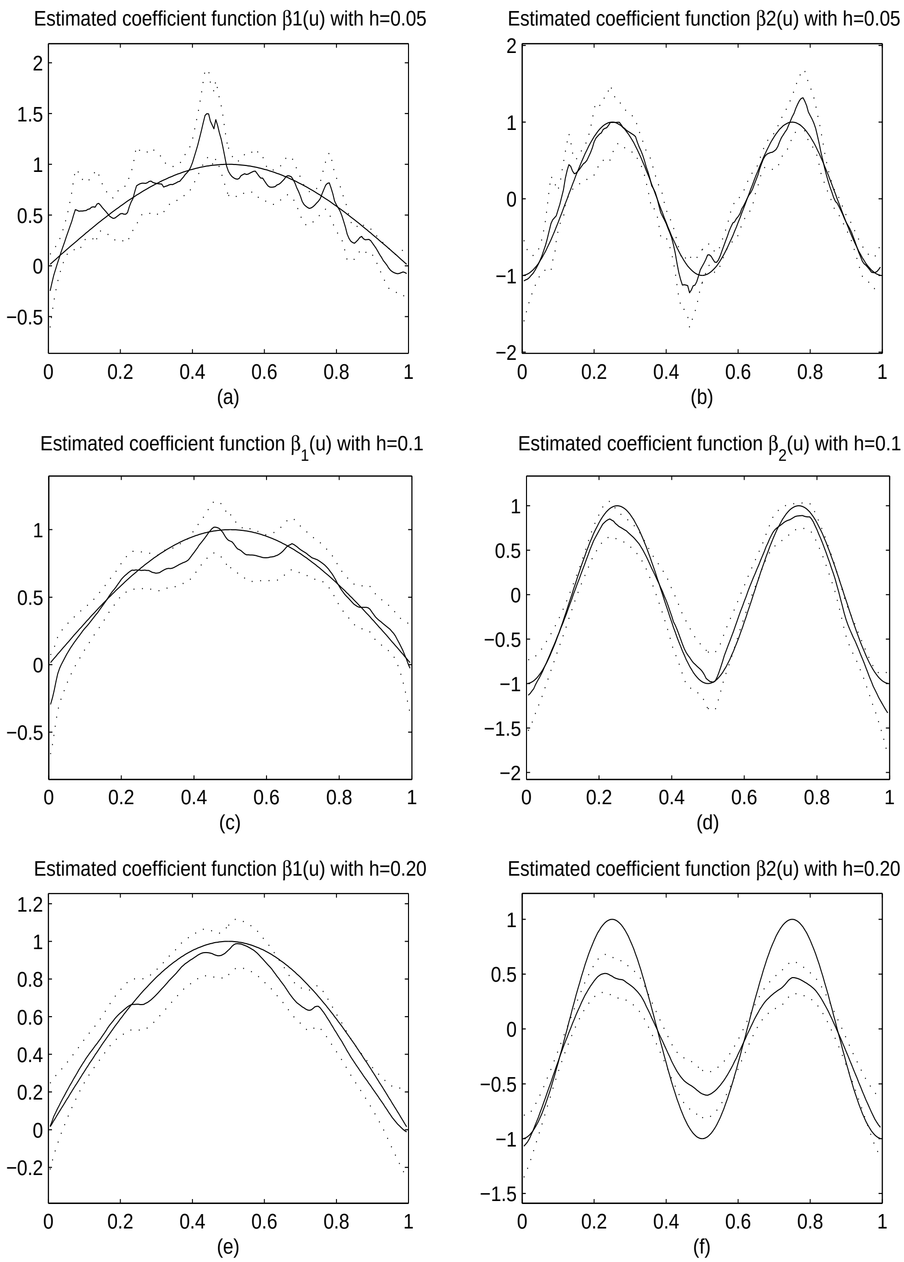

Figure 4.10: Estimated coefficients for Example 2 when $n = 250$. (a)-(b) are plots for $\beta_1(u)$ and $\beta_2(u)$ using bandwidth $h = 0.05$; (c)-(d) are for $h = 0.10$; (e)-(f) are $h = 0.20$. The fluctuated solid curves are estimated coefficient functions; the dotted curves are pointwise 95% confidence intervals obtained using standard error formulas; the one-mode and two-mode solid curves are the true coefficient functions.

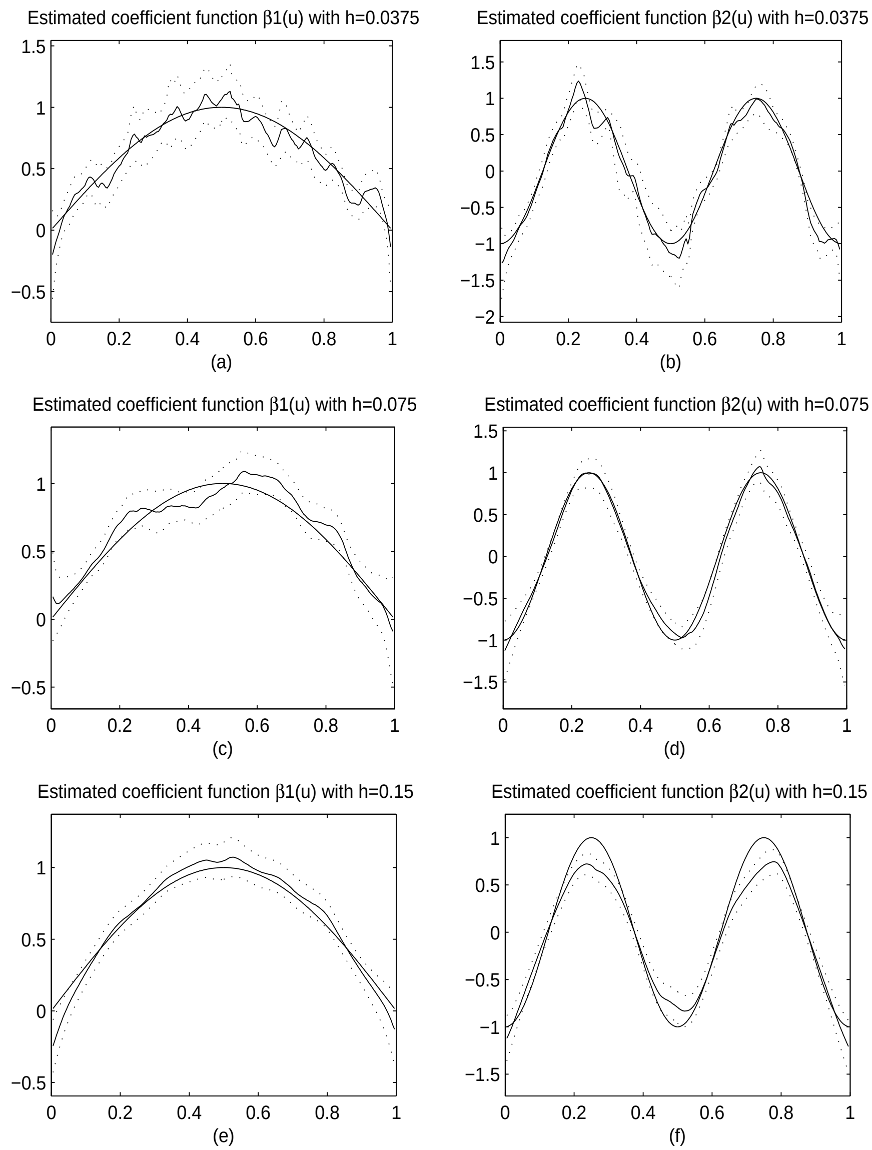

Figure 4.11: Estimated coefficients for Example 2 when $n = 500$. Caption is similar to Figure 4.10

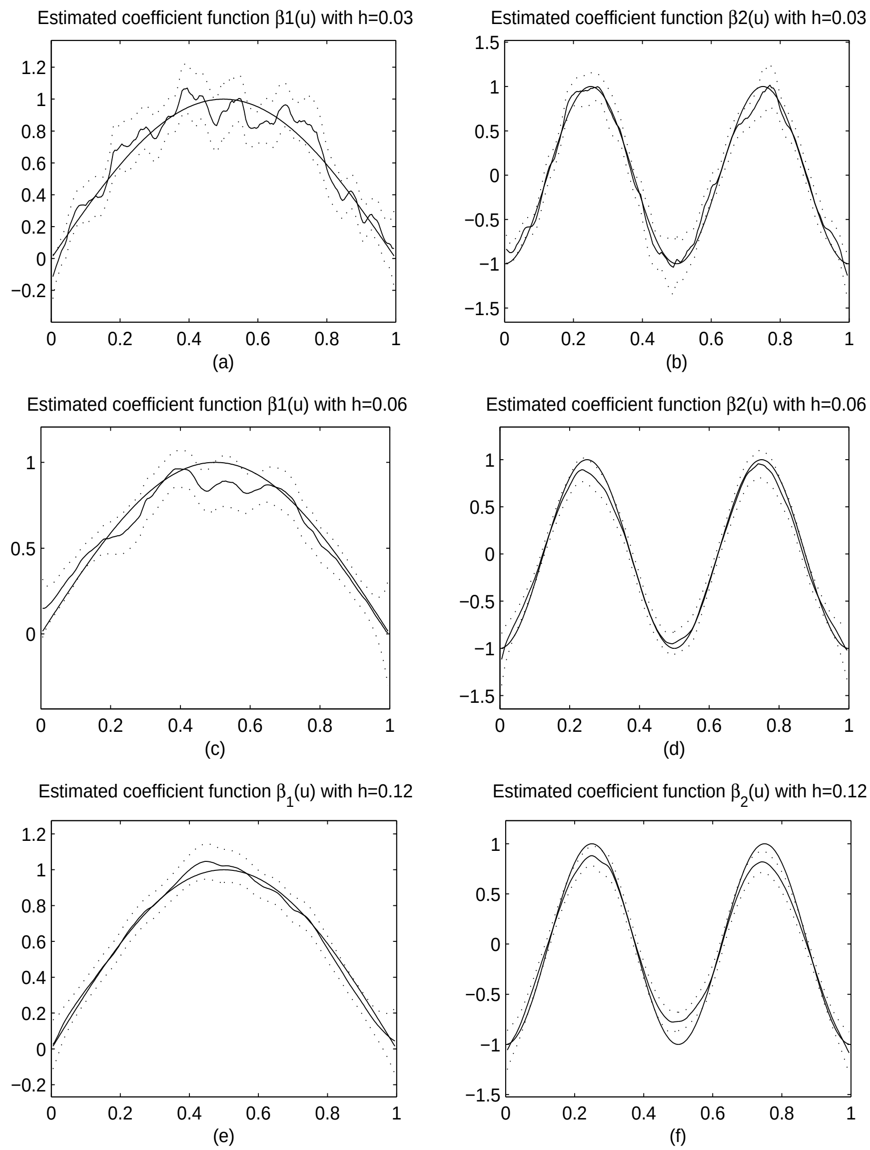

Figure 4.12: Estimated coefficients for Example 2 when $n = 500$. Caption is similar to Figure 4.10

Table 4.3: Summary of Simulation Output for Example 2

n	h	$\mu_{(RASE)}$	$\sigma_{(RASE)}$
	0.05	0.2442	0.1209
250	0.10	0.1877	0.1300
	0.20	0.3174	0.0732
	0.0375	0.1835	0.1117
500	0.075	0.1308	0.1005
	0.15	0.2116	0.0924
	0.03	0.1308	0.0771
1000	0.06	0.0990	0.0757
	0.12	0.1416	0.0163

The simulation results are summarized in Table 4.3, in which notation is the same as that in Table 4.1. The patterns in Table 4.1 and 4.3 are the same. We further test the accuracy of the proposed standard error formula. Table 4.4 displays the standard deviation of the 400 estimated $\widehat{\beta}_j(u_0)$ for $u_0 = 0.1,\ 0.3,\ 0.5,\ 0.7,\ 0.9$, based on 400 simulations, and the average of 400 estimated standard errors. It is expected that the pattern of Table 4.2 and 4.4 are the same.

In this example, we also examine the performance of the GLRT for hypothesis

$$H_0 : \beta_j(u) = \gamma_{j0}, \qquad j = 1 \cdots , p$$

versus

$$H_1 : \beta_j(u) \neq \beta_{j0}(u, \boldsymbol{\gamma}_j), \qquad \text{for at least one j}$$

where γ_{j0} is unknown constant. The null hypothesis implies that the $\beta_j(u)$ is a constant. Similar to Example 1, we consider both parametric bootstrap and nonparametric bootstrap to estimate the null distribution. Figure 4.13 illustrates the nonparametric estimated coef-

Table 4.4: Standard Deviation and Standard Errors for Example 2

n	h	u	SD	$SD_a(SD_{std})$	SD	$SD_a(SD_{std})$
		0.1	0.1483	0.0942 (0.0713)	0.1607	0.1005 (0.0697)
		0.3	0.1361	0.1168 (0.0320)	0.1798	0.1326 (0.0494)
	0.05	0.5	0.1742	0.1358 (0.0720)	0.2105	0.1523 (0.0923)
		0.7	0.2164	0.1243 (0.1411)	0.2268	0.1419 (0.1423)
		0.9	0.1319	0.0922 (0.0475)	0.3318	0.1091 (0.2072)
		0.1	0.1335	0.0691 (0.0390)	0.2527	0.0799 (0.0849)
		0.3	0.1283	0.0846 (0.0369)	0.1294	0.0961 (0.0311)
250	0.10	0.5	0.0988	0.0946 (0.0168)	0.1215	0.1067 (0.0233)
		0.7	0.0854	0.0826 (0.0148)	0.1007	0.0963 (0.0212)
		0.9	0.2633	0.0763 (0.1534)	0.3358	0.0858 (0.2009)
		0.1	0.2040	0.0635 (0.0712)	0.2373	0.0736 (0.0759)
		0.3	0.0670	0.0628 (0.0065)	0.0977	0.0811 (0.0123)
	0.20	0.5	0.2574	0.0985 (0.0648)	0.2501	0.1115 (0.0424)
		0.7	0.0617	0.0625 (0.0068)	0.0882	0.0813 (0.0132)
		0.9	0.0670	0.0602 (0.0083)	0.0775	0.0692 (0.0147)
		0.1	0.0817	0.0737 (0.0135)	0.0920	0.0764 (0.0215)
		0.3	0.1027	0.0956 (0.0207)	0.1313	0.1091 (0.0307)
	0.0375	0.5	0.1208	0.1107 (0.0237)	0.1387	0.1270 (0.0400)
		0.7	0.1066	0.0949 (0.0186)	0.1252	0.1064 (0.0287)
		0.9	0.0794	0.0747 (0.0139)	0.0945	0.0764 (0.0207)
		0.1	0.1490	0.0583 (0.0819)	0.3492	0.0687 (0.2027)
		0.3	0.0715	0.0681 (0.0100)	0.0863	0.0781 (0.0141)
500	0.075	0.5	0.0829	0.0771 (0.0109)	0.0931	0.0870 (0.0162)
		0.7	0.0699	0.0674 (0.0090)	0.0827	0.0764 (0.0137)
		0.9	0.0547	0.0542 (0.0067)	0.0668	0.0583 (0.0107)
		0.1	0.0432	0.0436 (0.0044)	0.0519	0.0502 (0.0072)
		0.3	0.0517	0.0490 (0.0049)	0.0623	0.0607 (0.0085)
	0.15	0.5	0.0575	0.0540 (0.0054)	0.0718	0.0678 (0.0096)
		0.7	0.0469	0.0484 (0.0045)	0.0641	0.0593 (0.0080)
		0.9	0.0445	0.0436 (0.0043)	0.0550	0.0504 (0.0076)
		0.1	0.0629	0.0589 (0.0080)	0.0708	0.0625 (0.0135)
		0.3	0.0861	0.0775 (0.0131)	0.0997	0.0877 (0.0199)
	0.03	0.5	0.0846	0.0885 (0.0155)	0.1023	0.1002 (0.0209)
		0.7	0.0800	0.0766 (0.0120)	0.0988	0.0875 (0.0185)
		0.9	0.0588	0.0588 (0.0080)	0.0665	0.0599 (0.0116)
		0.1	0.0433	0.0425 (0.0042)	0.0479	0.0463 (0.0074)
		0.3	0.0582	0.0545 (0.0063)	0.0669	0.0620 (0.0099)
1000	0.06	0.5	0.0591	0.0616 (0.0072)	0.0687	0.0698 (0.0098)
		0.7	0.0548	0.0542 (0.0060)	0.0646	0.0622 (0.0091)
		0.9	0.0418	0.0425 (0.0041)	0.0509	0.0450 (0.0063)
		0.1	0.0324	0.0321 (0.0023)	0.0378	0.0370 (0.0044)
		0.3	0.0375	0.0383 (0.0028)	0.0459	0.0457 (0.0048)
	0.12	0.5	0.0402	0.0425 (0.0033)	0.0539	0.0502 (0.0052)
		0.7	0.0387	0.0380 (0.0028)	0.0484	0.0457 (0.0048)
		0.9	0.0313	0.0320 (0.0022)	0.0402	0.0362 (0.0039)

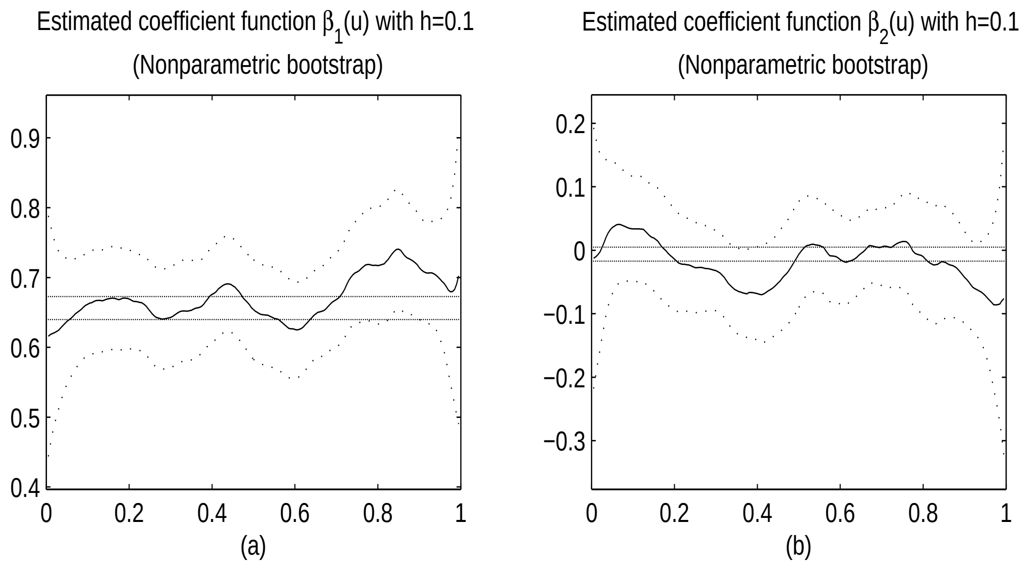

Figure 4.13: The estimated coefficient functions under null hypothesis. The fluctuated solid curves are estimated coefficient functions for Example 2; the dotted curves are pointwise 95% confidence intervals obtained using standard error formulas; the solid lines are true coefficient functions; the dashed lines are the true function coefficients.

ficient functions under null hypothesis. The dashed lines are the true function coefficients. The solid lines are the estimated function coefficients within the null space of constant coefficients; while the solid curves are the nonparametric estimated coefficient functions within the alternative space of varying-coefficients, with pointwise 95% confidence interval bands in dashed curves.

Figure 4.14 depicts the estimated null distribution of test statistic $T \equiv r_K T_0$ compared to a Chi-squared distribution with degree of freedom 20.51. The null distribution of test statistic T are estimated through both parametric bootstrap and nonparametric bootstrap described above. From the figure it shows that under null hypothesis, test statistic T has a Chi-squared distribution.

