5-2011

Nonparametric Methods in Varying Coefficient Models And Quantile Regression Models

Chinthaka Kuruwita
Clemson University, chinthk@g.clemson.edu

Nonparametric Methods in Varying Coefficient Models And Quantile Regression Models

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Mathematical Sciences

by
Chinthaka Kuruwita
May 2011

Accepted by:
Dr. Karunarathna Kulasekera, Committee Chair
Dr. Colin Gallagher
Dr. Robert Taylor
Dr. William Padgett
Dr. Xiaoqian Sun

Abstract

This dissertation aims to address two problems in nonparametric regression models. An estimation issue in generalized varying coefficient models and a hypothesis testing issue in nonparametric quantile regression models is discussed.

We propose a new estimation method for generalized varying coefficient models where the link function is specified up to some smoothness conditions. Consistency and asymptotic normality of the estimated varying coefficient functions are established. Simulation results and a real data application demonstrate the usefulness of the new method.

A new approach for testing the equality of nonparametric quantile regression functions is also presented. Based on marked empirical processes, we develop test statistics that possess $\sqrt{n}$ properties in contrast to all available procedures in the literature. Asymptotic distributions are given and the performance of the proposed tests is compared with existing methods in mean regression and quantile regression. Theoretical results show that our tests have superior local power properties over existing tests. Finite sample performance is analyzed through simulations under a variety of settings. A data analysis is given which highlights the usefulness of the proposed methodology.

Dedication

To my parents and to my wife and daughter.

Acknowledgments

I would like to express my gratitude to many individuals who helped me in many ways during this work.

First and my foremost, I am indebted to my advisor Dr. K.B. Kulasekera for his guidance and support. I thank him for giving me the confidence to tackle difficult problems in my dissertation. His knowledge and experience in statistics helped me immensely to shape up my own line of thinking as a researcher.

I thank Dr. Colin Gallagher, for inspiring me to become a teacher like him (he's the best teacher I have had in my graduate courses). His support in proving several theorems and his guidance in improving my English is much appreciated.

I would also like to thank the Department of Mathematical Sciences for providing me financial support me during my stay at Clemson. I'm grateful for the help and support from Brian Dandurand and the Clemson 'Condor' team in resolving computational issues in this work.

Finally, I thank my parents and my wife and daughter for giving me unconditional love and support during my entire graduate life. Without their support I couldn't have done this.

Table of Contents

List of Tables

List of Figures

Chapter 1

Estimation in Varying Coefficient Models

The varying coefficient model has gained considerable interest since the pioneering work by Hastie & Tibshirani (1993). Its ability to model dynamical systems led to applications in many areas such as functional data modeling (Ramsay & Silverman, 1998), time series analysis (Huang & Shen, 2004), longitudinal data analysis (Fan *et al.*, 2007), survival analysis (Cai *et al.*, 2008) and nonparametric quantile regression (Cai & Xu, 2009).

Suppose we have independent and identically distributed observations (Y_i, X_i, U_i) following the generalized varying coefficient model

$$Y_i = g\{\sum_{k=1}^{p} \beta_k(U_i)X_{ik}\} + \epsilon_i \tag{1.1}$$

with $E(\epsilon_i \mid U_i, X_i) = 0$ $(i = 1, \ldots, n)$. Here $g(\cdot)$ is called the link function and $\beta_k(\cdot)$s are the varying coefficient functions. The covariate X_i is p-dimensional and U_i is a univariate random variable and is called the effect modifier or the index variable. Cai *et al.* (2000) studied the estimation and hypothesis testing of the varying coefficient functions in model (1.1) with a known link function. However, assuming a parametric form for the link function is very restrictive and misspecification of the link can result in large bias in the estimated varying coefficient functions. Therefore, it is desirable to have the link function unspecified in model (1.1), especially at the exploratory stage of modeling. This first part of the dissertation discusses nonparametric estimation of the varying coefficient functions in model (1.1) when the link function is unknown.

Regression models with unknown link functions have been studied by several authors in the context of generalized linear models and generalized additive models. Li & Duan (1989) discuss asymptotic properties of the estimated regression coefficients under link violation and lay out specific conditions needed to attain consistency of the estimated regression parameters. Weisberg & Welsh (1994) proposed to estimate the unknown link function using a Nadaraya–Watson kernel smoother and used an iterative weighted least squares method to estimate the regression coefficients. Extending this idea Chiou & Müller (1998) proposed a quasilikelihood approach and established the asymptotic normality of the regression parameters with unknown link and variance functions. Horowitz (2001) studied the estimation in a generalized additive model with an unknown link function and showed that the additive components and the link function can be estimated consistently. However, no work has been done to extend these methods to varying coefficient models with unknown links.