Figure 4.15 compares the estimated null distribution of test statistic T for five different

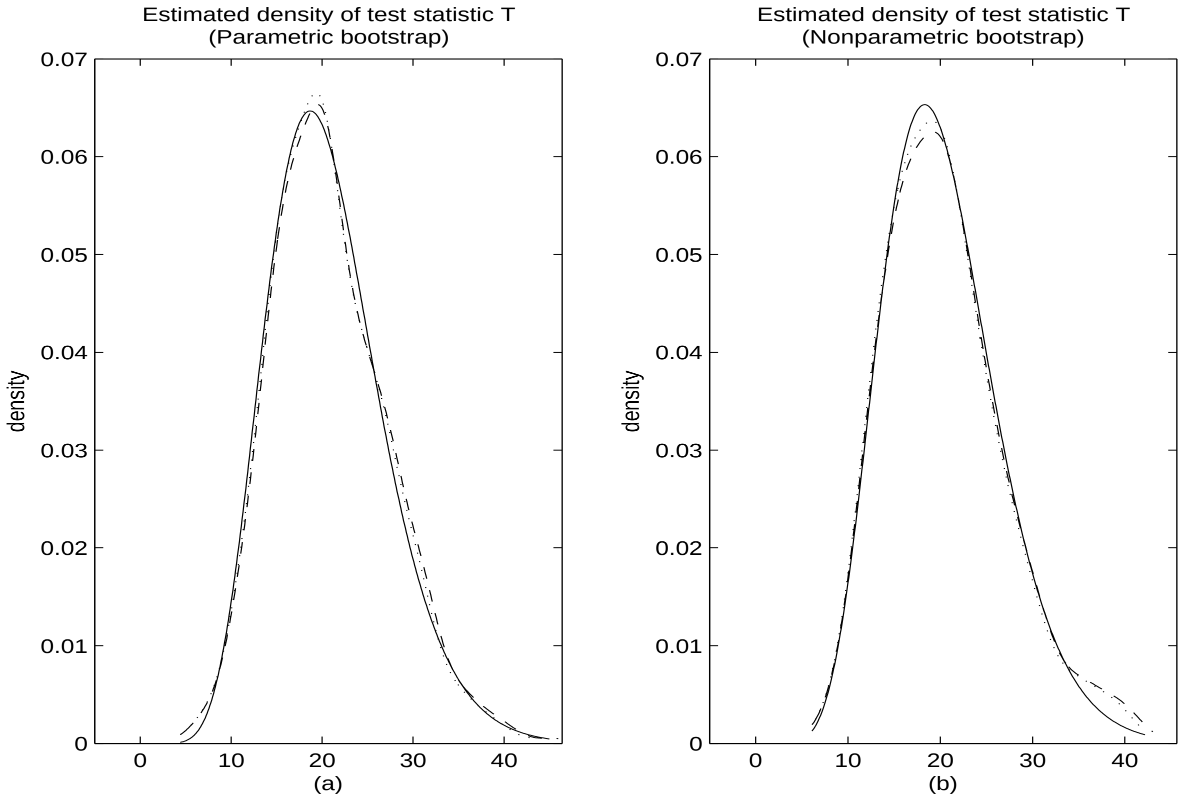

Figure 4.14: The estimated density of test statistic T by Monte Carlo Simulation. The dotted curve is the estimated density of generalized likelihood ratio test statistic T; the dash-dotted curve is the estimated density of generalized F-test statistic T_0. The solid curve is the density of Chi-squared distribution with degree of freedom 20.51.

sets

$$\begin{aligned}\{\gamma_{j0}\} \quad &= \quad \{\ (\beta_1,\ \beta_2),\ (\beta_1 - 2\cdot\mathrm{std}(\widehat{\beta}_1),\ \beta_2),\ (\beta_1 + 2\cdot\mathrm{std}(\widehat{\beta}_1),\ \beta_2),\\ &\qquad (\beta_1,\ \beta_2 - 2\cdot\mathrm{std}(\widehat{\beta}_2)\ ),\ (\beta_1,\ \beta_2 + 2\cdot\mathrm{std}(\widehat{\beta}_2)\ )\ \}\end{aligned}$$

where $\beta_j = E\{\beta_j(u)\}$ for $j = 1, 2$, and $\mathrm{std}(\widehat{\beta}_j)$ is the standard deviation of $\widehat{\beta}_j$ for $j = 1, 2$. Again, Figure 4.15 shows that the null distribution does not depend on the specific value of $\boldsymbol{\beta}$.

To examine the power of the proposed hypothesis test procedure, we consider the

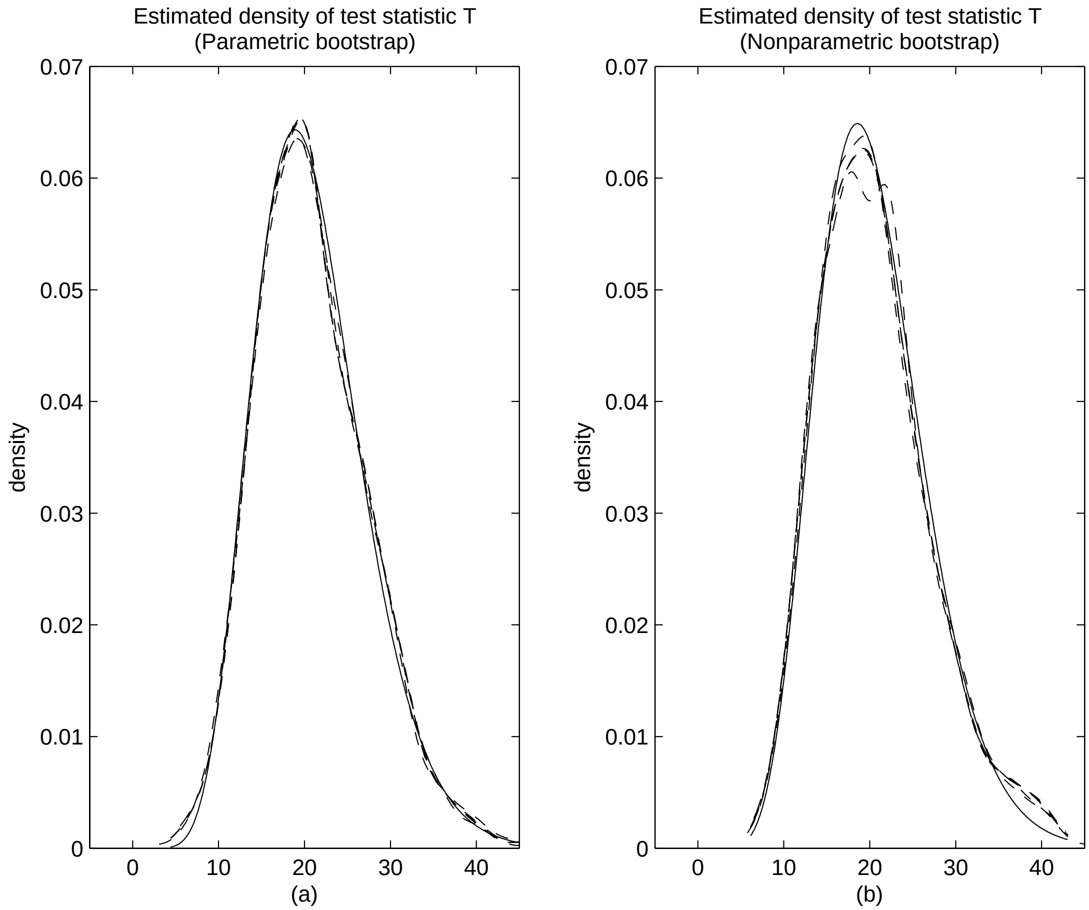

Figure 4.15: The estimated densities of test statistic T by Monte Carlo Simulation, using 5 different coefficient values. The five dashed curves are estimated densities. The solid curve is Chi-squared density with degree of freedom 20.51.

alternative hypothesis

$$H_1 : \beta_j(u, \delta) = (1 - \delta)\gamma_{j0} + \delta\beta_j^0(u), \qquad j = 1, 2,$$

where $\beta_1^0(u) = \sin(\pi \cdot u)$ and $\beta_2^0(u) = \sin(4\pi \cdot (u - 1/8))$ as defined previously in the simulation section for model estimation, and $\gamma_{j0} = E\{\beta_j^0(u)\}$. We then evaluate the power functions of the proposed test under a sequence of alternative models that are indexed by δ, The power functions are depicted in Figure 4.16. In Figure 4.16, we plot five power functions at five different significance levels: 0.50, 0.25, 0.10, 0.05, and 0.01, based on 1000 simulations

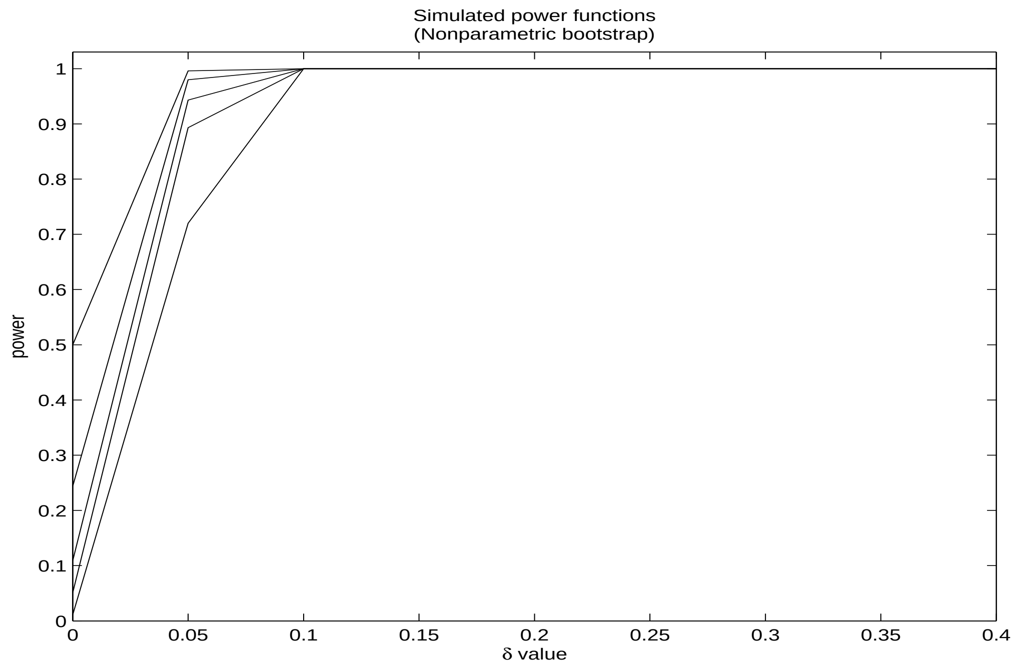

Figure 4.16: The power functions at five different significance levels: 0.5, 0.25, 0.10, 0.05, and 0.01.

for sample size n=1000. The alternative hypothesis is chosen such that when $\delta = 0$, the alternative collapses into the null hypothesis. In particular, the powers at $\delta = 0$ for the five significance levels are 0.505, 0.248, 0.105, 0.054, and 0.011. This shows that the nonparametric bootstrap method provides the correct levels of test. Also, as $\delta = 0$ value increases, the signal of varying-coefficient carried in alternative hypothesis is amplified. Therefore, the power should increase as the value of δ increases. The results shown in figure 4.16 exhibit that the powers increase very rapidly as δ increases, which justifies that our proposed testing procedures works well.

4.3 Example 3

In this section, we consider the following model

$$Y = \beta_1(u) - \frac{\beta_2(u)\ x}{x + \beta_3(u)} + \varepsilon \tag{4.10}$$

where $\beta(u) = \{\beta_1(u),\ \beta_2(u),\ \beta_3(u)\}$ with $\beta_1(u) = 7 + \exp(u-1)$, $\beta_2(u) = 10 + \sin(2\pi u)$ and $\beta_3(u) = 9 + 4(u-0.5)^2$, X is a log-normal random variable, of which the logarithm is normal with mean 2 and variance 3^2, and u is uniformly distributed on (0,1), independent of X. The coefficient functions $\beta_1(u)$, $\beta_2(u)$ and $\beta_3(u)$ are selected to represent typical functions, namely, $\beta_1(u)$ is a monotonic function, $\beta_2(u)$ is a one-mode function, and $\beta_3(u)$ is a two-mode function. The random error ε follows $N(0,1)$. This model is referred to as *ecology* model. In this example, we consider three sample sizes $n = 500, 1000$ and 2000.

The local least squares function is

$$\ell(\mathbf{a}, \mathbf{b}) = \frac{1}{2}\sum_{i=1}^{n}[\ y_i - \{(a_1 +\ b_1(u_i - u_0)) + \frac{(a_2 +\ b_2(u_i-u_0))\ x_i}{x_i + (a_3 +\ b_3(u_i - u_0))}\}\ ]^2 K_h(u_i - u_0)$$

where $\mathbf{x} = \{1,\ x\}$ and $\mathbf{a} = (a_1,\ a_2,\ a_3)^T$ and $\mathbf{b} = (b_1,\ b_2,\ b_3)^T$. For a given initial value $(\mathbf{a}_0^T, \mathbf{b}_0^T)^T$ of $(\mathbf{a}^T, \mathbf{b}^T)^T$, we locally and linearly approximate

$$\begin{aligned} f\{\mathbf{x}, \mathbf{a} + \mathbf{b}(u-u_0)\} \ \approx\ & \{(a_1 +\ b_1(u-u_0)) + \frac{(a_2 +\ b_2(u-u_0))\ x}{x + (a_3 +\ b_3(u-u_0))}\} + \\ & \{(\mathbf{a}-\mathbf{a}_0) + (\mathbf{b}-\mathbf{b}_0)(u-u_0)\}^T \begin{pmatrix} 1 \\ \\ \dfrac{x}{x + (a_3 +\ b_3(u-u_0))} \\ \\ -\dfrac{(a_2 +\ b_2(u-u_0))\ x}{\{x + (a_3 +\ b_3(u-u_0))\}^2} \end{pmatrix} \end{aligned}$$

With this approximation, we can use an iterated least squares algorithm to search the solution of (4.9). Specifically, at the k-th step during the course of iteration, the current value for $\mathbf{a}$

and $\mathbf{b}$ is $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$. Denote

$$\begin{aligned} y_{i,k} &= y_i - f\{\mathbf{x}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0)\} + \{\mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0)\} f'\{\mathbf{x}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0)\} \\ &= y_i - \{(a_1^{(k)} + \ b_1^{(k)}(u_i - u_0)) + \frac{(a_2^{(k)} + \ b_2^{(k)}(u_i - u_0))\ x_i}{x_i + (a_3^{(k)} + \ b_3^{(k)}(u_i - u_0))}\} + \\ &\quad \{(\mathbf{a} - \mathbf{a}^{(k)}) + (\mathbf{b} - \mathbf{b}^{(k)})(u - u_0)\}^T \begin{pmatrix} 1 \\ \\ \dfrac{x_i}{x_i + (a_3^{(k)} + \ b_3^{(k)}(u_i - u_0))} \\ \\ -\dfrac{(a_2^{(k)} + \ b_2^{(k)}(u_i - u_0))\ x_i}{\{x_i + (a_3^{(k)} + \ b_3^{(k)}(u_i - u_0))\}^2} \end{pmatrix}, \end{aligned}$$

$\mathbf{y}_k = (y_{1,k}, \cdots, y_{n,k})^T$, and

$$F_k = \begin{pmatrix} f'\{\mathbf{x}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_1 - u_0)\}, & \cdots & f'\{\mathbf{x}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_n - u_0)\} \\ (u_1 - u_0) f'\{\mathbf{x}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_1 - u_0)\}, & \cdots & (u_n - u_0) f'\{\mathbf{x}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_n - u_0)\}^T \end{pmatrix}$$

where

$$f'\{\mathbf{x}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0)\} = \begin{pmatrix} 1 \\ \\ \dfrac{x_i}{x_i + (a_3^{(k)} + \ b_3^{(k)}(u_i - u_0))} \\ \\ -\dfrac{(a_2^{(k)} + \ b_2^{(k)}(u_i - u_0))\ x_i}{\{x_i + (a_3^{(k)} + \ b_3^{(k)}(u_i - u_0))\}^2} \end{pmatrix}$$

We iteratively update $(\mathbf{a}, \mathbf{b})^T$ by

$$\begin{pmatrix} \mathbf{a}^{(k+1)} \\ \mathbf{b}^{(k+1)} \end{pmatrix} = (F_k^T W F_k)^{-1} F_k^T W \mathbf{y}_k,$$

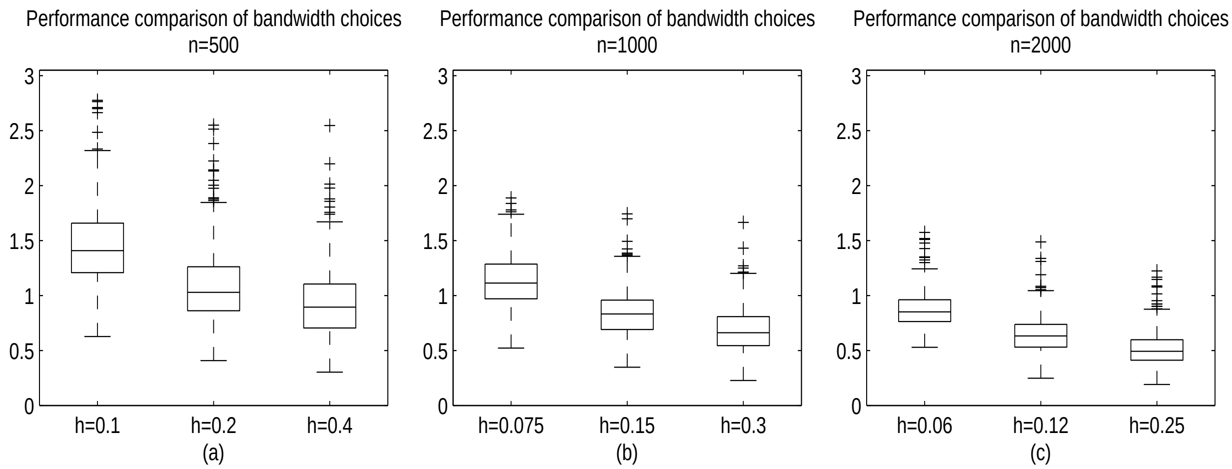

Figure 4.17: Boxplots of RASE values for Example 3. For $n = 500$, bandwidth $h = 0.10$, 0.20, and 0.40; for $n = 1000$, $h = 0.075$, 0.15, and 0.30; for $n = 2000$, $h = 0.06$, 0.12, and 0.25.

until it converges. When the algorithm converges, the solution is satisfied with

$$\ell'(\mathbf{a}, \mathbf{b}) = 0.$$

Denote the resulting estimate of $(\mathbf{a}, \mathbf{b})^T$ by $(\widehat{\mathbf{a}}, \widehat{\mathbf{b}})^T$. Then

$$\widehat{\boldsymbol{\beta}}(u_0) = \widehat{\mathbf{a}}, \quad \text{and} \quad \widehat{\boldsymbol{\beta}}'(u_0) = \widehat{\mathbf{b}}.$$

Figure 4.17(a) depicts the marginal distribution of the overall RASE values using sample size n=500 for bandwidth h=0.10, 0.20, and 0.40. Figure 4.18(a) through 4.18(f) show the estimate of the coefficient functions from a typical sample. The typical sample is selected such that its RASE value is the median of the RASE values from 400 simulated samples.

Similar to figure 4.17(a) and 4.18, figure 4.17(b) depicts the marginal distribution of the overall RASE values using sample size n=1000 for bandwidth h=0.075, 0.15, and 0.30. Figure 4.19(a) through 4.19(f) show the estimate of the coefficient functions from a typical

Table 4.5: Summary of Simulation Output for Example 3

n	h	$\mu_{(RASE)}$	$\sigma_{(RASE)}$
	0.10	1.4574	0.3543
500	0.20	1.0902	0.3454
	0.40	0.9240	0.3116
	0.075	1.1362	0.2335
1000	0.15	0.8454	0.2133
	0.30	0.6891	0.1966
	0.06	0.8696	0.1623
2000	0.12	0.6455	0.1632
	0.25	0.5146	0.1554

sample. The typical sample is selected such that its RASE value is the median of the RASE values from 400 simulated samples.

In the same fashion, figure 4.17(c) depicts the marginal distribution of the overall RASE values using sample size n=2000 for bandwidth h=0.06, 0.12, and 0.25. Figure 4.20(a) through 4.20(f) show the estimate of the coefficient functions from a typical sample. The typical sample is selected such that its RASE value is the median of the RASE values from 400 simulated samples.

The simulation results are summarized in Table 4.5. In Table 4.5, μ and σ denote the mean and standard deviation of the RASE values in 400 simulations. For any fixed sample size level, the RASE values generally increase as the bandwidth h increases. This is because a larger bandwidth value implies a wider neighborhood and thus more local data, which causes a greater bias. This can also be verified from the graphical comparisons of estimated coefficient function at different bandwidth values at a fixed sample size level. In the plots, we observe a pattern that a larger bandwidth h leads to a greater estimation bias.

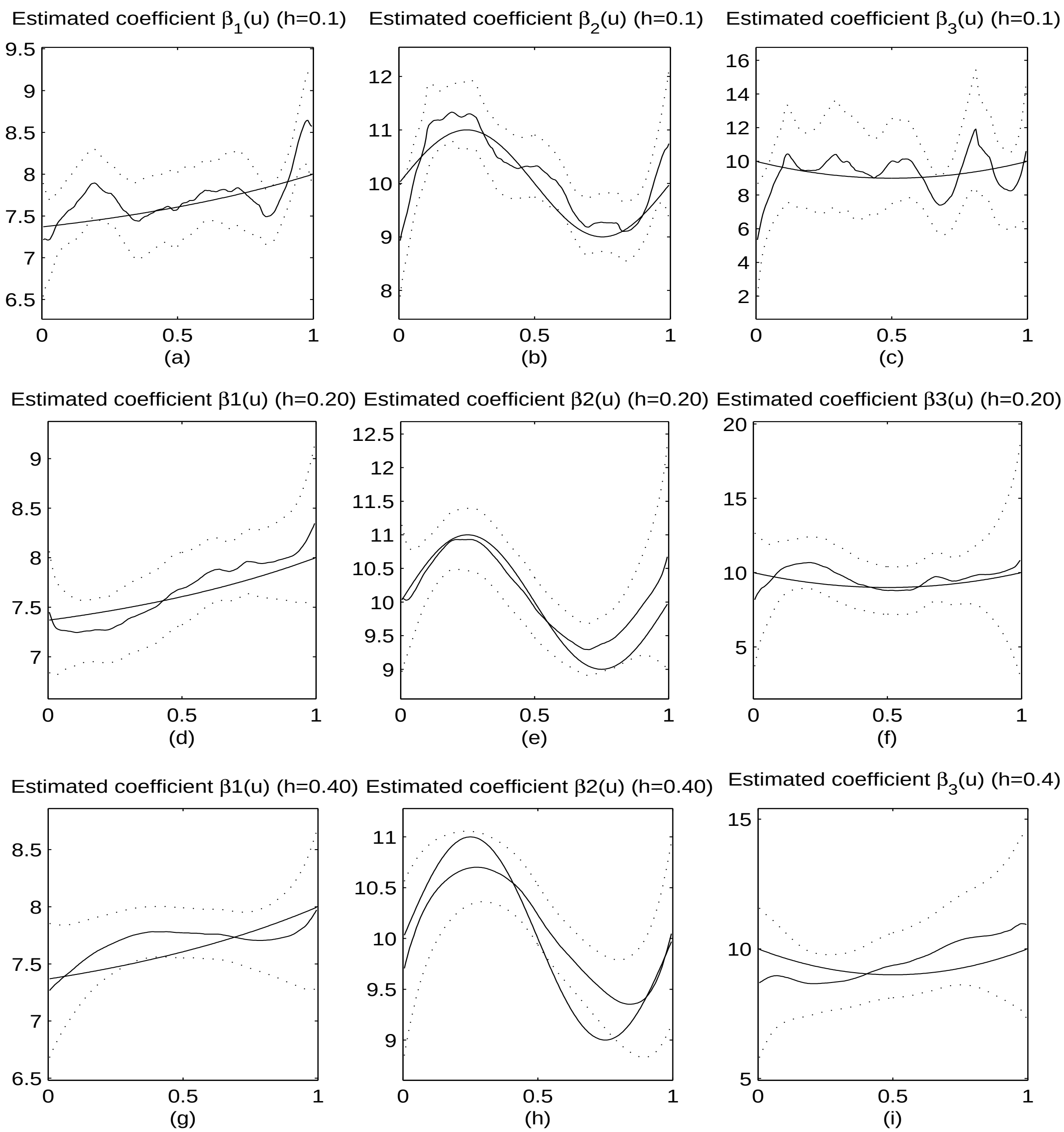

Figure 4.18: Estimated Coefficients for Example 3 with $n = 500$. (a)-(b) are plots for bandwidth $h = 0.10$; (c)-(d) are $h = 0.20$; (e)-(f) are plots for $h = 0.40$. The fluctuated solid curves are estimated coefficient functions for the exponential model; the dotted curves are pointwise 95% confidence intervals obtained using standard error formulas; the one-mode and two-mode solid curves are the true coefficient functions.