We introduce an estimation method that can be used to estimate the varying coefficient functions of model (1.1) with an unspecified link function. Our approach involves a simple localized least squares minimization which is non-iterative in the sense of estimate/update/re-estimate steps. It not only gives consistent estimates of the coefficient functions but also allows us to estimate the unknown link, which in itself is useful as it can suggest an appropriate parametric link function for the given data.

1.1 Estimation Method

For convenience, assume the coefficient functions $\beta_k(\cdot)$ $(k = 1, \ldots, p)$ of model (1.1) are defined on $[0, 1]$ with each $\beta_k(\cdot)$ having a continuous derivative. Our aim is to estimate these coefficient functions pointwise. Therefore, let $\eta_i(u_0)$ be the local constant approximation of the linear predictor $\eta(U_i) = \sum_{k=1}^{p} \beta_k(U_i)X_{ik}$ of model (1.1), for $0 \leq u_0 \leq 1$. Then $\eta_i(u_0) = \theta^{\mathrm{T}} X_i$, where $\theta = (a_1, \ldots, a_p)^{\mathrm{T}}$ and $X_i = (X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$. If the link function g was known, one could obtain local estimates of the coefficients by a straightforward weighted least squares minimization. Since g is unspecified in our model, a natural strategy is to estimate the link function nonparametrically. However, since the coefficient functions are unknown, the linear predictor is unknown, so we cannot simply smooth the responses against the linear predictor to estimate the link function. Noting that $\eta_i(u_0) = \theta^{\mathrm{T}} X_i$ is a local approximation of the linear predictor, an estimate of the link function can be obtained by smoothing $\{\theta^{\mathrm{T}} X_i, Y_i\}$ $(i = 1, \ldots, n)$ instead. For example, let $\hat{g}_{NW}(t, \theta) = \sum_{i=1}^{n} w_i Y_i$ be the Nadaraya–Watson estimator of the link function at t, where $w_i = K_h\left(\theta^{\mathrm{T}} X_i - t\right) / \sum_{j=1}^{n} K_h\left(\theta^{\mathrm{T}} X_j - t\right)$ with $K_h(\cdot) = K(\cdot/h)$ a symmetric kernel function. However, if we simply use a standard one dimensional smoother of this form, we are ignoring the fact that $\eta_i(u_0)$ is only a localized estimate of the true linear predictor for a given $U = u_0$, and as a result the link function estimate will exhibit very poor finite sample performance as shown in Fig. 1.1. Furthermore, such an estimate will be inconsistent. Suppose we know the true values $\theta_0 = \{\beta_1(u_0), \ldots, \beta_p(u_0)\}^{\mathrm{T}}$ of the coefficient functions of model (1.1) at $U = u_0$. Under suitable conditions on the smoothing parameter h we know that $\hat{g}_{NW}(t, \theta_0) \rightarrow E\left(Y \mid \theta_0^{\mathrm{T}} X = t\right)$ in probability as $n \rightarrow \infty$. However,

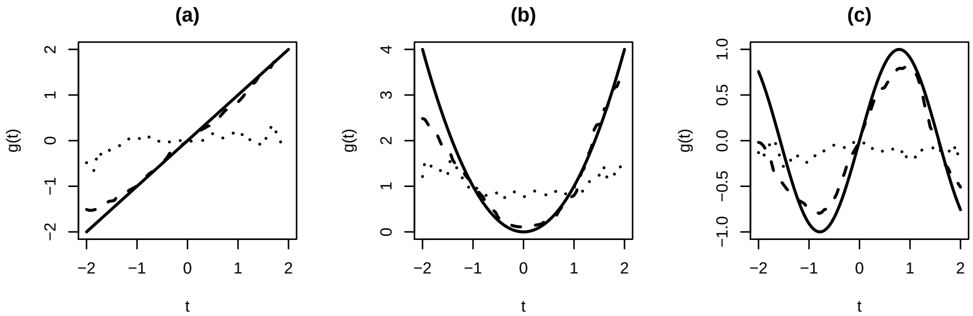

Figure 1.1: Link function estimates for (a) $g(t) = t$, (b) $g(t) = t^2$ and (c) $g(t) = sin(2t)$ with sample size $n = 200$. Simple one dimensional smoother $\hat{g}_{NW}$ (dotted). Proposed method (dashed).