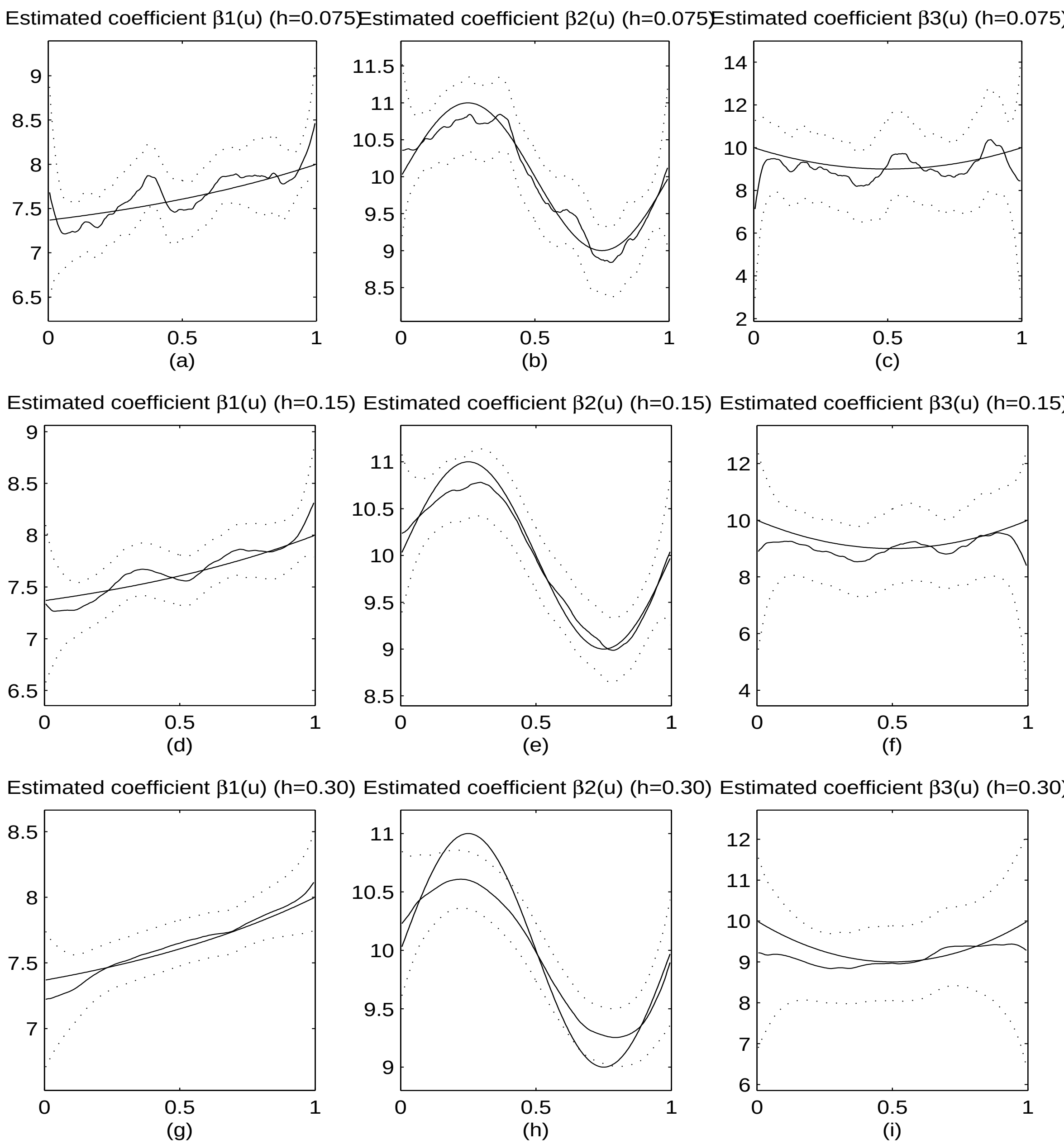

Figure 4.19: Estimated Coefficients for Example 3 with $n = 1000$. Caption is similar to that in 4.18

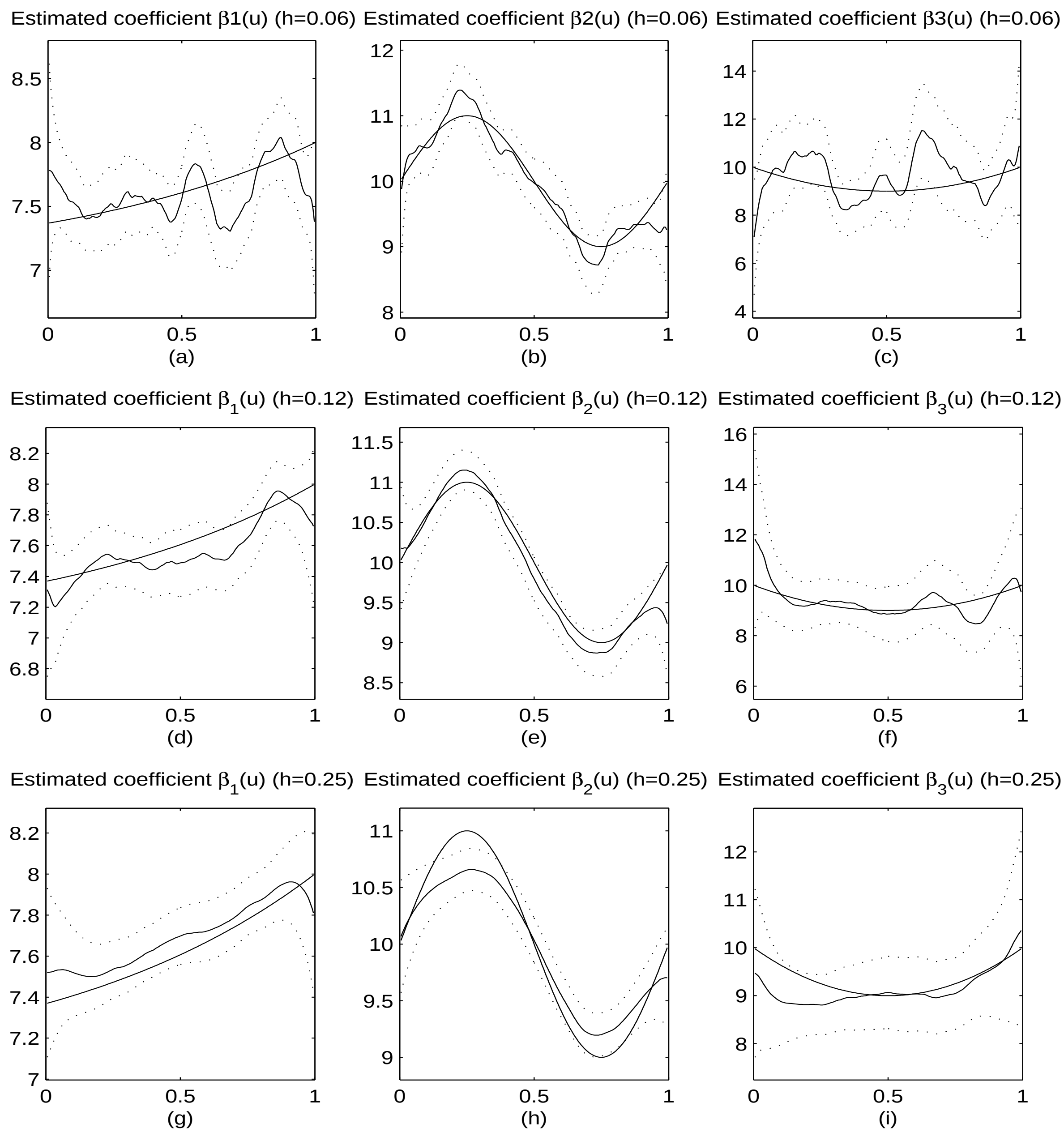

Figure 4.20: Estimated Coefficients for Example 3 with $n = 1000$. Caption is similar to that in 4.18

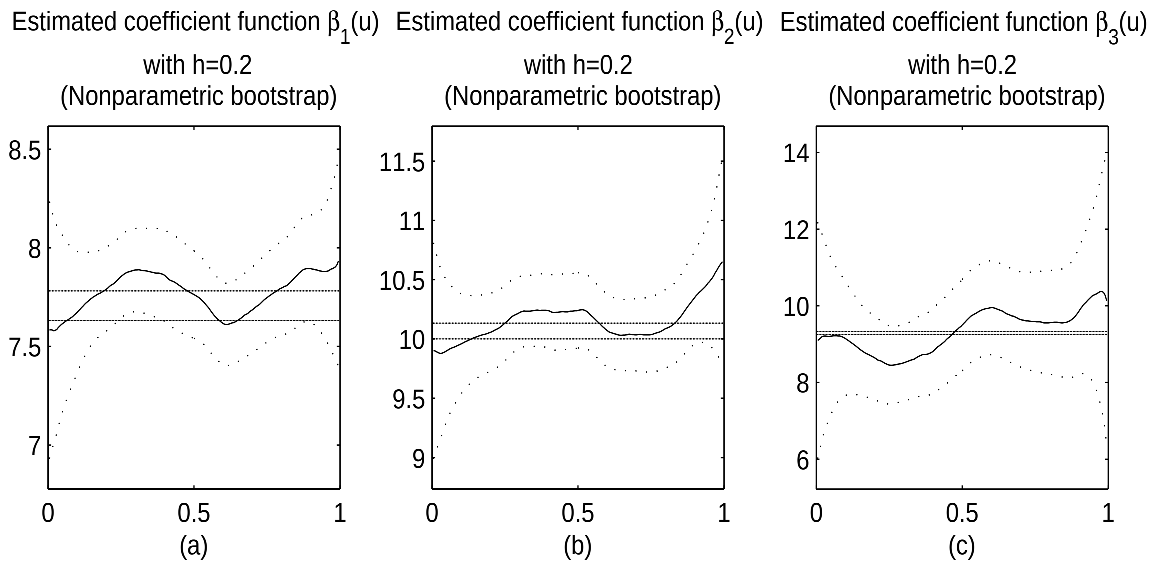

Figure 4.21: The estimated coefficient functions under null hypothesis. The fluctuated solid curves are estimated coefficient functions for Example 3; the dotted curves are pointwise 95% confidence intervals obtained using standard error formulas; the solid lines are true coefficient functions; the dashed lines are the true function coefficients.

We have also tested the accuracy of the proposed standard error formula. The results are summarized in Table 4.6, in which the notation is the same as that in Table 4.2. From Table 4.6, the proposed standard error formula performs well.

We next demonstrate the GLRT for hypothesis:

$$H_0 : \beta_j(u) = \gamma_{j0}, \quad j = 1 \cdots, p$$

versus

$$H_1 : \beta_j(u) \neq \gamma_{j0}, \quad \text{for at least one j}$$

where γ_{j0} is unknown constant. The null hypothesis implies that the $\beta_j(u)$ is a constant. As in Examples 1 and 2, the null distribution of the GLRT can be obtained by either parametric bootstrap or nonparametric bootstrap. Here we estimate the distribution of test statistic T under H_0 using $n = 1000$ via 1000 Monte Carlo simulations.

Table 4.6: Standard Deviation and Standard Error for Example 3

			$\hat{\beta}_1(u)$		$\hat{\beta}_2(u)$		$\hat{\beta}_2(u)$	
n	h	u	SD	$SD_a(SD_{std})$	SD	$SD_a(SD_{std})$	SD	$SD_a(SD_{std})$
		0.1	0.2196	0.2178 (0.0371)	0.3120	0.2930 (0.0384)	1.3063	1.1676 (0.2260)
		0.3	0.2374	0.2239 (0.0374)	0.3128	0.2981 (0.0374)	1.1592	1.0745 (0.2178)
	0.10	0.5	0.2396	0.2258 (0.0409)	0.3132	0.2931 (0.0383)	1.2624	1.1384 (0.2294)
		0.7	0.2286	0.2167 (0.0372)	0.3113	0.2898 (0.0359)	1.2979	1.2857 (0.2544)
		0.9	0.2220	0.2162 (0.0380)	0.2939	0.2975 (0.0379)	1.3460	1.3053 (0.2597)
		0.1	0.1870	0.1916 (0.0307)	0.2610	0.2569 (0.0306)	1.0997	1.0228 (0.1803)
		0.3	0.1671	0.1595 (0.0182)	0.2144	0.2126 (0.0181)	0.8086	0.7647 (0.0992)
500	0.20	0.5	0.1663	0.1608 (0.0189)	0.2285	0.2100 (0.0178)	0.8481	0.8093 (0.1103)
		0.7	0.1607	0.1566 (0.0182)	0.2185	0.2084 (0.0179)	0.8825	0.8973 (0.1239)
		0.9	0.1950	0.1882 (0.0301)	0.2589	0.2595 (0.0297)	1.1655	1.1135 (0.2072)
		0.1	0.1773	0.1878 (0.0283)	0.2537	0.2543 (0.0280)	1.0152	0.9808 (0.1496)
		0.3	0.1186	0.1183 (0.0097)	0.1642	0.1616 (0.0098)	0.6486	0.5991 (0.0577)
	0.40	0.5	0.1184	0.1129 (0.0086)	0.1638	0.1496 (0.0082)	0.5767	0.5814 (0.0511)
		0.7	0.1185	0.1167 (0.0097)	0.1617	0.1594 (0.0095)	0.6575	0.6690 (0.0683)
		0.9	0.1940	0.1841 (0.0275)	0.2531	0.2557 (0.0269)	1.1587	1.0850 (0.1835)
		0.1	0.1908	0.1795 (0.0269)	0.2642	0.2417 (0.0265)	0.9936	0.9487 (0.1539)
		0.3	0.1798	0.1844 (0.0267)	0.2426	0.2439 (0.0254)	0.8987	0.8661 (0.1292)
	0.075	0.5	0.1908	0.1831 (0.0256)	0.2457	0.2432 (0.0255)	0.9711	0.9360 (0.1429)
		0.7	0.1962	0.1812 (0.0254)	0.2563	0.2421 (0.0253)	1.1230	1.0478 (0.1753)
		0.9	0.1870	0.1812 (0.0266)	0.2637	0.2429 (0.0269)	1.0663	1.0569 (0.1642)
		0.1	0.1395	0.1382 (0.0160)	0.2009	0.1867 (0.0153)	0.7418	0.7360 (0.0913)
		0.3	0.1295	0.1302 (0.0133)	0.1687	0.1728 (0.0124)	0.6340	0.6138 (0.0659)
1000	0.15	0.5	0.1342	0.1301 (0.0127)	0.1751	0.1727 (0.0122)	0.6905	0.6625 (0.0726)
		0.7	0.1394	0.1295 (0.0123)	0.1867	0.1725 (0.0121)	0.7614	0.7378 (0.0815)
		0.9	0.1485	0.1392 (0.0150)	0.1991	0.1870 (0.0151)	0.7986	0.8077 (0.0962)
		0.1	0.1329	0.1341 (0.0155)	0.1931	0.1821 (0.0150)	0.6985	0.7089 (0.0816)
		0.3	0.0918	0.0921 (0.0065)	0.1229	0.1240 (0.0061)	0.4615	0.4510 (0.0328)
	0.30	0.5	0.0933	0.0922 (0.0064)	0.1228	0.1223 (0.0057)	0.4859	0.4722 (0.0362)
		0.7	0.0978	0.0921 (0.0059)	0.1317	0.1238 (0.0058)	0.5339	0.5176 (0.0396)
		0.9	0.1451	0.1349 (0.0147)	0.1900	0.1822 (0.0152)	0.7811	0.7807 (0.0955)
		0.1	0.1443	0.1428 (0.0155)	0.1951	0.1923 (0.0147)	0.7639	0.7480 (0.0881)
		0.3	0.1378	0.1442 (0.0157)	0.1979	0.1925 (0.0158)	0.6756	0.6892 (0.0797)
	0.06	0.5	0.1535	0.1455 (0.0179)	0.1942	0.1921 (0.0162)	0.7310	0.7296 (0.0866)
		0.7	0.1565	0.1443 (0.0160)	0.2036	0.1926 (0.0161)	0.8840	0.8315 (0.1101)
		0.9	0.1434	0.1420 (0.0160)	0.1768	0.1904 (0.0153)	0.9075	0.8325 (0.1030)
		0.1	0.0998	0.1034 (0.0077)	0.1362	0.1391 (0.0072)	0.5576	0.5437 (0.0466)
		0.3	0.1018	0.1026 (0.0083)	0.1398	0.1366 (0.0077)	0.4913	0.4886 (0.0404)
2000	0.12	0.5	0.1092	0.1035 (0.0082)	0.1387	0.1362 (0.0078)	0.5195	0.5171 (0.0438)
		0.7	0.1073	0.1026 (0.0083)	0.1441	0.1364 (0.0078)	0.6100	0.5817 (0.0519)
		0.9	0.1018	0.1033 (0.0081)	0.1302	0.1382 (0.0078)	0.6147	0.6012 (0.0517)
		0.1	0.0929	0.0955 (0.0071)	0.1248	0.1291 (0.0068)	0.5234	0.5064 (0.0423)
		0.3	0.0661	0.0715 (0.0039)	0.0963	0.0954 (0.0034)	0.3259	0.3463 (0.0190)
	0.25	0.5	0.0745	0.0717 (0.0036)	0.0938	0.0946 (0.0035)	0.3628	0.3626 (0.0208)
		0.7	0.0739	0.0714 (0.0039)	0.0924	0.0952 (0.0036)	0.4190	0.3992 (0.0240)
		0.9	0.0953	0.0956 (0.0074)	0.1230	0.1284 (0.0070)	0.5649	0.5539 (0.0487)

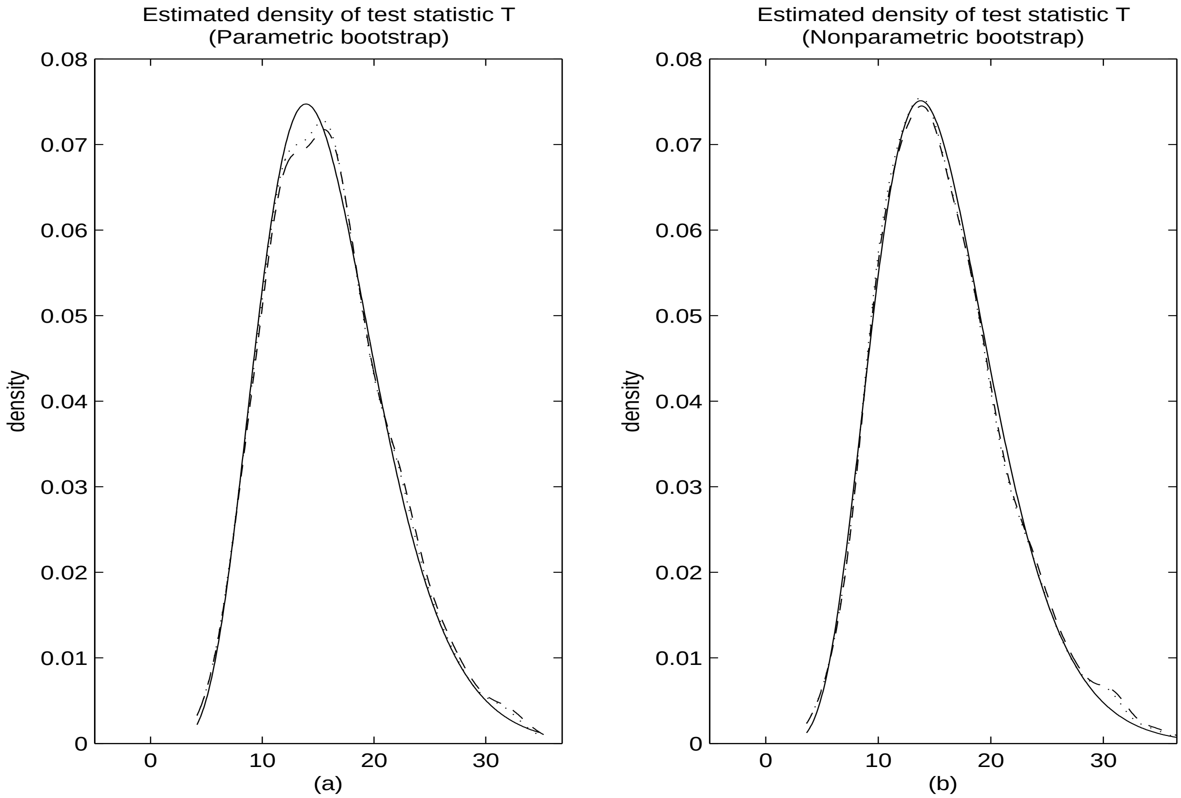

Figure 4.22: The estimated density of test statistic T by Monte Carlo Simulation. The dotted curve is the estimated density of generalized likelihood ratio test statistic T; the dash-dotted curve is the estimated density of generalized F-test statistic T_0. The solid curve is the density of chi-squared distribution with degree of freedom 15.90

Figure 4.21 illustrates the nonparametric estimated coefficient functions under null hypothesis. The dashed lines are the true function coefficients. The solid lines are the estimated function coefficients within the null space of constant coefficients; while the solid curves are the nonparametric estimated coefficient functions within the alternative space of varying-coefficients, with pointwise 95% confidence interval bands in dashed curves.

Figure 4.22 depicts the estimated null distribution of test statistic $T \equiv r_K T_0$ compared to a Chi-squared distribution with degree of freedom by Wilks phenomenon. The null distribution of test statistic T are estimated through both parametric bootstrap and nonparametric bootstrap described above. From the figure it shows that under null hypothesis, test statistic T has a chi-squared distribution with degree of freedom 15.90.

Figure 4.23 compares the estimated null distribution of test statistic T for five different sets

$$\begin{aligned}
\{\gamma_{j0}\} \;\; = \;\; & \{\ (\beta_1,\ \beta_2,\ \beta_3),\ (\beta_1 - 2\cdot\mathrm{std}(\widehat{\beta}_1),\ \beta_2,\ \beta_3),\ (\beta_1 + 2\cdot\mathrm{std}(\widehat{\beta}_1),\ \beta_2,\ \beta_3), \\
& (\beta_1,\ \beta_2 - 2\cdot\mathrm{std}(\widehat{\beta}_2)\ ,\ \beta_3),\ (\beta_1,\ \beta_2 + 2\cdot\mathrm{std}(\widehat{\beta}_2)\ ,\ \beta_3), \\
& (\beta_1,\ \beta_2\ ,\ \beta_3 - 2\cdot\mathrm{std}(\widehat{\beta}_3)\ ),\ (\beta_1,\ \beta_2\ ,\ \beta_3 + 2\cdot\mathrm{std}(\widehat{\beta}_2)\ )\}
\end{aligned}$$

where $\beta_j = E\{\beta_j(u)\}$ for $j = 1, 2, 3$, and $\mathrm{std}(\widehat{\beta}_j)$ is the standard deviation of $\widehat{\beta}_j$ for $j = 1, 2, 3$. The null distribution of test statistic T are estimated through both parametric bootstrap and nonparametric bootstrap described above. Figure 4.23 provides us an empirical justification for that the null distribution of test statistic T follows Chi-squared distribution with degree of freedom and is independent of the specific values of $\boldsymbol{\beta}$.

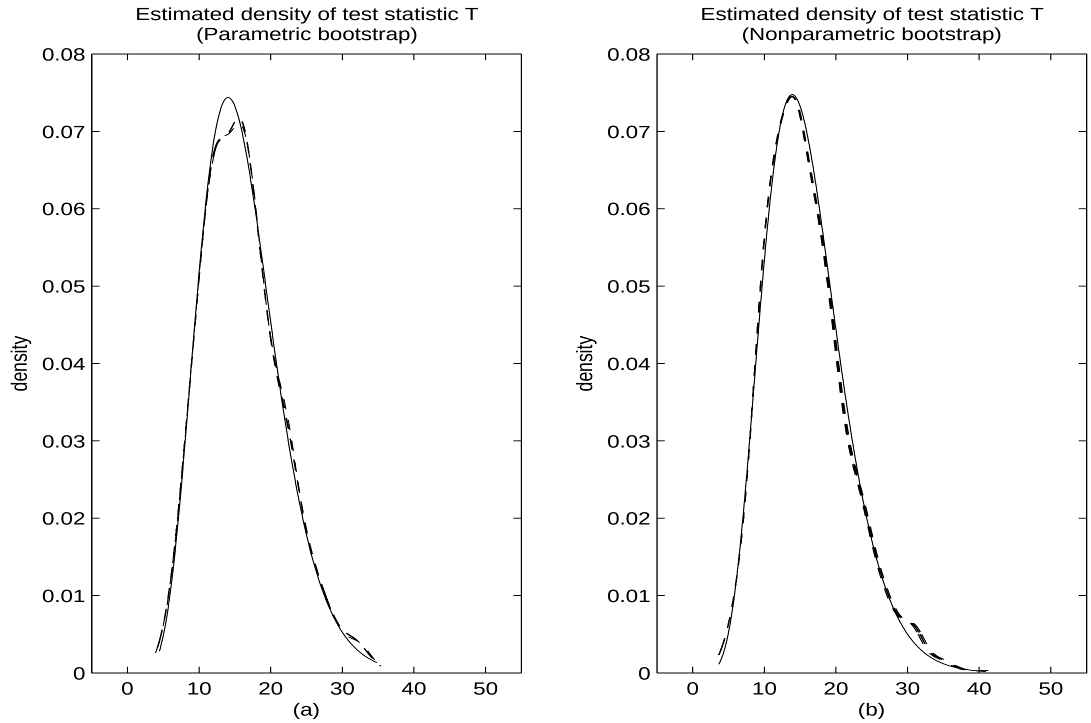

Figure 4.23: The estimated densities of test statistic T by Monte Carlo Simulation, using 5 different coefficient values. The five dashed curves are estimated densities. The solid curve is Chi-squared density with degree of freedom 15.90.

To examine the power of the proposed hypothesis test procedure, we consider the alternative hypothesis

$$H_1 : \beta_j(u, \delta) = (1-\delta)\gamma_{j0} + \delta\beta_j^0(u), \qquad j = 1, 2, (0 \leq \delta \leq 0.8)$$

where $\beta_1^0(u) = 7 + exp(u-1)$, $\beta_2^0 = 10 + sin(2 \cdot \pi \cdot u)$ and $\beta_3^0(u) = 9 + 4 \cdot (u - 0.5)^2$ as defined previously in the simulation section for model estimation, and $\gamma_{j0} = E\{\beta_j^0(u)\}$. The power functions are depicted in Figure 4.24. In Figure 4.24, we plot five power functions at five different significance levels: 0.50, 0.25, 0.10, 0.05, and 0.01, based on 1000 simulations for sample size n=1000. The alternative hypothesis is chosen such that when $\delta = 0$, the alternative collapses into the null hypothesis. In particular, the powers at $\delta = 0$ for the five significance levels are 0.498, 0.258, 0.107, 0.053, and 0.011. This shows that the nonpara-

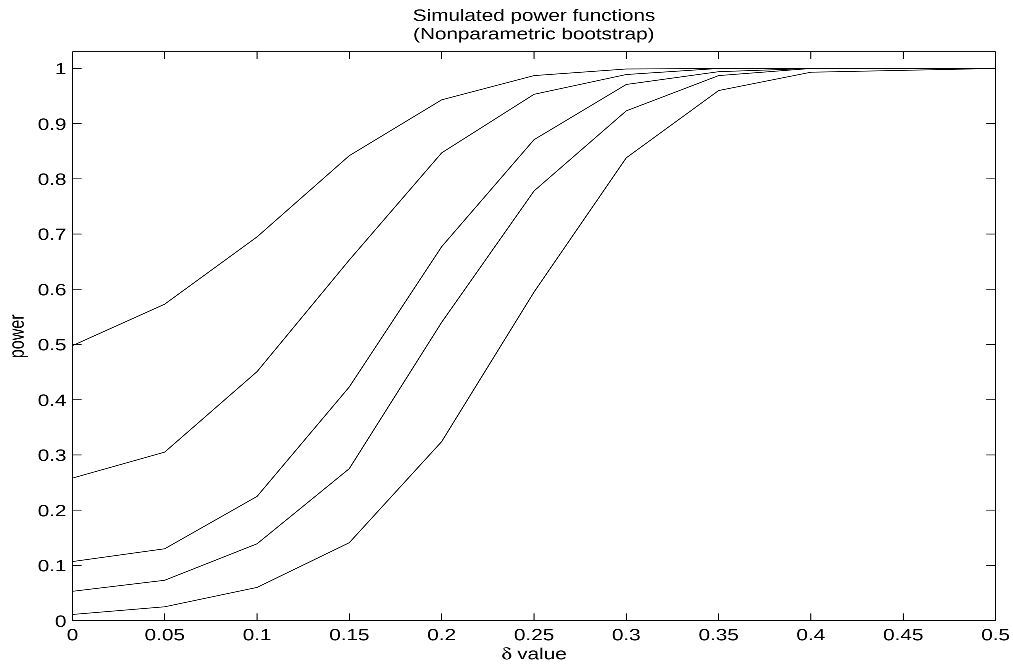

Figure 4.24: The simulated power functions at five different significance levels: 0.5, 0.25, 0.10, 0.05, and 0.01.

metric bootstrap method provides the correct levels of test. Also, as $\delta = 0$ value increases, the signal of varying-coefficient carried in alternative hypothesis is amplified. Therefore, the power should increase as the value of δ increases. The results shown in figure 4.24 exhibit that the powers increase very rapidly as δ increases, which justifies that our proposed testing procedures works well.

Chapter 5

Application to Ecological Data

In this chapter, we illustrate the proposed methodology by an analysis of a ecological data set, briefly described in Chapter 1. Specifically, we first apply the proposed estimation procedures for the data, and further conduct hypothesis testing whether the coefficient functions really change over temperature.

5.1 Data and Model

As mentioned in Chapter 1, the data set was collected within the AmeriFlux network during summer growth seasons (from June 1 to August 31) of years 1993 to 1995 at the Walker Branch Watershed Site in eastern Tennessee (35.96^{o}N, 84.29^{o}W). It is known that sunlight intensity affects the rate of photosynthesis in an ecosystem. Since leaves absorb carbon dioxide (CO_2) during the course of photosynthesis, the *Net Ecosystem Exchange* of CO_2, denoted by NEE, is used to measure the level of photosynthetic activity in a natural ecosystem. Photosynthetic rate as measured by NEE is dependent on the amount of *Photosynthetically Active Radiation* available to an ecosystem, denoted by PAR.

Based on empirical studies, scientists believe that the relationship between NEE and PAR is nonlinear and can be characterized by the following model

$$\text{NEE} = R - \frac{\beta_1 \text{PAR}}{\text{PAR} + \beta_2} + \varepsilon, \tag{5.1}$$

where ε is random error with zero mean, and R, β_1 and β_2 are unknown parameters. The physical interpretation for the unknown parameters are as follows: R is the dark respiration

Table 5.1: Percentiles for the Ecological Data

Percentiles	NEE	PAR	Temp
0^{th}	-37.9000	0	5.00
10^{th}	-21.0000	103	14.80
30^{th}	-15.5000	324	18.00
50^{th}	-10.6000	683	20.30
70^{th}	-4.8000	1114	22.40
90^{th}	0.5000	1551	25.40
100^{th}	21.9000	1955	32.80

rate, β_1 is the light-saturated net photosynthetic rate, and β_1/β_2 is the apparent quantum yield. From the contour plot in Figure 1.1, we consider an alternative model that allows the unknown model parameters changing over temperature and explore whether the alternative model can better characterize the relationship between NEE and PAR. Thus, we fit the data by a *nonlinear varying-coefficient model*

$$\text{NEE} = R(T) - \frac{\beta_1(T)\text{PAR}}{\text{PAR} + \beta_2(T)} + \varepsilon, \tag{5.2}$$

where T stands for temperature. This model takes into consideration of the dynamic feature of variable temperature. For simplicity of notation, let y denote the response variable NEE, u denote temperature and $\mathbf{x}$ denote PAR, then model (5.2) takes the form of

$$Y = \beta_1(u) - \frac{\beta_2(u)\ x}{x + \beta_3(u)} + \varepsilon \tag{5.3}$$

Let us begin with some exploratory data analysis for this data set. Table 5.1 displays some percentiles of the ecological data, Figure 5.2 depicts the histogram of the data, and Figure 5.1 depicts boxplots of the three variables by year. It appears that the distributions of NEE, PAR and Temp are roughly similar across year 1993, 1994, and 1995. Thus, we will use all three year data in our analysis.

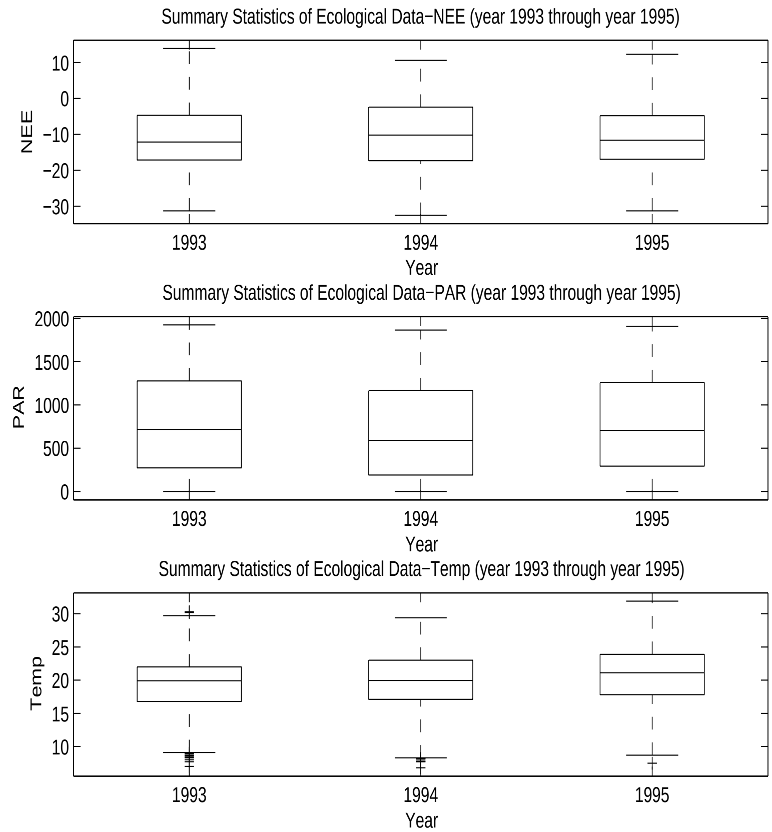

Figure 5.1: Poxplots for the Ecological Data of year 1993 through year 1995. (a) the boxplot for variable NEE; (b) the boxplot for variable PAR; (c) the boxplot for temperature.

5.2 Estimation of Coefficient Functions

We have conducted simulation for model 5.3. Let us describe the estimation procedure in details. Using Taylor's expansion in a neighborhood of given u_0, for $j = 1, 2$,

$$\beta_j(u) \approx \beta_j(u_0) + \beta_j(u_0)(u - u_0) \equiv a_j + b_j(u - u_0).$$

Denote $\mathbf{a} = (a_1, a_2)^T$ and $\mathbf{b} = (b_1, b_2)^T$. Thus, we obtain a local linear regression estimator $(\widehat{\mathbf{a}}^T, \widehat{\mathbf{b}}^T)^T$ by minimizing the following local least squares function:

$$\ell(\mathbf{a}, \mathbf{b}) = \frac{1}{2} \sum_{i=1}^{n} [\ y_i - \{(a_1 + \ b_1(u_i - u_0)) + \frac{(a_2 + \ b_2(u_i - u_0))\ x_i}{x_i + (a_3 + \ b_3(u_i - u_0))}\}\ ]^2 K_h(u_i - u_0)$$

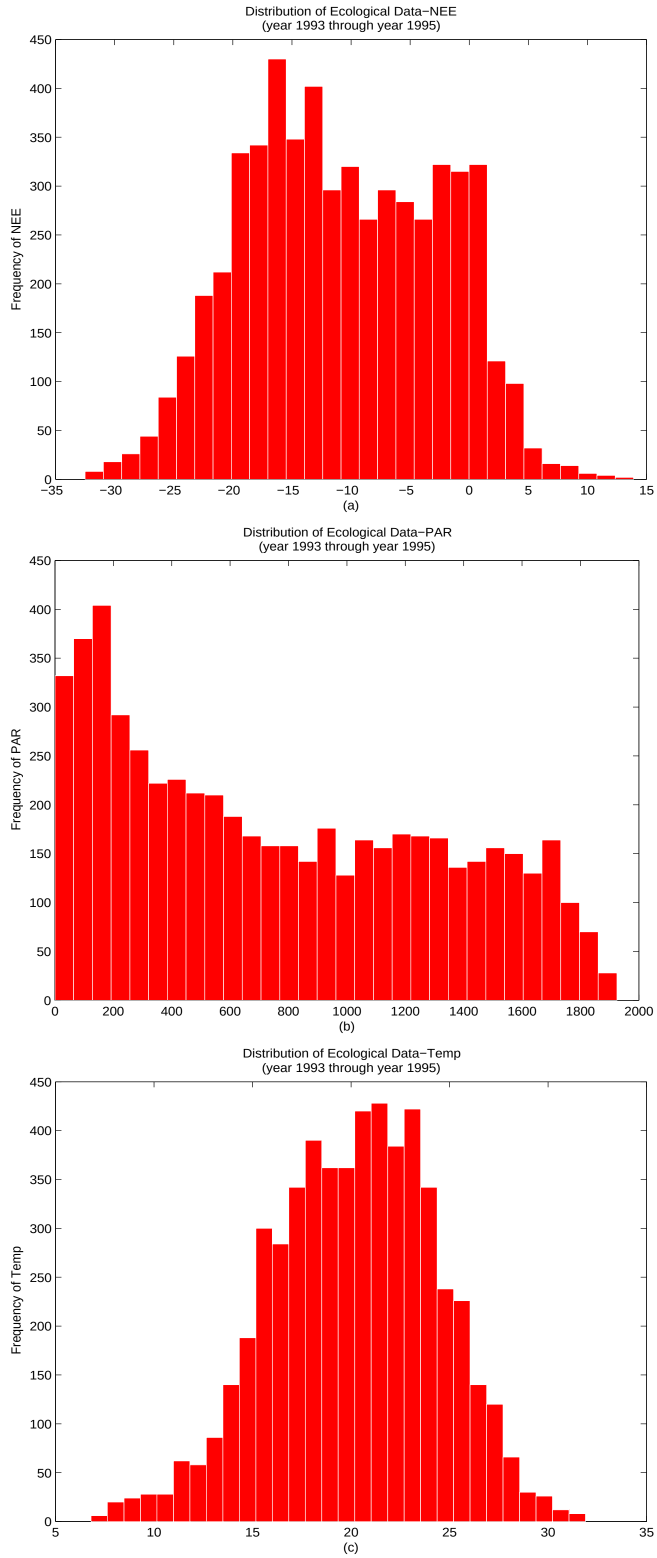

Figure 5.2: Histograms for the Ecological Data Set. (a) the histogram for NEE; (b) the histogram for PAR; (c) the histogram for temperature

where $\mathbf{x} = \{1,\ x\}$ and $\mathbf{a} = (a_1,\ a_2,\ a_3)^T$ and $\mathbf{b} = (b_1,\ b_2,\ b_3)^T$. To minimize the local least squares function, we locally and linearly approximate $f\{\mathbf{x}, \mathbf{a} + \mathbf{b}(u - u_0)\}$ by

$$\{(a_1 +\ b_1(u-u_0)) + \frac{(a_2 +\ b_2(u-u_0))\ x}{x + (a_3 +\ b_3(u-u_0))}\} +$$

$$\{(\mathbf{a}-\mathbf{a}_0) + (\mathbf{b}-\mathbf{b}_0)(u-u_0)\}^T \begin{pmatrix} 1 \\ \\ \dfrac{x}{x + (a_3 +\ b_3(u-u_0))} \\ \\ -\dfrac{(a_2 +\ b_2(u-u_0))\ x}{\{x + (a_3 +\ b_3(u-u_0))\}^2} \end{pmatrix}$$

for a given initial value $(\mathbf{a}_0^T, \mathbf{b}_0^T)^T$ of $(\mathbf{a}^T, \mathbf{b}^T)^T$, With this approximation, we can use the proposed iterated least squares algorithm for the local least squares function. Specifically, at the k-th step during the course of iteration, the current value for $\mathbf{a}$ and $\mathbf{b}$ is $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$. Denote

$$y_{i,k} \;=\; y_i - \{(a_1^{(k)} +\ b_1^{(k)}(u_i-u_0)) + \frac{(a_2^{(k)} +\ b_2^{(k)}(u_i-u_0))\ x_i}{x_i + (a_3^{(k)} +\ b_3^{(k)}(u_i-u_0))}\} +$$

$$\{(\mathbf{a}-\mathbf{a}^{(k)}) + (\mathbf{b}-\mathbf{b}^{(k)})(u-u_0)\}^T \begin{pmatrix} 1 \\ \\ \dfrac{x_i}{x_i + (a_3^{(k)} +\ b_3^{(k)}(u_i-u_0))} \\ \\ -\dfrac{(a_2^{(k)} +\ b_2^{(k)}(u_i-u_0))\ x_i}{\{x_i + (a_3^{(k)} +\ b_3^{(k)}(u_i-u_0))\}^2} \end{pmatrix}$$

and

$$F_k \;=\; \begin{pmatrix} f'\{\mathbf{x}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_1-u_0)\}, & \cdots & f'\{\mathbf{x}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_n-u_0)\} \\ (u_1-u_0)f'\{\mathbf{x}_1, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_1-u_0)\}, & \cdots & (u_n-u_0)f'\{\mathbf{x}_n, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_n-u_0)\}^T \end{pmatrix}$$

where

$$f'\{\mathbf{x}_i, \mathbf{a}^{(k)} + \mathbf{b}^{(k)}(u_i - u_0)\} = \begin{pmatrix} 1 \\ \\ \dfrac{x_i}{x_i + (a_3^{(k)} + \ b_3^{(k)}(u_i - u_0))} \\ \\ -\dfrac{(a_2^{(k)} + \ b_2^{(k)}(u_i - u_0))\ x_i}{\{x_i + (a_3^{(k)} + \ b_3^{(k)}(u_i - u_0))\}^2} \end{pmatrix}$$

We iteratively update $(\mathbf{a}, \mathbf{b})^T$ by

$$\begin{pmatrix} \mathbf{a}^{(k+1)} \\ \mathbf{b}^{(k+1)} \end{pmatrix} = (F_k^T W F_k)^{-1} F_k^T W \mathbf{y}_k,$$

where $\mathbf{y}_k = (y_{1,k}, \cdots, y_{n,k})^T$ and

$$W = \text{diag}\{K_h(u_1 - u_0), \cdots, K_h(u_n - u_0)\}$$

Denote the resulting estimate of $(\mathbf{a}, \mathbf{b})^T$ by $(\widehat{\mathbf{a}}, \widehat{\mathbf{b}})^T$. Then

$$\widehat{\boldsymbol{\beta}}(u_0) = \widehat{\mathbf{a}}, \quad \text{and} \quad \widehat{\boldsymbol{\beta}}'(u_0) = \widehat{\mathbf{b}}.$$

To implement the proposed estimation procedures, we need to select a bandwidth. In this analysis, we will employ multiple fold cross validation to select a bandwidth. Specifically, we first partition the data into Q groups, with the j^{th} group consisting datum points with indices

$$d_j = \{Qk + j, k = 1, 2, 3, ...\}, \quad j = 0, 1, 2, ..., Q - 1$$

We fit the model and obtain the j^{th} estimate, using data without including data in the j^{th} group. Now, we denote by $\widehat{y}_{-d_j}(u_i, \mathbf{x}_i)$ the fitted value using the data with j^{th} group deleted. Define the cross-validation score to be

$$CV(h) = \sum_{j=0}^{Q-1} \sum_{i \in d_j} [y_i - \widehat{y}_{-d_j}(u_i, \mathbf{x}_i)]^2 \tag{5.4}$$

We choose the bandwidth h that minimizes the $CV(h)$; that is, we select the bandwidth h that provide the model that best fits of the data. In our analysis, we set $Q = 20$. The selected bandwidth is $h = 0.15$.