$E(Y \mid \theta_0^{\mathrm{T}} X = t) \neq g(t) = E(Y \mid \theta_0^{\mathrm{T}} X = t, U = u_0)$. Therefore, in order to consistently estimate the link function, we need to ensure that only the observations close to u_0 are used in the smoothing process. To achieve this objective we use two kernels in a Nadaraya–Watson type estimator to get a localized estimate of the link function.

For t on T, the support of $\theta^{\mathrm{T}} X$, let $\hat{g}_{u_0}(t, \theta) = A_{n,u_0}(t, \theta)/B_{n,u_0}(t, \theta)$ where

$$A_{n,u_0}(t, \theta) = (nh_1h_2)^{-1} \sum_{j=1}^{n} Y_j K_1 \left(\frac{\theta^{\mathrm{T}} X_j - t}{h_1} \right) K_2 \left(\frac{U_j - u_0}{h_2} \right) \quad (1.2)$$

and $B_{n,u_0}(t, \theta)$ is $A_{n,u_0}(t, \theta)$ with $Y_j \equiv 1$ $(j = 1, \ldots, n)$. The kernel $K_1(\cdot)$ with smoothing parameter h_1 localizes the observations around the point of estimation t as in the Nadaraya–Watson estimator, while the kernel $K_2(\cdot)$ with smoothing parameter h_2 localizes the observations around u_0. The estimator $\hat{g}_{u_0}(t, \theta)$ is similar to that of Ichimura (1993). However, the weight function in his estimator does not depend on the sample size and is used as a means of handling heteroscedasticity. In contrast, the weights given by the extra kernel $K_2(\cdot)$ in our estimator depend on the sample size and serve the purpose of localizing the observations around the point u_0.

Given this estimator of the link function we use the following estimation procedure to estimate the varying coefficient functions.

Step 1: On a grid of u_0 values in $[0,1]$, minimize the localized least squares criterion

$$M_n(\theta) = (nh_2)^{-1} \sum_{i=1}^{n} \{Y_i - \hat{g}_{u_0}(\theta^{\mathrm{T}} X_i, \theta)\}^2 \, K_2\left(\frac{U_i - u_0}{h_2}\right) \tag{1.3}$$

to estimate $\beta_k(\cdot)$ $(k = 1, \ldots, p)$. For each u_0 the minimizer $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_p)$ of (1.3) is the estimate of $\beta_k(u_0)$. In our simulations and in the data example we normalize this by setting $\mathrm{sign}(\hat{\theta}_1)\hat{\theta}/||\hat{\theta}||$ to impose the identifiability restriction in condition A3 in appendix A. Here $\mathrm{sign}(x)$ denotes the algebraic sign of a real number x and $||x||$ denotes the Euclidean norm of a vector x.

Step 2: Once the coefficient functions are recovered, an improved estimate of the link function $\hat{g}(\cdot)$ is constructed in the following way. Noting that the model shares a common link function for each fixed u_0 let $\hat{\eta}(U_i) = \sum_{k=1}^{p} \hat{\beta}_k(U_i) X_i$ be an estimate of the linear predictor of model (1.1). Smoothing $\{\hat{\eta}(U_i), Y_i\}$ $(i = 1, \ldots, n)$ with a Nadaraya–Watson estimator we get

$$\hat{g}(t) = \frac{\sum_{i=1}^{n} K_1\left\{\frac{\hat{\eta}(U_i) - t}{h_1}\right\} Y_i}{\sum_{i=1}^{n} K_1\left\{\frac{\hat{\eta}(U_i) - t}{h_1}\right\}} .$$

***Remark* 1.** *In our estimation of the coefficient functions and the link function, we have used local constant approximation. However, one could easily use local linear smoothing in estimating the coefficient functions and the link function. For example, a local linear approximation of the linear predictor can be written as* $\eta_i^{\mathrm{LOL}}(u_0) = \gamma^{\mathrm{T}} Z_i$ *where* $\gamma = (a_1, \ldots, a_p, b_1, \ldots, b_p)^{\mathrm{T}}$ *and* $Z_i = \{X_{i1}, \ldots, X_{ip}, (U_i - u_0)X_{i1}, \ldots, (U_i - u_0)X_{ip}\}^{\mathrm{T}}$*. Substituting* γ *for* θ *and* Z_i *for* X_i *in (1.2) and (1.3) will yield local linear estimates of the coefficient functions.*