Figure 5.3 depicts the estimated coefficient functions $\beta_1(u)$, $\beta_2(u)$, and $\beta_3(u)$, based on a bandwidth of $h = 0.15$, which implies that we use about 15% of the data set for estimation at a given u_0. Figure 5.3 describes the extent to which the association of *Net Ecosystem Exchange* (NEE) and *Photosynthetically Active Radiation*(PAR) over different temperature level. It shows very clearly that the coefficient functions vary with temperature, suggesting the NEE-PAR relationship depends on a covariate level, that is the temperature level. The 95% pointwise confidence interval is also depicted in Figure 5.3.

5.3 Testing Procedure

It is a scientific question whether the coefficient function really depends on temperature. To address this question, we employ the GLRT procedure for the following hypothesis.

$$H_0 : \beta_j(u) = \gamma_{j0}, \qquad j = 1, 2, 3 \quad \text{versus} \quad H_1 : \beta_j(u) \neq \gamma_{j0}, \qquad \text{for at least one j}$$

where γ_{j0} is a constant.

Figure 5.4 illustrates the estimated coefficient functions $\tilde{\boldsymbol{\beta}}$ under null hypothesis. The solid lines are the estimated function coefficients within the null space, H_0 of constant coefficients; while the solid curves are the nonparametric estimated coefficient functions within the alternative space, H_1, of varying-coefficients, with pointwise 95% confidence interval bands in dashed curves. Under the null hypothesis, $\tilde{\boldsymbol{\beta}} = (5.2904, \quad 32.2159, \quad 537.0713)$ with standard error $(0.2229, \quad 0.4056, \quad 27.0715)$. We further calculate the GLRT, which equals 36.0414.

We now use nonparametric bootstrap procedure to obtain the null distribution. The estimated density function of the test statistic T under H_0 is depicted in Figure 5.5, based on

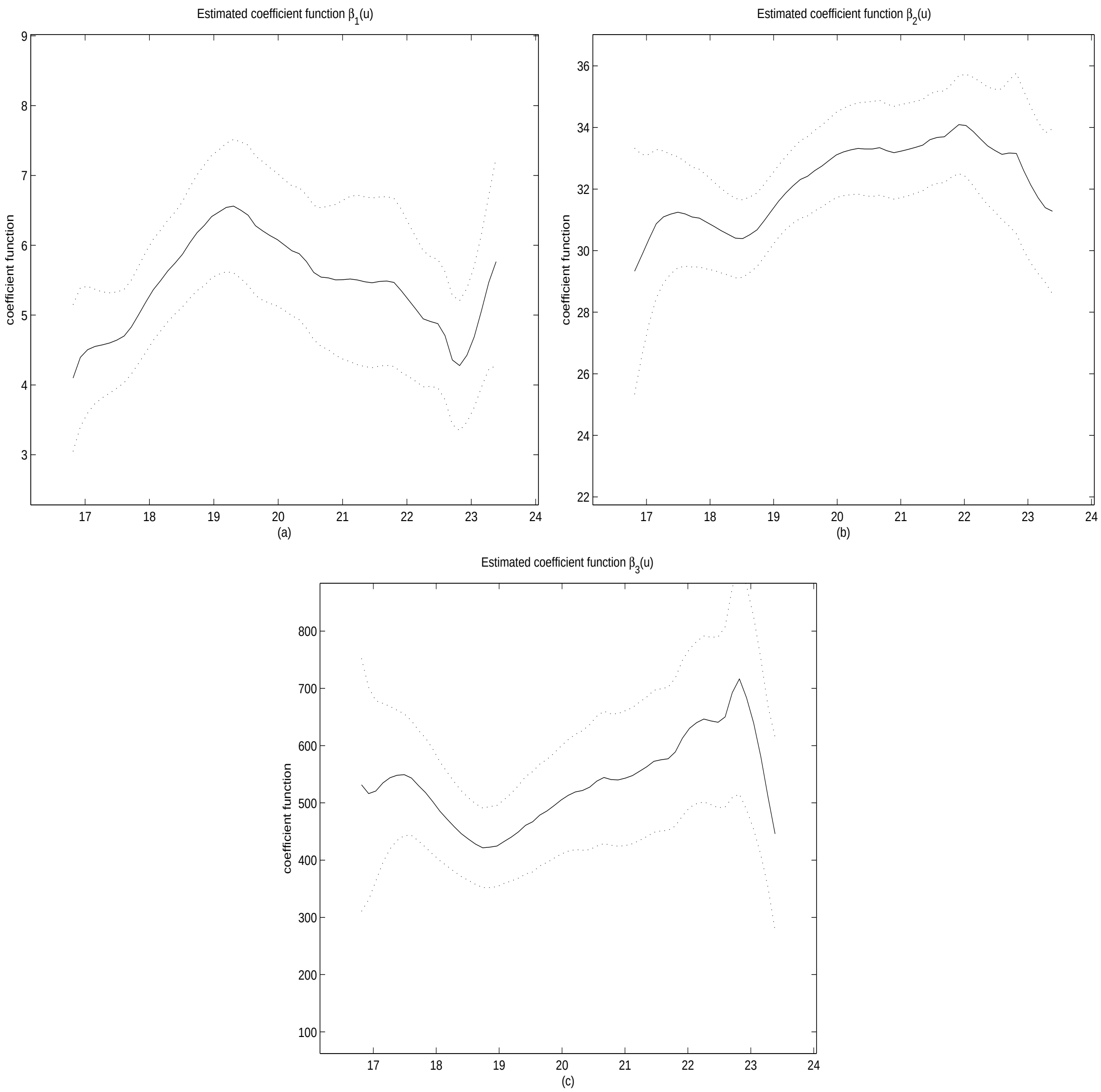

Figure 5.3: Plots of estimated coefficient functions of nonlinear varying-coefficient model. The solid curves are estimated coefficient functions, and the dotted curves are pointwise 95% confidence intervals.

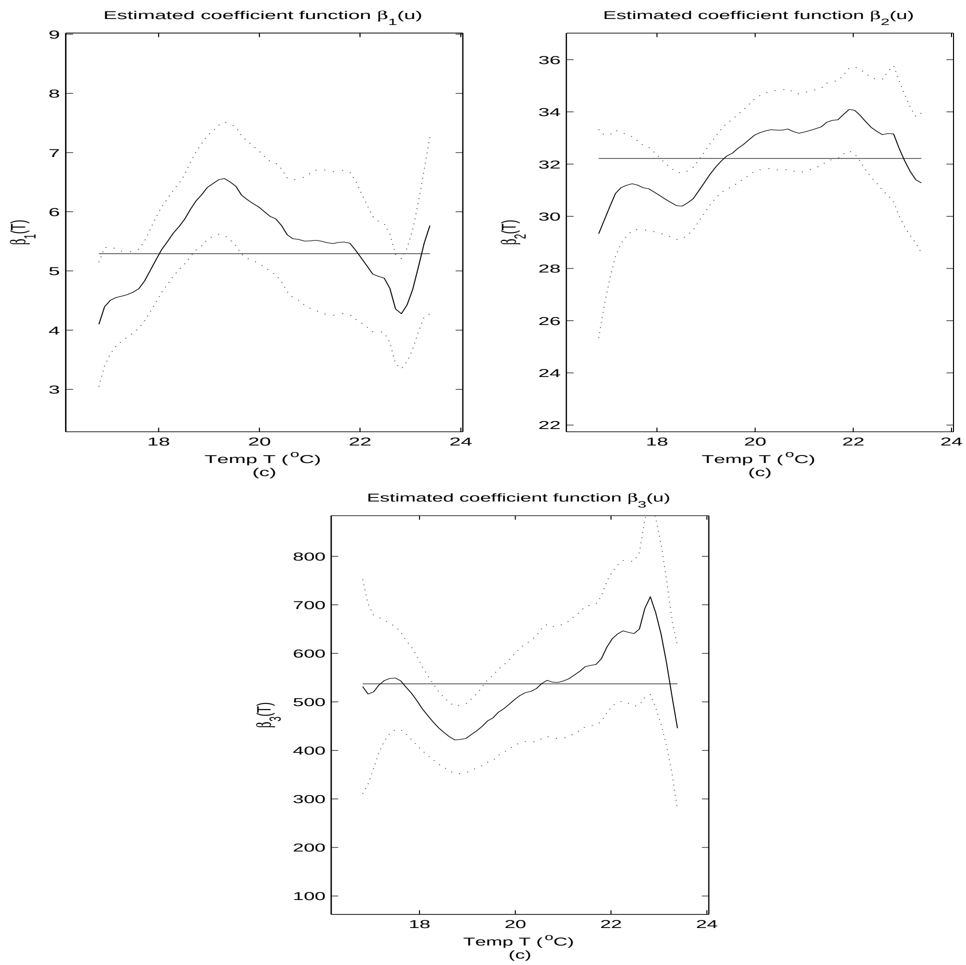

Figure 5.4: Plots of estimated coefficient functions of nonlinear varying-coefficient model using ecological data. The fluctuated solid curves are estimated coefficient functions for the exponential model; the solid straight lines are estimated coefficient functions under null hypothesis; the dotted curves are pointwise 95% confidence intervals obtained using standard error formulas.

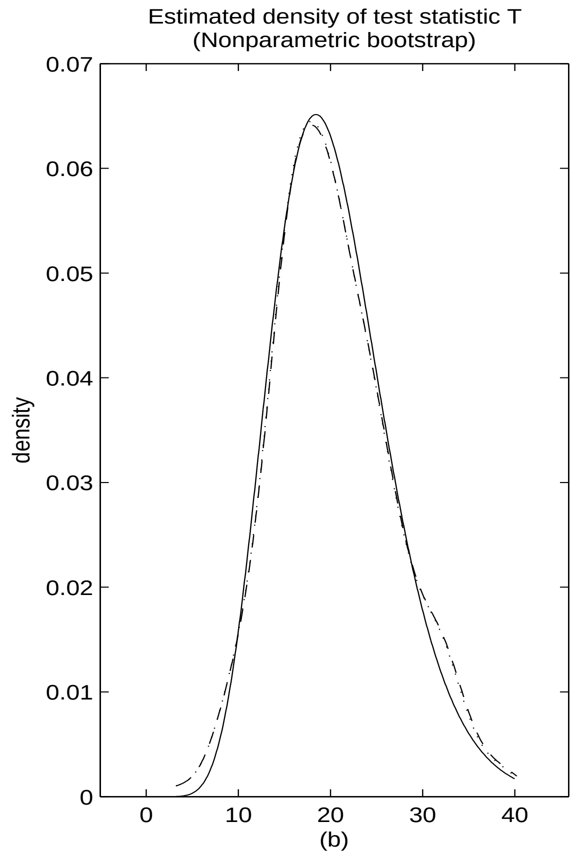

Figure 5.5: The estimated density of test statistic T by 1000 bootstrap samples. The dotted curve is the estimated density of generalized likelihood ratio (GLR) test statistic T; the dash-dotted curve is the estimated density of generalized F-test statistic T_0. The solid curve is the density of Chi-squared distribution with 20.41 degree of freedom.

1000 bootstrap samples. The estimated null distribution of test statistic T is close to a Chi-squared distribution with 20.41 degrees of freedom, and the GLRT has P-value of 0.0177, which rejects the null hypothesis at level 0.05. The testing result concludes a nonlinear varying-coefficient model yields a better fit to the data.

Chapter 6

New Application of Varying Coefficient Models to Financial Data

In this chapter, we apply varying-coefficient models for a financial data set. We demonstrate that varying-coefficient models are superior to ordinary linear regression models in terms of modeling fitting and model prediction.

6.1 Introduction

In the finance academics, practitioners, and regulators, it has been a major area of interest and concern to assess and manage credit risk of corporate bonds. See Caouette, Altman, and Narayanan (1998), Saunders and Allen (2002), and Duffie and Singleton (2003). Finding the significant factors that affect credit yield spreads of corporate bonds is an important issue.

Huang and Kong (2003) analyzed the credit spread data set for nine Merrill Lynch corporate bond indexes and investigated the important factors affecting credit yield spreads of corporate bonds. Their study was on credit spread data set which is the option-adjusted bond indexes from January 1997 through July 2002. In Huang and Kong (2003), it has been found that Russell 2000 index historical return volatility and the Conference Board composite leading and coincident economic indicators have significant power in explaining credit spread changes. Using linear regression model, these three variables plus the interest rate level, the historical interest rate volatility, the yield curve slope, the Russell 2000 index return, and a high-minus-low factor together explain more than 40% of credit spread changes.

Huang and Kong (2003) built a foundation for finding the important factors affecting credit spread. Based on their study, our purpose is to extend the study on credit spreads of corporate bonds by proposing new statistical modeling and testing techniques, in aim to improve model fitting and to construct a novel forecasting technique and improve forecasting accuracy. We will employ time-varying coefficient models(Hastie and Tibshrani, 1993). As we discuss later, the new modeling procedure can enhance model fitting and forecasting accuracy of future credit spread change, which is crucial to financial practitioners, researcher, and policy makers.

The chapter is organized as follows. In section 2, we discuss previous findings and describe the credit spread data set. In section 3, we conduct Chow's structure change test, and focus on motivations of utilizing the proposed time-varying-coefficient model. In section 4, we apply proposed modeling and testing procedure to credit spread data. Model fitting is compared to that from linear regression model. A novel forecasting procedure using time-varying-coefficient model is discussed and applied to the credit spread data. A modification of the forecasting procedure is proposed in which a weighted forecasting procedure is studied. We compare both model fitting and forecasting accuracy between time-varying-coefficient model and linear regression model. Generalized likelihood ratio test is applied to determine whether some coefficient is invariant over covariate and whether some coefficient is statistically insignificant. In section 5, we conduct extensive Monte Carlo simulations to further assess the proposed modeling and testing procedure. Section 6 summarizes the study and discusses future work.

6.2 Discussion on previous work and description of the credit spread data

6.2.1 Discussion on previous work

The credit spreads data set analyzed in this chapter has previously been analyzed in Huang and Kong (2003), where the determinant of corporate bond credit spreads is examined using monthly option-adjusted spreads for nine Merrill Lynch corporate bond indexes from January 1997 through July 2002. In Huang and Kong (2003), there are nine Merrill Lynch corporate bond indexes include six OAS series for investment-grade corporate bonds: AA-AAA and BBB-A rated series with maturities of 1-10 years, 10-15 years, and 15+ years, also include three series for high-yield corporate bonds with ratings of BB, B, C. Linear regression and ordinary-least-squares (OLS) estimation approach are used in Huang and Kong (2003) to find the most powerful factors that affect the credit spreads, i.e. the powerful factors that affect each of nine series of corporate bond indexes. To account for model errors with serial correlation or heteroscedasticity of unknown form found through preliminary analysis, Newly-West (1987) heteroscedasticity and autocorrelation-consistent covariance matrix estimator is used in testing the estimators from ordinary-least-square (OLS) estimation. Huang and Kong (2003) finds the Russell 2000 index historical return volatility and the Conference Board composite leading and coincident economic indicators have significant power in explaining credit spread changes, especially for high-yield indexes. The three variables together with the interest rate level, the historical interest rate volatility, the yield curve slope, the Russell 2000 index return, and a high-minus low factor together can explain more than 40% of credit spread changes for five bond indexes. These eight variables together can explain 67.68% and 60.82% of credit spread changes for the B and BB-rated indexes. The analysis confirms that credit spread changes for high-yield bonds are more closely related to equity market factors and also provides evidence in favor of incorporating macroeconomic factors into credit risk models. Huang and Kong (2003) build a foundation for finding important factors affecting change of credit spreads.

6.2.2 Description of the credit spread data set

The available credit spread data are monthly option-adjusted spreads for nine Merrill Lynch corporate bond indexes from January 1997 through August 2002 and contain 68 continuous observations in time. The data set contains August 2002 observation in addition to the data set used in Huang and Kong (2003). Variables included in the data set are option-adjusted spreads for nine Merrill Lynch corporate bond indexes, change in yield of Merrill Lynch Treasury master Index denoted by Δr, change in CBOE VIX denoted by Δvix, change in yield of Merrill Lynch 15+ years Treasury Index minus yield of Merrill Lynch 1-3-year treasury Index denoted by Δslope, change in Russell 2000 index return denoted by Δrusrtn, and change in S&P index return denoted by Δsprtn. Our goal is to model how credit spreads are affected by other financial and macroeconomic factors, therefore, the nine changes in credit spreads are considered response variables and the rest five variables, treated as proxies of financial market and macroeconomic environment, are considered predictor variables.

We will further assign notations Y_1 through Y_9 to the nine response variables and X_0 through X_5 to the five predictor variables, respectively. We summarize their correspondence in Table 6.1 and Table 6.2, where Table 6.1 contains response variables, i.e. the nine Merrill Lynch corporate bond indexes; Table 6.2 contains predictor variables, i.e. the financial market and macroeconomic variables that may potential affect changes in credit spreads.

Using the above notations, the credit spreads data set can be expressed as

$$(Y_i^j,\ \mathbf{X_i},\ t_i) \quad \text{for } i = 1, 2, 3, \ldots, 68 \text{ and } j = 1, 2, 3, \ldots, 9$$

where the index j represent the nine Merrill Lynch corporate bond indexes, $\mathbf{X_i} = (\mathbf{1}, X_1, X_2, X_3, X_4, X_5)$ are the set of predictor variables, and covariate t_i, $i = 1, 2, 3, \ldots\ldots, 68$ is the underlying time variable that represent dates from January 30th, 1997 to August 30th, 2002 (1997.01.30 to 2002.08.30).

Table 6.1: Response variables

Notation	Description
Y_1=Δhs	AA-AAA 1-10 yrs
Y_2=Δhi	AA-AAA 10-15 yrs
Y_3=Δhl	AA-AAA 15+ yrs
Y_4=Δls	BBB-A 1-10 yrs
Y_5=Δli	BBB-A 10-15 yrs
Y_6=Δll	BBB-A 15+ yrs
Y_7=Δbb	BB
Y_8=Δb	B
Y_9=Δc	C

Table 6.2: Description of predictor variables

Notation	Description
X_0=1	Intercept for the Linear Regression Model
X_1=Δr	Changes in yield of Merrill Lynch Treasury Master Index
X_2=Δslope	Changes in yield of Merrill Lynch 15+ years Treasury Index minus yield of Merrill lynch 1-3-year Treasury Index
X_3=Δvix	Change in CBOE VIX
X_4=Δrusrtn	Russell 2000 Index Return
X_5=Δsprtn	S&P 500 Index Return

Also note that since X_4 (Russell 2000 Index Return) and $X5$ (S&P 500 Index Return) preserve very high linear correlation, in future analysis X_5 is extracted from the model to avoid multi-collinearity.

6.3 Chow structure change test and motivation of time-varying-coefficient model

We introduce the motivation for proposing time-varying-coefficient model, a new modeling technique, and its model estimation procedure, to the financial study on credit spreads data through discussion on Chow's structure change test.

Empirical considerations According to Chow (1960), when a linear regression is used to represent an economic relationship, the question often arises as to whether the relationship remains stable in two periods of time; quite often, there is no economic rationale in assuming that two relationships are completely the same. It is natural to extend and generalize the question to whether an economic relationship remains stable throughout the subintervals of a time period; and under many circumstances, there is no economic rationale to assume the relationships are completely the same.

In the credit spread data, we aim to find the significant factors that affect credit yield spreads of corporate bonds and to identify the structures or relationships through which the significant factors affect credit yield spreads. These relationships were represented by linear regression functions in Huang and Kong (2003). Since the credit spread data are monthly and weekly observations and span an extended period of time from January 1997 through August 2002, it is rational not to assume that the structures or relationships remain completely the same throughout the time period during which credit spread data were collected and, of particular importance to financial practitioners and regulators, the future time period for which the prediction of future credit spread is of interest.

Statistical considerations As the most commonly used statistical technique, regression analysis has been utilized to explore the association between response variables and predictor variables and to identify how predictor variables may impact on response variable. A linear predictor and constant model coefficients are assumed in linear regression models. However, there exist potential problems when any of the model assumptions are significantly violated. If the model structure changes significantly throughout the time period under investigation and therefore the model coefficients do not remain constants, the ordinary model estimators are no longer maintain some nice statistical properties: model estimators are biased and the efficiency of linear regression estimators does not hold.

6.3.1 Chow test for structural breaks

Chow test(Chow 1960) is an econometric test for structural or parameter stability of the regression model, which tests to determine whether the coefficients in a regression are the same in separate subsamples.

Chow test indeed is an application of the F-test. We consider regression function of two time periods.

$$Y = \mathbf{X}^T \boldsymbol{\beta}^{1st} + \varepsilon, \tag{6.1}$$

where $\boldsymbol{\beta}^{1st}$ represent the model coefficient of the first period of time;

$$Y = \mathbf{X}^T \boldsymbol{\beta}^{2nd} + \varepsilon, \tag{6.2}$$

where $\boldsymbol{\beta}^{2nd}$ represent the model coefficient of the second period of time;

$$Y = \mathbf{X}^T \boldsymbol{\beta}^{ALL} + \varepsilon, \tag{6.3}$$

where $\boldsymbol{\beta}^{All}$ represent the model coefficient of the entire period of time.

Using Chow test, we test for the hypothesis that the model structures of the first time period and the second time period are the same; that is the regression functions of the first and second time periods have the same set of model coefficients, or say $\boldsymbol{\beta}^{1st}$ and $\boldsymbol{\beta}^{2nd}$ are equal.

We will construct the F-statistic using the sum of squares of restricted and unrestricted regressions, and obtain the sum of squared residuals. Define

RSS_R=RSS_{ALL}, the sum of squared residuals when we restrict two periods having equal intercepts and slopes.

RSS_{UR}=RSS_1 + RSS_2, the sum of squared residuals when we allow each period having its own intercepts and slopes.

The F-statistic to evaluate the structural change is

$$F^*_{k,T_1+T_2-2k} = \frac{(RSS_R - RSS_{UR})/k}{RSS_{UR}/(T_1+T_2-2k)} = \frac{(RSS_{ALL} - (RSS_1 + RSS_2))/k}{(RSS_1 + RSS_2)/(T_1+T_2-2k)}, \tag{6.4}$$

where k is the number of parameters included in the linear regression function, and T_1 and T_2 are the number of observations in the first and second time period, respectively.

Now we conduct Chow test on the credit spread data to determine, during the time period of January 1997 through August 2002, whether there exist any structural changes on the linear relationship between the explanatory variables and the monthly option-adjusted spreads of nine Merrill Lynch corporate bond indexes. Chow test is conducted for the spreads of each of the nine Merrill Lynch corporate bond indexes. Note that for exploratory purposes, for the spreads of each corporate bond indexes, we partition the entire time period into two subsamples through three different ways by using three cut-off points: the 1st quartile, the 2nd quartile, and the 3rd quartile. The reason for this is that we expect certain structural change in the regression model along the time period from January 1997 through August 2002; here we choose three typical cut-off points within the duration of the data, and test the stability of the regression functions from the resulting two subsamples obtained in three typical ways.

In Table 6.3, the exploratory Chow structural break test results for the spreads of nine Merrill Lynch corporate bond indexes is shown. We use significance level of 0.10 to test for the possible change in regression functions. Using the three typical partitions, we find seven out of the nine spreads have significant structural changes during the time period from January 1997 through August 2002. The Chow test results also show strong indication of structural changes for the AA-AAA 10-15 yrs (p-value=0.0047 at break point 52 on 3rd

Table 6.3: Chow Structural Break Test Results

Chow Structural Break Test (at 0.10 significance level)					
Response Variable Y_1					
Break Point	Num DF	Dum DF	F-Value	Pr > F	Significant Break Point(s)
18	5	57	2.15	0.0728	*yes*
34	5	57	2.16	0.0710	*yes*
52	5	57	0.62	0.6880	
Response Variable Y_2					
Break Point	Num DF	Dum DF	F-Value	Pr > F	Significant Break Point(s)
18	5	57	2.76	0.0265	*yes*
34	5	57	1.32	0.2687	
52	5	57	3.83	0.0047	*yes*
Response Variable Y_3					
Break Point	Num DF	Dum DF	F-Value	Pr > F	Significant Break Point(s)
18	5	57	1.48	0.2104	
34	5	57	4.68	0.0012	*yes*
52	5	57	1.26	0.2955	
Response Variable Y_4					
Break Point	Num DF	Dum DF	F-Value	Pr > F	Significant Break Point(s)
18	5	57	0.90	0.4903	
34	5	57	2.03	0.0875	*yes*
52	5	57	2.07	0.0828	*yes*
Response Variable Y_5					
Break Point	Num DF	Dum DF	F-Value	Pr > F	Significant Break Point(s)
18	5	57	1.11	0.3628	
34	5	57	2.15	0.0727	*yes*
52	5	57	3.54	0.0074	*yes*
Response Variable Y_6					
Break Point	Num DF	Dum DF	F-Value	Pr > F	Significant Break Point(s)
18	5	57	1.67	0.1560	
34	5	57	3.75	0.0052	*yes*
52	5	57	1.56	0.1870	
Response Variable Y_7					
Break Point	Num DF	Dum DF	F-Value	Pr > F	Significant Break Point(s)
18	5	57	0.82	0.5393	
34	5	57	1.48	0.2117	
52	5	57	4.51	0.0016	*yes*
Response Variable Y_8					
Break Point	Num DF	Dum DF	F-Value	Pr > F	Significant Break Point(s)
18	5	57	0.96	0.4527	
34	5	57	1.53	0.1961	
52	5	57	1.02	0.4143	
Response Variable Y_9					
Break Point	Num DF	Dum DF	F-Value	Pr > F	Significant Break Point(s)
18	5	57	0.62	0.6849	
34	5	57	0.69	0.6360	
52	5	57	0.99	0.4346	

$*$ For Y_8, highly significant break point found at 46 (p-value=00491); for Y_9, relatively large discrepancy or structural change found at point 55 (p-value=0.1578). and 57 (p-value=0.1790).

quartile), AA-AAA 15+ yrs (p-value=0.0012 at break point 34 on 2nd quartile), BBB-A 10-15 yrs (p-value=0.0074 at break point 52 on 3rd quartile), BBB-A 15+ yrs (p-value=0.0052 at break point 34on 2nd quartile), BB (p-value=0.0016 at break point 52 on 3rd quartile).

Although at the three chosen exploratory cut-off points, we do not find any significant structural changes for the spreads of B and C corporate bond indexes, we can not conclude there exist no structural changes for the two spreads. It is worthy of noting that, in fact, a full exploration of the date shows that for spread of B corporate bond, we find highly significant break point at 46 (p-value=00491), and for spread of C corporate bond, we find a relatively large discrepancy or structural change at point 55 (p-value=0.1578). and 57 (p-value=0.1790).

The exploratory study by Chow test for structural changes has shown that there exist strong to moderate structural changes of the regression functions during the time period from January 1997 through August 2002. The results strongly suggest that it is not appropriate to use constant coefficient regression models to describe the relationship between the explanatory variables and the credit spreads for the nine Merrill Lynch corporate bond indexes, which motivates us to search for alternative models.

6.4 Time-Varying coefficient models

When Chow's test for structural changes shows that there exist strong to moderate violation of constant coefficient assumption for linear regression model, we natural consider alternative statistical models that have more flexible model assumptions. In this section, we fit the data by using time-varying coefficient models. The estimation procedures and generalized likelihood ratio test for time-varying coefficient models have been introduced in Chapter 2.

Time-varying-coefficient model is a special case of varying-coefficient model, where the underlying covariate is time and model coefficients vary as covariate time varies. We will apply this newly proposed model to our credit spread data set $(Y_i^j,\, \mathbf{X_i},\, t_i)$ for $i = 1,\, 2,\, 3$,..., 68 and $j = 1,\, 2,\, 3$,..., 9, where the index j represent the nine Merrill Lynch corporate bond indexes, $\mathbf{X}_i = (X_1,\, X_2,\, X_3,\, X_4,\, X_5)$ for $X_1 = 1$ are the set of predictor variables, and covariate t_i, $i = 1,\, 2,\, 3$,......, 68 is the underlying time variable that represent dates from January 30th, 1997 to August 30th, 2002 (1997.01.30 to 2002.08.30).

Based on financial studies, the variable Russell 2000 Index Return (X_4) and S&P 500 Index Return ($X5 =$) preserve very high linear correlation, therefore, we extract X_5 from the model to avoid multi-collinearity. The time-varying-coefficient model takes the form of

$$Y = \mathbf{X}^T \boldsymbol{\beta}(t) + \varepsilon, \tag{6.5}$$

where

- ε is the error term with E(ε)=0 and Var(ε)=σ^2 .
- $\boldsymbol{\beta}(t) = \{\beta_1(t),\, \beta_2(t),\, \beta_3(t),\, \beta_4(t),\, \beta_5(t)\}$ are the time-varying-coefficient functions, and are functions of time t. We also assume $\beta_1(t)$, $\beta_2(t)$, $\beta_3(t)$, $\beta_4(t)$, and $\beta_4(t)$ are smooth nonparametric functions of t.
- Y is response variable notation, representing Y_1 through Y_9, whose interpretation have been explained in Table 1 and Table 2 of section 2.

Model (6.5) is a special type of varying-coefficient model, and it is referred to as time-varying coefficient model with covariate t for model coefficient functions. Time-varying coefficient models assume that model coefficient functions vary as time t varies, and they are used to explore the possible time-dependent effects. In financial studies, we expect data are collected throughout a period of time, and therefore it is reasonable to consider possible time-dependency into modeling procedures for financial data, which has been discussed in details in section 3. At fixed time t, model (6.5) is a constant coefficient model; as time t moves along, we expect model structure (i.e. relationship between Y and $\mathbf{X}$) changes in financial study.

To identify the model structure (i.e. relationship between Y and $\mathbf{X}$) at all time points, we estimate the coefficient functions $\boldsymbol{\beta}(t) = \{\beta_1(t), \beta_2(t), \beta_3(t), \beta_4(t), \beta_5(t)\}$ of model:

$$Y^j = \mathbf{X}^T \boldsymbol{\beta}(t) + \varepsilon, \tag{6.6}$$

for $j = 1, ..., 9$ representing credit spread changes for nine Merrill Lynch corporate bond indexes.

In this section, we present the estimated coefficient functions and their corresponding pointwise confidence intervals through graphs. In order to demonstrate that varying coefficient models are superior to ordinary linear regression models in terms of model fitting, we summarize the statistics obtained from the two models; to demonstrate this superiority in terms of model prediction, we conduct in sample prediction and compare prediction sum of squares of error of the two models in two different occasions.

6.4.1 Estimated Coefficient Functions

For the nine time-varying-coefficient models with dependent variables Y_0, Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8, and Y_9, respectively, we estimate their nonparametric coefficient functions $\beta_0(t)$, $\beta_1(t)$, $\beta_2(t)$, $\beta_3(t)$, and $\beta_4(t)$ for t=1, 2, 3, ..., 68, and present them in the following graphs.

Note that in the graphs, the observations are monthly data with original indexes Jan 30th, 1997 to Aug. 30th, 2002 (1997.01.30 to 2002.08.30). These 68 observations of form (Y_i, X_i, t_i) for $i = 1, 2, 3$,......, 68 and $X_i = (1, X_1, X_2, X_3, X_4, X_5)$ are then indexed by time t using $t_i = i$, $i = 1, 2, 3, ..., 68$.

Also note that for different dependent variables Y_1 through Y_9, the $y - axes$ in x-y coordinate systems have different limits, which adaptively accommodate the range of coefficient

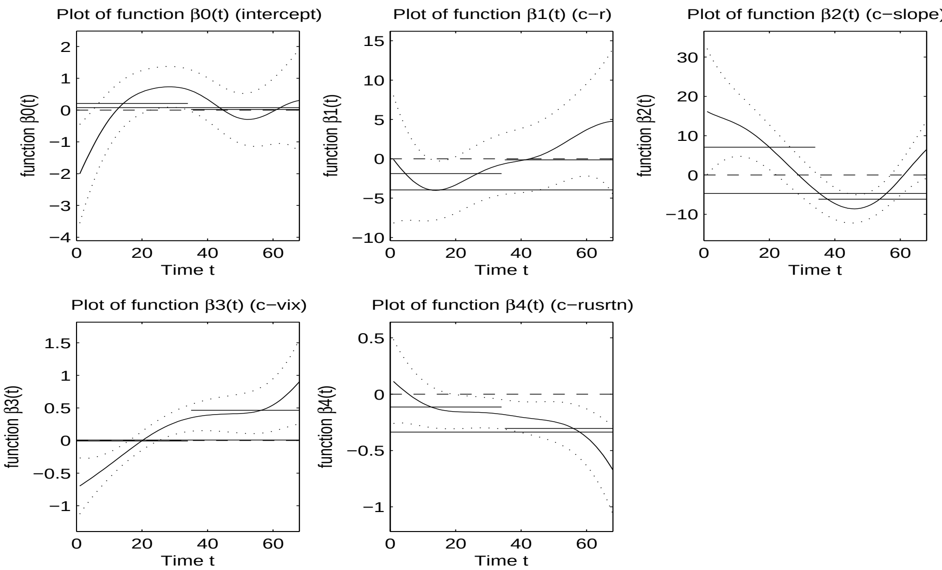

Figure 6.1: Plot of estimated coefficient functions for response variable Y_1. Solid curve: estimated coefficient functions (nonparametric); dotted curves: 95% pointwise confidence intervals; solid line: ordinary least-squares estimates; solid line (in half): ordinary least-squares estimates using first-half and second-half of the data, respectively.

functions $\beta_p(t)$, $p = 0,\ 1,\ 2,\ 3,\ 4$.

Summary from the plots In general, coefficient functions in models Y_7, Y_8, and Y_9 fluctuate more significantly than the coefficient functions in models Y_1, Y_2, Y_3, Y_4, Y_5, and Y_6. Also, the coefficient functions $\beta_0(t)$, $\beta_3(t)$, and $\beta_5(t)$ tend to be more constant over time than the coefficient functions for $\beta_1(t)$ and $\beta_2(t)$. That is to say, for each model the coefficients of the intercept and Rusell 2000 Index Return are more stable over time than for the coefficients of Change in the yield of Merill Lynch Index, Change in the yield of Merill Lynch Slope, and Change in CBOE VIX.

6.4.2 Comparisons in terms of R-squars

In assessment of how well time-varying coefficient (TVC) model and linear regression model fit the credit spread data by ordinary least squares (OLS) method, we compute the R^2 statistics for each of the nine models with dependent variables Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7,

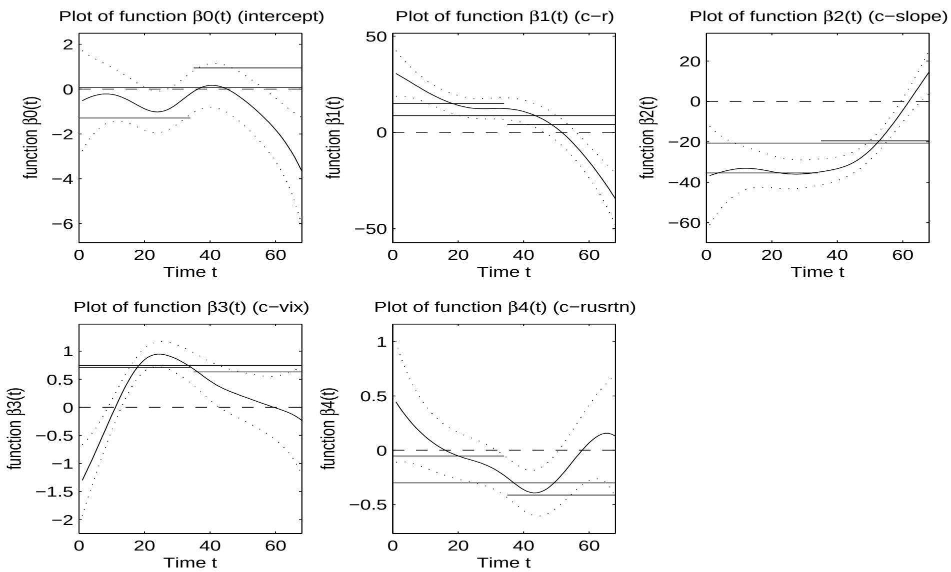

Figure 6.2: Plot of estimated coefficient functions for response variable Y_2. Caption is the same as Figure 6.1

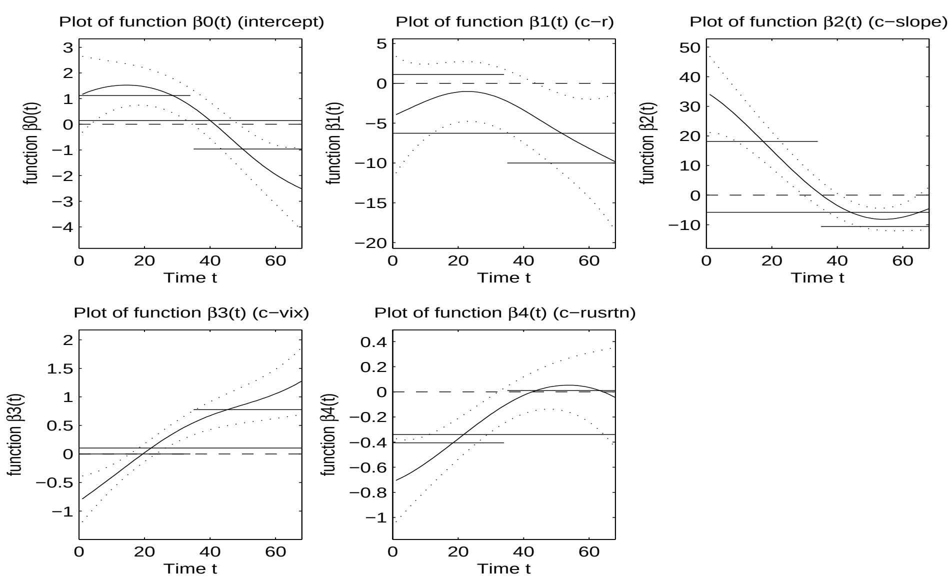

Figure 6.3: Plot of estimated coefficient functions for response variable Y_3. Caption is the same as Figure 6.1

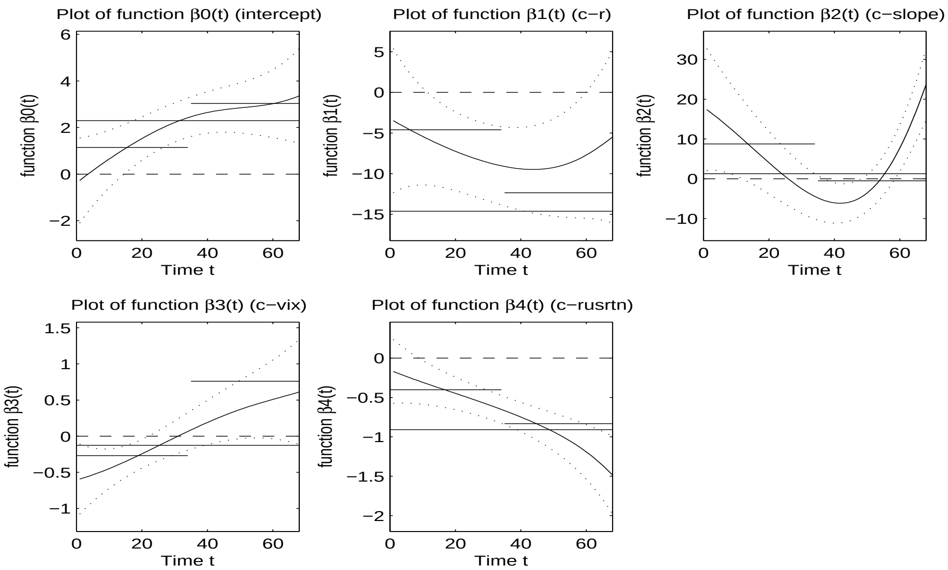

Figure 6.4: Plot of estimated coefficient functions for response variable Y_4. Caption is the same as Figure 6.1

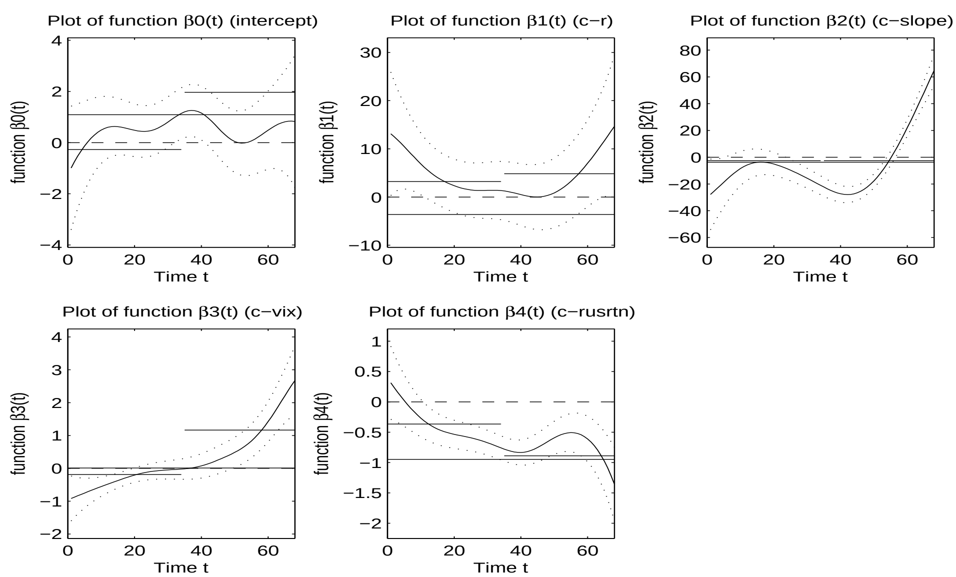

Figure 6.5: Plot of estimated coefficient functions for response variable Y_5. Caption is the same as Figure 6.1

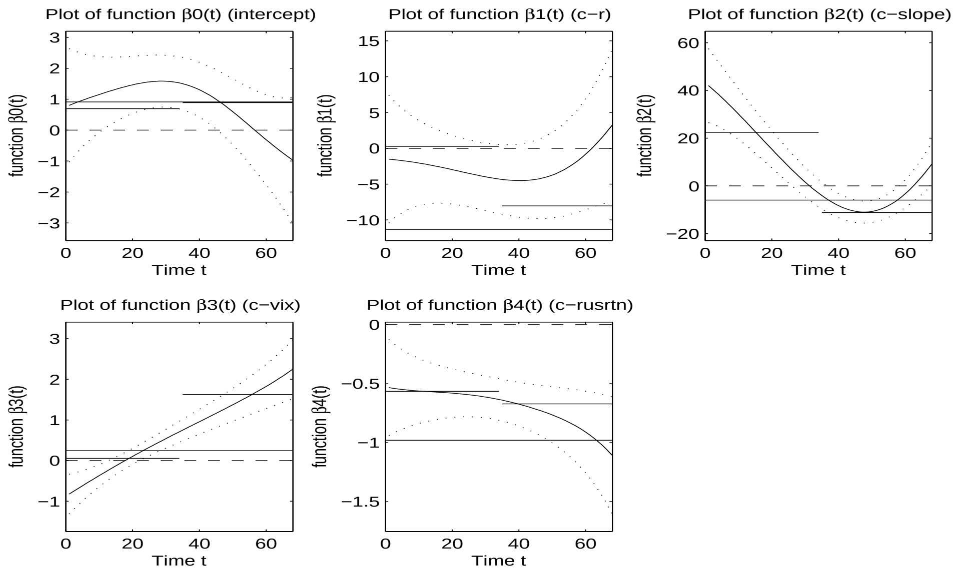

Figure 6.6: Plot of estimated coefficient functions for response variable Y_6. Caption is the same as Figure 6.1

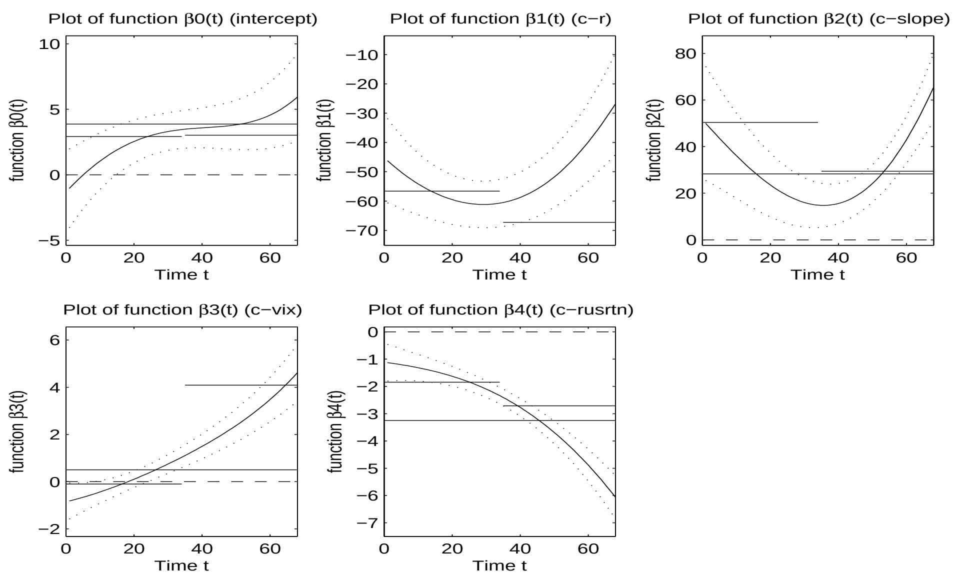

Figure 6.7: Plot of estimated coefficient functions for response variable Y_7. Caption is the same as Figure 6.1

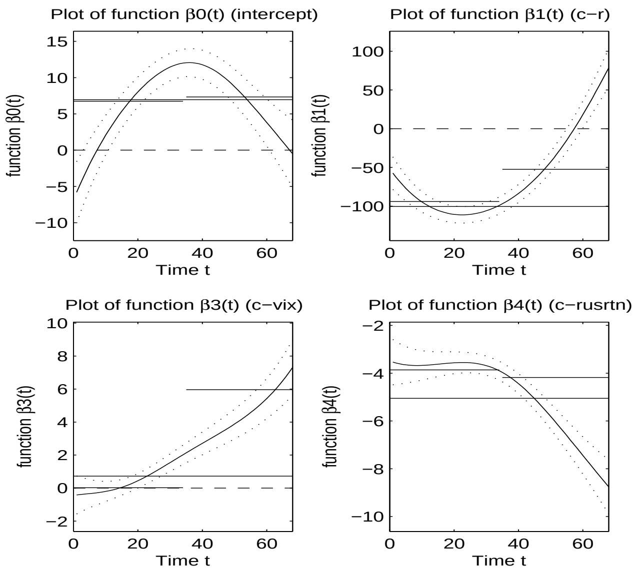

Figure 6.8: Plot of estimated coefficient functions for response variable Y_8. Caption is the same as Figure 6.1

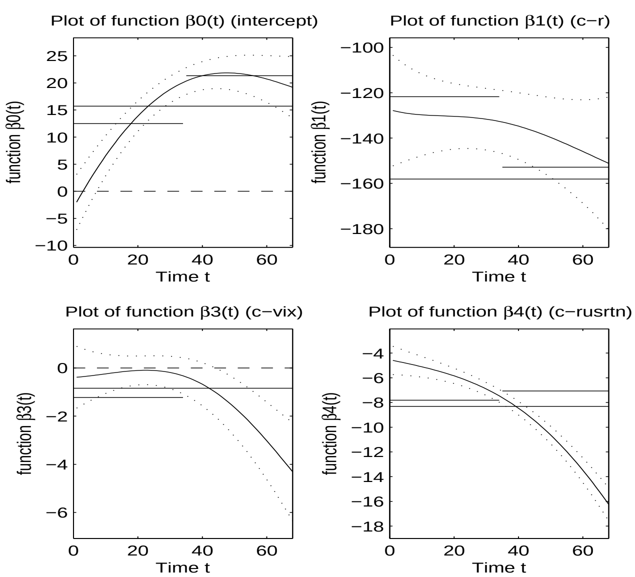

Figure 6.9: Plot of estimated coefficient functions for response variable Y_9. Caption is the same as Figure 6.1

Table 6.4: Comparison of R^2 statistics for TVC and OLS

Variable	R^2 **for TVC**	**MSE for TVC**	R^2 **for OLS**	OLS estimated coefficients $\widehat{\beta}=(\widehat{\beta}_0,\widehat{\beta}_1,\widehat{\beta}_2,\widehat{\beta}_3,\widehat{\beta}_4)^T$
Y_1	0.5591	3.3146	0.2303	$(0.0778, -3.9496, -4.6977, 0.0083, -0.3365)$
Y_2	0.6573	6.0335	0.4033	$(0.0783, 8.6817, -20.6318, 0.7445, -0.3011)$
Y_3	0.4857	4.6422	0.2170	$(0.1439, -6.2681, -5.8002, 0.1048, -0.3393)$
Y_4	0.5168	7.8716	0.3380	$(2.2969, -14.6168, 1.2810, -0.1266, -0.9091)$
Y_5	0.6702	7.0354	0.2349	$(1.0894, -3.6256, -3.7134, 0..0121, -0..9474)$
Y_6	0.5935	7.6779	0.3981	$(0.9109, -11.3222, -5.9744, 0.2445, -0.9802)$
Y_7	0.6605	25.7282	0.5343	$(3.8734, -82.0956, 28.2909, 0.5021, -3.2504)$
Y_8	0.6625	36.5587	0.5537	$(6.9534, -100.5260, 56.3680, 0.7257, -5.0531)$
Y_9	0.4731	76.1791	0.4287	$(15.7131, -158.0897, 74.2536, -0.8441, -8.3115)$

Y_8, and Y_9 fitting to the two proposed models, and summarize them in Table 6.4

Note that the statistics for time-varying-coefficient model in the summary table are based on a bandwidth h computed through the cross-validation criterion described in the following section. Also, note that the proposed local linear modeling is referred to as a one-step estimation procedure, which is simple and useful. See Cleveland, Grosse and Shyu (1991). However, it implicitly assumes that the coefficient functions $\beta_j(u), j = 1, 2,..., p$ possess approximately the same degree of smoothness. This assumption allows the coefficient functions to be estimated equally well in the same interval of covariate u.

6.4.3 Comparison in terms of prediction

To evaluate the prediction power of time-varying coefficient (TVC) model for our credit spread data set, it is natural to imagine the situation where we do not observe the last m values for Y, the change of credit spread, and we predict these m values of Y using models fitted from the first n-m observations, where n is the size of original data. Therefore, by pretending not observing the last m values of Y, we can compare the prediction power of the two candidate models. This is referred to as out-of-sample (OOS) prediction, and the comparison criterion is prediction mean square errors (PMSE) obtained from the two candidate models.

The credit spread data include $n = 68$ observations indexed by time t. To conduct out-of-sample (OOS) prediction, we fit the two candidate models using the first S ($S < n$) observations and evaluate the prediction mean squared error (PMSE) of observation $S+m$, for $m = 1, 2, ..., n-S$. For each of the nine models with dependent variables Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7, Y_8, and Y_9, respectively, for illustration purpose, we leave out the last $m(m = 3 or 5)$ observations of the data set; that is, we assume the last $m(m = 3$ or $5)$ observations on time indices $t = 66$, 67, 68 or $t = 64$, 65, 66, 67, 68 are unknown. We then apply the same estimation procedure on the rest of the observations and obtain estimated model coefficient function for time-varying-coefficient model and coefficient estimates for ordinary linear regression model. The last m (m=3 or 5) left-out observations are then predicted using the estimated model coefficient functions for time-varying-coefficient (TVC) model and the coefficient estimates for the ordinary linear regression (OLS) model, respectively. After we obtain the predicted value for the last m left-out observations, predicted Y at $t = 64$, 65, 66, 67, 68 for the nine models, we obtain the prediction sum of square of errors (PSSE) and then the prediction mean square errors (PMSE) using formulas:

$$PSSE = \sum_{i=64}^{68} (y_i^{pre} - y_i)^2 \tag{6.7}$$

$$PMSE = \sqrt{\frac{\sum_{i=64}^{68} (y_i^{pre} - y_i)^2}{m}} \tag{6.8}$$

where $m = 3 \;\; or \;\; 5$ is the number of predicted observations.

Demonstration of Prediction Procedure We illustrate by Figure (6.10) the prediction procedure using time-varying models. In predicting new observations using time-varying-coefficient model, we used the value of model coefficient $\beta(S)$ at time $t = S$, to obtain the predicted value for dependent variable Y; this is considered as leveling-out the estimated coefficient functional value at $t = S$, where S is the last observation used in estimating the predictive model. Let us demonstrate of prediction procedure using time-varying models by

$$Y_{pred} = \beta_0(t^*) \cdot X_0 + \beta_1(t^*) \cdot X_1 + \beta_2(t^*) \cdot X_2 + \beta_3(t^*) \cdot X_3 + \beta_4(t^*) \cdot X_4 + \varepsilon$$

where t^* represent the last time point, and the predicted response is obtained using the functional value of estimated coefficient functions at the last time point t^*.

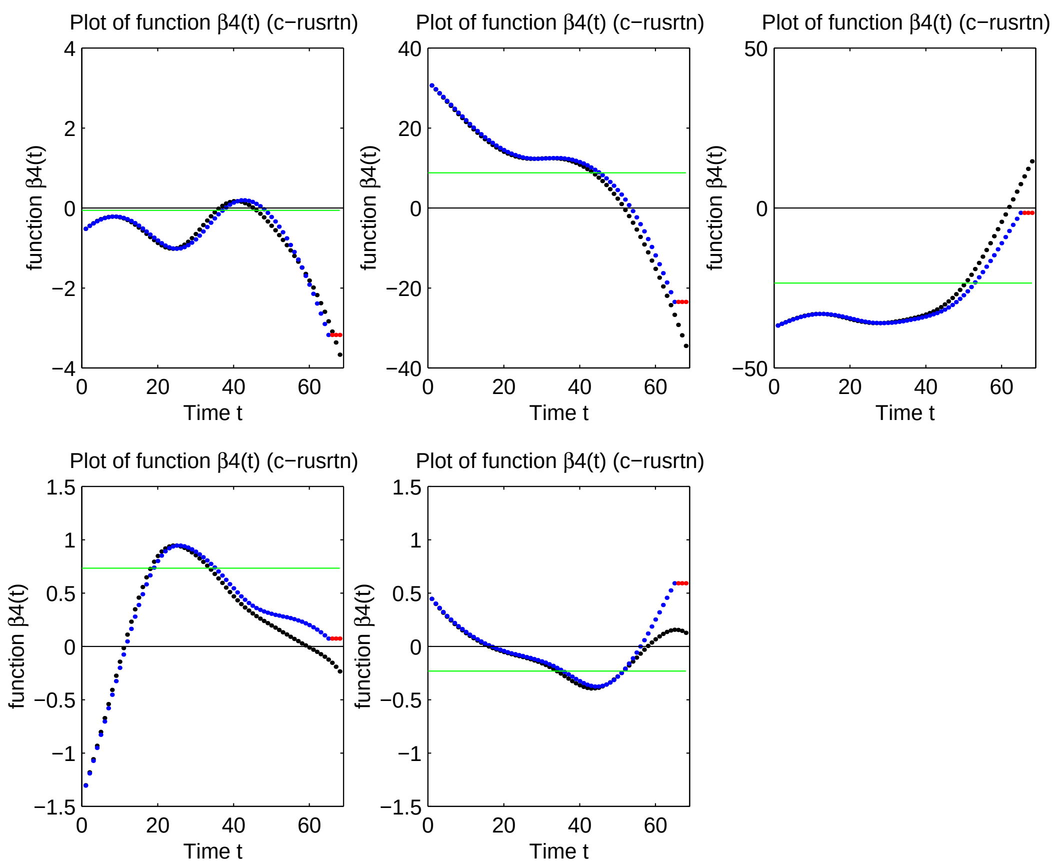

Figure 6.10: Demonstration plot for prediction procedures using time-varying models. Dotted-blue curves: estimated coefficient functions from data set with last m observations left out; dotted-red curves: predicted coefficient functions at the left-out time points; dotted-black curves: estimated coefficient functions from the complete data set; solid-green lines: least-square constant coefficient from data set with last m observations left out

To compare the prediction power of time varying-coefficient model and ordinary linear regression model, we summarize their corresponding prediction mean square errors (PMSE)

Table 6.5: Comparison of PMSE for TVC and OLS

Variable	**TVC model**	**OLS model**	$PMSE_{OLS}/PMSE_{TVC}$
Left-out last $m = 3$ observations are predicted			
Y_1	10.4354	10.4213	0.9986
Y_2	9.5689	11.1018	1.1602
Y_3	5.3647	4.4651	0.8323
Y_4	21.5439	24.2497	1.1256
Y_5	30.4192	36.3741	1.1958
Y_6	16.6999	18.4126	1.1026
Y_7	89.5060	97.9869	1.0948
Y_8	51.3387	73.0777	1.4234
Y_9	127.2548	115.0377	0.9040
Left-out last $m = 5$ observations are predicted			
Y_1	8.5239	8.1475	0.9558
Y_2	8.3525	10.3304	1.2368
Y_3	4.6893	3.9674	0.8460
Y_4	17.1070	18.9927	1.1102
Y_5	25.9667	29.7612	1.1461
Y_6	13.7520	15.3769	1.1182
Y_7	69.3116	75.6464	1.0914
Y_8	48.0922	58.5415	1.2173
Y_9	107.4774	112.2471	1.0444

in Table 6.5.

In Table 6.5 when last 3 left-out observations are predicted, the $PMSE_{OLS}/PMSE_{TVC}$ ratio shows that 7 out of 9 cases time-varying-coefficient model out-performs ordinary linear regression model by at least around 10% to 40%, and the only two cases when time-varying-coefficient model out-performs ordinary linear regression model, the different is not very significant and probably due to the random error inherent in the data set. For the summary table when last 5 left-out observations are predicted, the $PMSE_{OLS}/PMSE_{TVC}$ ratio shows that 6 out of 9 cases time-varying-coefficient model out-performs ordinary linear regression model by at least around 5% to 25%, and the only two cases when time-varying-coefficient model out-performs ordinary linear regression model, the different is not very significant

Table 6.6: Ratios of Prediction Mean Square Errors (**PMSE)**–OLS vs TVC models– for number of left-out observations $m = 1$ to 8

Y	**m=1**	**m=2**	**m=3**	**m=4**	**m=5**	**m=6**	**m=7**	**m=8**
$Y1$	0.6510	1.0601	0.9986	0.9398	0.9558	0.9436	0.9741	0.9769
$Y2$	15.2937	1.1889	1.1602	1.0534	1.2368	1.1985	1.1920	1.0930
$Y3$	0.8079	0.7310	0.8323	0.8094	0.8460	0.9505	0.8772	0.8990
$Y4$	2.2173	1.1345	1.1256	1.1277	1.1102	1.0928	1.0769	1.0638
$Y5$	4.5951	1.3127	1.1958	1.1138	1.1461	1.1057	1.0655	1.0332
$Y6$	2.4851	1.0726	1.1026	1.1359	1.1182	1.0970	1.0551	1.0370
$Y7$	0.2825	1.8194	1.0948	1.0859	1.0914	1.0940	1.0785	1.0659
$Y8$	17.0045	2.2915	1.4234	1.2312	1.2173	1.1589	1.1572	1.0498
$Y9$	0.9052	0.7596	0.9040	0.9484	1.0444	1.0478	1.0375	1.0294

either and again probably due to the random error inherent in the data set.

In order to further evaluate the prediction performance of time-varying-coefficient model versus ordinary least-squares model, we conduct the out-of-sample (OOS) prediction using reduced data set, namely, we use data

$$(Y_i, X_i, t_i) \text{ where } X_i = (1, X_1, X_2, X_3, X_4) \text{ and } i = 1, 2, 3, ..., 68 - R$$

and R is the number of observations trimmed off from the original data set.

One reason for conducting out-of-sample prediction using reduced data set is to make full use the available limited data and to explore the prediction power of time-varying-coefficient (TVC) under various situations. Another reason is that in practise of financial prediction, the out-of-sample prediction is up to only a few or a couple of future values; while the entire data set only provide us one such situation, namely, when the left-out observations m is very small.

The following tables summarize the comparison of PMSE from time-varying-coefficient and ordinary least-squares model using reduced data for $R = 1, 2, 3, 4$.

Note that all the summary statistics from time-varying-coefficient model in the comparison tables are based on an optimal bandwidth h computed through the cross-validation criterion described in the preceding section.

6.4.4 Testing for Time-delay Effect and Time Lag in Response

It is natural to further investigate whether the effect of factors affecting change of credit spread is immediate or there exists certain pattern of time delay in such effect. So the question arises if there is any time lag in the response variable. In order to study this question, we fit a time-varying-coefficient model for each time lag τ using the data

$$(Y_{i+\tau}, X_i, t_i) \text{ where } X_i = (1, X_1, X_2, X_3, X_4) \text{ and } i = 1, 2, 3, ..., 68.$$

We present the resulting residual sum of squares (RSS) for each of the time lag in figure (6.11). The RSS versus time lag τ plot suggests, as τ gets larger, so the RSS of the time-varying-model. This in turn implies that there exists n evidence for time delay effect in the response variable.

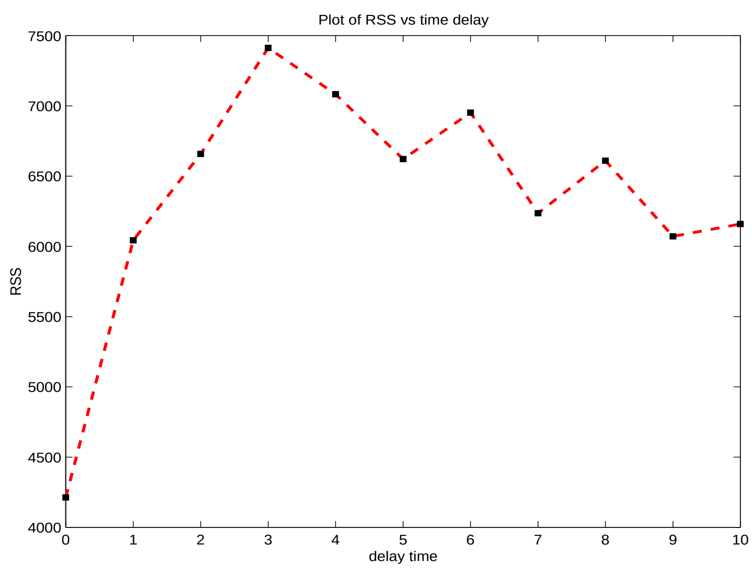

Figure 6.11: Testing whether there exists any time delay in the response variable

Table 6.7: Ratios of Prediction Mean Square Errors (**PMSE**)–OLS vs TVC models– using reduced model of size 67. For R=1, time-varying-coefficient model outperforms ordinary least-squares regression model in 72.2% of the cases.

Y	**m=1**	**m=2**	**m=3**	**m=4**	**m=5**	**m=6**
$Y1$	1.0709	0.9812	0.9219	0.9348	0.9225	0.9508
$Y2$	1.1328	1.1170	1.0109	1.2131	1.1886	1.2060
$Y3$	0.6891	0.8401	0.7797	0.8249	0.9579	0.8438
$Y4$	1.1064	1.1150	1.1439	1.1316	1.1140	1.0949
$Y5$	1.1643	1.1177	1.0348	1.0946	1.0761	1.0381
$Y6$	0.9996	1.0668	1.1553	1.1482	1.1406	1.1241
$Y7$	2.7154	1.0951	1.0845	1.0897	1.0926	1.0772
$Y8$	0.2233	1.1600	1.0822	1.1317	1.1145	1.1227
$Y9$	0.3500	0.8643	0.9243	1.0374	1.0446	1.0349

Table 6.8: Ratios of Prediction Mean Square Errors (**PMSE**)–OLS vs TVC models– using reduced model of size 66. For R=2, time-varying-coefficient model outperforms ordinary least-squares regression model in 96.3% of the cases.

Y	**m=1**	**m=2**	**m=3**	**m=4**	**m=5**	**m=6**
$Y1$	1.1773	1.1186	1.1776	1.134	1.2153	1.1943
$Y2$	1.9021	1.0806	2.7855	1.3264	1.2893	1.1337
$Y3$	1.5114	1.3614	1.3844	1.6830	1.2759	1.2946
$Y4$	1.4503	1.2247	1.2447	1.2444	1.2293	1.2042
$Y5$	1.5992	1.3767	1.6930	1.4714	1.3390	1.2340
$Y6$	1.7312	1.4327	1.4435	1.3738	1.2302	1.2083
$Y7$	1.0576	1.0466	1.0600	1.0672	1.0580	1.0490
$Y8$	1.1636	1.0982	1.1415	1.1201	1.1206	1.0661
$Y9$	0.8898	0.9417	1.0504	1.0518	1.0401	1.0217

Table 6.9: Ratios of Prediction Mean Square Errors (**PMSE**)–OLS vs TVC models– using reduced model of size 65. For R=3, time-varying-coefficient model outperforms ordinary least-squares regression model in 70.4% of the cases.

Y	**m=1**	**m=2**	**m=3**	**m=4**	**m=5**	**m=6**
$Y1$	1.0273	1.0409	0.7777	1.4811	1.3878	1.3312
$Y2$	0.7771	2.8099	1.2984	1.2719	1.1203	1.0977
$Y3$	1.3489	1.1234	1.7929	1.1276	1.1907	1.5744
$Y4$	0.8903	0.8034	0.9122	0.9528	0.9409	0.9698
$Y5$	0.4765	3.3772	2.7207	2.0239	1.6474	1.1646
$Y6$	1.3035	1.2355	1.2037	1.0837	1.0586	1.0232
$Y7$	0.4636	0.9590	1.2455	1.1653	1.14444	0.8076
$Y8$	0.6739	1.2635	1.2699	1.2889	0.9240	0.8333
$Y9$	1.5938	1.1322	1.0864	1.0570	0.9983	1.0127

Table 6.10: Ratios of Prediction Mean Square Errors (**PMSE**)–OLS vs TVC models– using reduced model of size 64. For R=4, time-varying-coefficient model outperforms ordinary least-squares regression model in 68.5% of the cases.

Y	**m=1**	**m=2**	**m=3**	**m=4**	**m=5**	**m=6**
$Y1$	1.1646	*0.6796*	1.4825	1.3342	1.2552	1.2663
$Y2$	39.5777	1.3291	1.2705	1.1144	1.0902	0.8778
$Y3$	1.0200	1.9860	1.1255	1.1593	1.5040	1.1921
$Y4$	0.1519	0.7959	0.8575	0.8717	0.9811	1.2656
$Y5$	14.0794	3.2671	2.2445	1.7116	1.1612	1.0869
$Y6$	0.4887	0.6685	0.9030	0.8877	0.8884	1.6315
$Y7$	1.0096	1.2932	1.1787	1.1528	*0.8138*	1.1211
$Y8$	2.4954	2.0093	1.6283	0.9490	0.8576	1.0018
$Y9$	1.1227	1.0830	1.0547	0.9868	0.9939	1.0050

Chapter 7

Conclusions and Future Work

Motivated by a real data analysis of data in ecology, we proposed nonlinear varying-coefficient models, a new class of varying-coefficient model. We developed the statistical inference procedures for nonlinear varying-coefficient models. Specifically, we proposed an estimation procedure to the nonlinear varying coefficient models by local linear regression techniques, derived the asymptotic normality of the resulting estimate. We proposed a standard error estimate by the conventional sandwich formula. We further extend the generalized likelihood ratio test for the nonlinear varying coefficient models to test whether the coefficients really depend on a covariate. For nonlinear regression models, the optimization is more challenging than that for existing varying coefficient models due to the fact that there is no closed form for the resulting estimate and/or the objective function is typically nonconvex. The finite sample performance are empirically examined by Monte Carlo simulation studies. From our simulation studies, the proposed procedures perform well.

In Chapter 5, we applied the proposed estimation and inference procedure for the empirical analysis of an ecological data. By applying the generalized likelihood ratio test, we showed that model coefficients for the nonlinear ecological model vary as temperature varies and conclude the nonlinear relationship between NEE and PAR really depends on temperature. As new applications of varying coefficient models, Chapter 6 deals with applying some existing procedures to some financial data sets. By comparing predictive mean squared error (PMSE) values, it is demonstrated that varying-coefficient models are superior to an ordinary linear regression model, the commonly used model in finance research.

Some future research is needed in this topic.

1. The present work showed empirically through simulation studies that Wilks' phenomenon valid for the proposed generalized likelihood ratio test; however, it will be meaningful to provide theoretical foundations of the result by proving, for nonlinear varying-coefficient models, the asymptotic null distribution of Generalized Likelihood Ratio test statistic has a chi-square distribution with degrees of freedom which do not depend on the unknown parameters presented in the model under the null hypothesis.

2. The hypothesis testing problem concerned is to test whether there is at least one coefficient varies as a covariate varies. To extend the testing problem, we will construct hypothesis testing procedures on testing whether a subset of coefficients vary as a covariate, with the aid of using backfitting algorithm.

3. The current work is concerned about the estimation of a known form of the nonlinear model f. In the situation when the functional form f is unknown, we can establish statistical procedures to identify this unknown functional form of f.

4. Our models consider varying-coefficients and independent error terms; however, when dealing with panel data in economic or financial studies, error terms are in most cases correlated instead being approximately independent. Therefore, it may be practical to construct varying-coefficient models with time-series-like correlated error terms, and develop the estimation and hypothesis testing procedures for the model to provide superior predictive models.

5. It is of interest to apply the proposed nonlinear varying-coefficient model estimation and hypothesis testing procedures to pharmacokenetic and pharmacodynamic models widely used in pharmacology. Many pharmacological models are nonlinear in nature

but overlook the underlying characteristics such as subject's age, blood pressure, and weight, etc.. Nonlinear varying-coefficient models with coefficient depending on these underlying subject's characteristics. We can apply our inference procedures in hoping to improve modeling fitting and thus lower the expenses of pharmaceutical companies.

Bibliography

Barnhill, T.M., Jr., F.L.Joutz, and W.F. Maxwell. (2000). Factors affecting the yields on noninvestment grade bond indices: a cointegration analysis. *Journal of Empirical Finance*, **7** , pp.57-86.

Brown, D. (2001). An empirical analysis of credit spread innovation. *The Journal of Fixed Income*, **9** , pp.9-27.

Brumback, B. and Rice, J. (1998), Smoothing spline models for the analysis of nested and crossed samples of curves, *J. Amer. Statist. Assoc.*, **93**, 961–976.

Cai, Z, Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.*, **95**, 888-902.

Campbell, J. Y., and G. Taksler. (2003). Equity volatility and corporate bond yields. *Journal of Finance*, **58**.

Caouette, J., E. Altman, and P. Narayanan. *Managing Credit Risk: The next Great Financial Challenge.* New York: John Wiley & Sons, 1998.

Carey, E.V., Callaway, R.M. and DeLucia, E.H. (1998). Increased photosynthesis offsets costs of allocation to sapwood in an arid environment, *Ecology*, **79**, 2281-2291.

Chen, R. and Tsay, R.S. (1993), Functional-coefficient autoregressive models, *J. Amer. Statist. Assoc.*, **88**, 298-308.

Chen, W.J., Black, T.A., Yang,P.C., Barr, A.G., Neumann, H.H., Nesic, Z., Blanken, P.D., Novak, M.D., Eley,J. Ketler, R.J. and Cuenca, A. (1999) Effects of climatic variability on the annual carbon sequestration by a boreal aspen forest, *Global Change Biology*, **5**, 41-53.

Chiang, C.-T., Rice, J.A. and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Amer. Statist. Assoc.*, **96**, 605-619.

Collin-Dufresne, P., and R. Goldstein. (2001). Do credit spreads reflect stationary leverage ratios? *Journal of Finance*, **56**, pp. 1929-1957.

Collin-Dufresne, P., R. Goldstein, and S. Martin. (2001). The determinants of credit spread changes. *Journal of Finance*, **56** , pp. 2177-2207.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**, 377-403.

Curtis, P.S. and Wang, X. (1998). A meta-analysis of elevated CO_2 effects on woody plant mass, form, and physiology, *Oecologia*, **113**, 299-313.

Curtis, P.S., Vogel,C.S., Wang,X., Pregitzer, K.S., Zak, D.R., Lussenhop, J., Kubiske, M., and Teeri, J.A. (2000). Gas exchange, leaf nitrogen, and growth efficiency of Populus tremuloides in a CO_2 enriched atmosphere, *Ecological Applications*, **10**, 3-17.

Das, S., and P. Tufano. (1996). Pricing credit sensitive debt when interest rates, credit ratings and credit spreads are stochastic. *Journal of Financial Engineering*, **5**, pp.161-198.

Donoho, D. L. (2000). High-dimensional data analysis: the curses and blessings of dimensionality. *Aide-Memoire of the lecture in AMS conference "Math challenges of 21st Centrury.* Available at http://www-stat.stanford.edu/~donoho/Lectures/AMS2000/AMS2000.html.

Duffee, G. (1998). The relation between treasury yields and corporate bond yield spreads. *Journal of Finance*, **54**, pp. 2225-2241.

Duffie, D., and K. Singleton. *Credit Risk: Pricing, Measurement and Management.* Princeton: Princeton University Press, 2003.

Duffie, D., and K. Singleton. (1999). Modeling the term structure of defaultable bonds. *The Review of Financial Studies*, **12**, pp. 687-720.

Elton, E., M. Gruber, D. Agrawal, and C. Mann. (2001). Explaining the rate spread on corporate bonds. *Journal of Finance*, **56**, pp. 247-277.

Fama, E.F., and K.R. French. (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, **51**, pp. 55-84.

Fan, J. (1992). Design-adaptive nonparametric regression. *Annals of Statistics*, **87**, 998-1004.

Fan, J. (1993). Local linear regression smoothers and their minimax efficiencies. *Journal of American Statistical Association*, **21**, 196-216.

Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear smoothers. *Annals of Statistics*, **20**, 2008-2036.

Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its applications.* Chapman and Hall, London.

Fan, J., Heckman, N.E. and Wand, M.P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Amer. Statist. Assoc.*, **90**, 141-150.

Fan, J. and Huang, L. (2001). Goodness-of-fit test for parametric regression models, *J. Amer. Statist. Assoc.*, **96**, 640-652.

Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. *Journal of American Statistical Association*, **99**, 710-723.

Fan, J., Yao, Q. and Cai, Z. (2003). Adaptive varying-coefficient linear models. *J. Royal Statist, Soc., Ser. B*, **65**, 57-80.

Fan, J., Zhang, C. and Zhang, J. (2001). Generalized likelihood ratio statistics and Wilks Phenomenon, *Ann. Statist.*. **29**, 153-193.

Fan, J. and Zhang, J. (2000), Functional linear models for longitudinal data, *J. Royal Statist, Soc., Ser. B*, **62**, 303-332.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying-coefficient models. *Annals of Statistics* **27**, 1491 - 1518.

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics* **27**, 715 - 731.

Field, C. B. (1991) Ecological scaling of carbon gain to stress and resource availability, In *Response of Plants to Multiple Stresses* (Ed. By H. A. Mooney, W. E. Winner and E. J. Pell), pp. 35-65, Academic Press, San Diego, Field, C., and Mooney, H. A. (1986). The photosynthesis-nitrogen relationship in wild plants, In *On the Economy of Plant Form and Function* (T. J. Givnish, ed.), pp. 25-55, Cambridge University Press, Cambridge.

Flanagan, L.B., Wever, L.A., and Carlson, P.J. (2002) Seasonal and interannual variation in carbon dioxide exchange and carbon balance in a northern temperate grassland, *Global Change Biolog*, **8**, 599-615.

Fridson, M.S., and J.G. Jonsson. (1995). Spread versus treasuries and the riskiness of high-yield bonds. *The Journal of Fixed Income*, **5**, pp. 79-88.

Gallant, A. N. (1987). *Nonlinear Statistical Models*, Wiley, New York.

Geske, R. (1977). The valuation of corporate liabilities as compound options. *Journal of Financial and Quantitative Analysis*, **12**, pp. 541-552.

Goetz, S.J. and Prince, S.D. (1999). Modeling terrestrial carbon exchange and storage: the evidence for and implications of functional convergence in light use efficiency, *Advances in Ecological Research*, 28, 57-92.

Gu, L., Baldocchi, D.D., Verma, S. B., Black, T. A., Vesala, T., Falge, E. M., Dowty, P.R. (2002). Superiority of diffuse radiation for terrestrial ecosystem productivity, *Journal of Geophysical Research*, 107.

Gutschick, V. P. (1987). *A Functional Biology of Crop Plants*, Croom Helm, London.

Härdle, W., Liang, H. and Gao, J. (2000). *Partially Linear Models*, Springer-Verlag, New York.

Hastie, T.J. and Tibshirani, R.J. (1993), Varying-coefficient models (with discussion), *J. Royal Statist. Soc., Ser. B*, **55** , 757-796.

Heckman, N. (1986). Spline smoothing in partly linear models, *J. Royal Statist. Soc. B* **48**, 244-248.

Helwege, J., and P.Kleiman. (1997). Understanding high-yield bond default rates. *The Journal of Fixed Income*, **5**, pp. 79-88.

Hollinger D.Y., Kelliher,F.M., Byers,J.N., Hunt, J.E., McSeveny,T.M., and Weir, P. L. (1994). Carbon dioxide exchange between an undisturbed old-growth temperate forest and the atmosphere, *Ecology*, 75, 134-150.

Hoover, D.R., Rice, J.A., Wu, C.O., and Yang, L.-P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85**, 809-822.

Huang, J.Z., and Kong, W.P., (2003). Explaining Credit Spread Changes: New Evidence from Option-Adjusted Bond Indexes. *Journal of Derivatives*,**Fall 2003**, 30-44.

Huang, J.Z., Wu, C. O. and Zhou, L. (2002). Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, **89**, 111-128.

Jarrow, R., D. Lando, and S. Turnbull. (1997). A markov model for the term structure of credit spreads. *The Review of Financial Studies*, **10**, pp. 481-523

Jarrow, R., and S. Turnbull. (2000). The intersection of market risk and credit risk. *Journal of Banking and Finance*, **24**, pp. 271-299.

Jarrow, R., and S. Turnbull. (1995). Pricing derivatives on financial securities subject to default Risk. *Journal of Finance*, **50**, pp. 53-86

Jiang, J. (1999). Conditional inference about generalized linear mixed models. *Ann. Statist.*, **27**, 1974–2007.

Kao, D.L. (1993). Estimating and pricing credit risk: An default risk in coupons affect the valuation of corporate bonds? *Financial Management*, **22**, pp. 117-131.

Lambers, H., Chapin, F. S. III, and Pons, T. L. (1998). *Plant Physiological Ecology*, Springer-Verlag, New York.

Leland, H., and K. Toft. (1996). Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *Journal of Finance*, **51**, pp. 987-1019.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22

Lin, D.Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussions). *J. Amer. Statist. Assoc.*, **96**, 103-126.

Longstaff, F., and E. Schwartz. (1995). A simple approach to valuing risky fixed and floating rate debt. *Journal of Finance*, **50**, pp. 789-820.

Madan, D., and H. Unal. (1998). Pricing the risks of default. *Review of Derivatives Research*, **2**, pp. 120-160.

Malinvaud, E. (1970). The consistency of nonlinear regression, *Ann. Math. Statist.*, **41**, 956-969.

Martinussen, T., and Scheike, T.H. (2001). Sampling adjusted analysis of dynamic additive regression models for longitudinal data. *Scandinavian Journal Of Statistics* **28**, 303 - 323.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models.* 2nd ed. Chapman and Hall, London.

Merrill Lynch Bond Index Rules and Definitions. Merrill Lynch & Co., October 2000.

Merton, R.C. (1974). On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, **29**, pp. 449-470.

Monteith, J.L., (1972). Solar radiation and productivity in tropical ecosystems, *Journal of Applied Ecology*, **9**, 747–766,

Newey, W., and K. West. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, **61** , pp. 631-653.

Newey, W., and K. West. (1987). A Simple Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55**, pp. 777-787.

Perelson, A.S., Essunger, P., Can, Y., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M., and Ho, D.D. (1997), Decay characteristics of HIV-1-infected compartments during combination therapy. *Nature*, **387**, 188-191.

Ruppert, D., Wand, M.P., Holst, U. and Hössjer, O. (1997). Local polynomial variance function estimation. *Technometrics*, **39**, 262-73.

Ruimy, A., Jarvis, P.G., Baldocchi, D.D. and Saugier, B. (1995). CO_2 fluxes over plant canopies: a literature review, *Advances in Ecological Research*, **26**, 1-68.

Ruimy, A., Kergoat, L., Bondeau, A., The. Participants OF. ThE. Potsdam NpP. (1999). Model intercomparison, comparing global models of terrestrial net primary productivity (NPP): analysis of differences in light absorption and light-use efficiency, *Global Change Biology*, **5** , 56-64.

Ruppert, (1997). Empirical-bias bandwidths for local polynomial nonparametric regression and density estimation. *J. Amer. Statist. Assoc.*, **92**, 1049-062.

Ruppert, D., Sheather, S.J. and Wand, M.P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, **90**, 1257-1270.

S&P Credit Indices: Overview and Methodology. New York: McGraw-Hill Co., 1999.

Saunders, A., and L. Allen. *Credit Risk Management*, 2nd ed. New York: John Wiley & Sons, 2002.

Scheike, T. H. and Martinussen, T. (2002). Nonparametric and semiparametric dynamic additive regression models. Technical Report, R-02-2008, Department of Mathematical Sciences, Aalborg University, Denmark.
Available at http://www.math.auc.dk/research/reports/R-02-2008.ps

Schmid, H. P., Grimmond, C. S. B., Cropley, F., Offerele, B. and Su, H.-B. (2000). Measurements of CO_2 and Energy Fluxes over a Mixed Hardwood Forest in the Midwestern United States, *Agricultural and Forest Meteorology*, **103**, 355-373.

Schwarz, G. (1978). Estimating the dimension of a model. *Ann. Statist.*, **6**, 461-464.

Seber, G.A.F and Wild, C. J. (1989). *Nonlinear Regression*, Wiley, New York.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. Royal Statist. Soc. B*, **50**, 413-436.

Van Horne, J. *Financial Market Rates and Flows*, 6th ed. Englewood Cliffs, NJ: Princeton-Hall, 2001.

Vassalou, M., and Y. Xing. Default risk in equity returns. Working paper, Columbia University, 2002.

Wofsy, S.C., Goulden, M.L., Munger, J.W., Fan, S.M., Bakwin, P.S., Daube, B.C., Bassow, S.L. and Bazzaz, F.A. (1993). Net exchange of CO_2 in a mid-latitude forest, *Science*, **260**, 1314-1317.

Wu, C.O., and Chiang, C.-T. (2000). Kernel smoothing on varying coefficient models with longitidual dependent variable, *Statistica Sinica*, **10**, 433-456.

Wu, C. O., Chiang, C.-T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.*, **93**, 1388-1402.

Wu, C.F. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.*, **9**, 501-513.

Yatchew, A. (1997). An elementary estimator of the partial linear model. *Economics Letters*, **57**, 135-143.

Yi, C., Li, R., Bakwin, P. S., Desai, A., Ricciuto, D. M., Burns, S., Turnipseed, A. A., Munger, J.W., Wofsy, S. C., Wilson, K., Meyers, T. P., Anderson, D. E., and Monson, R. K. (2004). A nonparametric method for separating photosynthesis and respiration components in CO_2 flux measurements. *Geophysical Research Letters.* **31**, L17107, doi:10.1029/2004GL020490

Zhang, W., Lee, S. and Song, X. (2002). Local polynomial fitting in semivarying coefficient models. *J. Multivariate Anal.*, **82** , 166-188.

Zhang, and W., Lee, S. (2000). Variable bandwidth selection in varying-coefficient models. *Journal of Multivariate Analysis* **74**, 116 - 134.

Vita

Yang Wang was born on June 18, 1975 in Changchun, Jilin Province, P.R. China. In 1998, he received his B.A. degree in Economics from People's University of China, Beijing, China. In 2001, he received his M.S. degree in Applied Mathematics from University of Dayton, Dayton, Ohio. He then came to the Pennsylvania State University (University Park), in 2001, to continue his Ph.D. study in Statistics.

During his graduate studies, he taught courses in advanced calculus, probability and statistics. He also worked as teaching assistant and research assistant. He had internship in Hershey Medical Center, Hershey, PA during summer of 2002, and internship in GlaxoSmithKline, Philadelphia, PA in 2005.

Throughout his dissertation research, he has been working with Dr. Runze Li and Dr. Jingzhi Huang on nonlinear varying-coefficient models and their applications. His research interests are varying-coefficient models, semiparametric models, and developing new predictive statistical models and applying their estimation and inference procedures to the fields of finance, economics, and pharmacology.