Bandwidth selection is an important aspect of any nonparametric estimation problem. For our estimation procedure we need to select two bandwidths h_1 and h_2. In our data example we use leave-one-out cross validation to select them, using the cross validation function

$$CV(h_1, h_2) = \sum_{i=1}^{n} (Y_i - \hat{Y}^{-i})^2$$

where $\hat{Y}^{-i}$ is the fitted value with the ith observation removed and using bandwidth combination (h_1, h_2). We minimize this function over a two-dimensional grid of bandwidth values and choose the bandwidth combination which yields the minimum cross validation score. When the sample size is large computing the leave-one-out cross validated bandwidths is time consuming. Therefore in our simulations we used a five-fold cross validation method. The k-fold cross validation is defined as follows. We partition the data into k groups, where the jth group consists of the data points $D_j = \{1 + (j-1)n/k, \ldots, jn/k\}$ $(j = 1, \ldots, k)$. Then we minimize

$$CV_k(h_1, h_2) = \sum_{j=1}^{k} \sum_{i \in D_j} (Y_i - \hat{Y}_i^{-D_j})^2$$

over a two-dimensional grid of bandwidth values. Here $\hat{Y}_i^{-D_j}$ is the fitted value of Y_i computed with observations in D_j removed.

1.2 Asymptotic Properties

For a specified link function $g(\cdot)$, Cai *et al.* (2000) established pointwise asymptotic normality of the estimated varying coefficient functions in model (1.1). Extending their work, we show that our estimated varying coefficient functions with an unspecified link are consistent and asymptotically normal. We have the following results.

THEOREM 1. *Under assumptions A1-A8 in appendix A, the minimizer* $\hat{\theta}$ *of (1.3) is a consistent estimator of* θ_0.

Proof. The proof is given in Appendix A □

THEOREM 2. *Under assumptions A1-A9 in appendix A, the minimizer* $\hat{\theta}$ *of (1.3) converges in distribution to a normal random variable with mean vector* θ_0 *and co-variance matrix* Σ_{u_0} *where* $\Sigma_{u_0} = \{M_2(\theta_0)\}^{-1} f_U(u_0)\nu_0\Delta(u_0)$ *and* $\nu_j = \int s^j K_2^2(s)ds$.

Proof. The proof is given in Appendix A □

***Remark* 2.** *The order of the bias in the coefficient function estimator is the same as that of Cai* et al. *(2000) and is given by (20) in appendix A. Due to the restrictions on the bandwidth sequences, the asymptotic bias of our coefficient function estimator becomes zero at the expense of a relatively slower convergence rate than that of the coefficient function estimator with known link. In practice one could use a method of moment estimator of* Σ_{u_0} *similar to Cai* et al. *(2000) for inference. In our data example we used bootstrap methods to estimate the standard errors of the coefficient function estimates.*

1.3 Empirical Study

1.3.1 Simulation Results

We generated 1000 random samples of size n responses from model (1.1) with three covariates and link functions $g(t) = t$, $g(t) = t^2$ and $g(t) = \sin(2t)$. The coefficient functions $\beta_1(\cdot)$, $\beta_2(\cdot)$ and $\beta_3(\cdot)$ were chosen to be the normalized versions of t^2+1, $\cos^2(\pi t)+0\cdot 5$ and $2\sin^2(\pi t)-0\cdot 5$ that satisfy the identifiability condition A3 in appendix A. We took X_1, X_2 and X_3 to be independent standard normal covariates, the effect modifier U to be a uniform random variable over $[0,1]$ and ϵ to be a normal random variable with mean zero and standard deviation $0\cdot 1$ independent of both X and U. We used the standard Gaussian kernel for $K_1(\cdot)$ and the Epanechnikov kernel $K(s)=0\cdot 75(1-s^2)_+$ as $K_2(\cdot)$. The performance of the estimated varying coefficient functions and the link function is assessed using average squared errors

$$\mathrm{ASE}_\beta = N_\beta^{-1}\sum_{k=1}^{3}\sum_{j=1}^{N_\beta}\left\{\hat{\beta}_k(u_j)-\beta_k(u_j)\right\}^2, \quad \mathrm{ASE}_g = N_g^{-1}\sum_{j=1}^{N_g}\left\{\hat{g}(t_j)-g(t_j)\right\}^2$$

respectively. Here $0\le u_j\le 1$ $(j=1,\ldots,N_\beta)$ and $-2\le t_j\le 2$ $(j=1,\ldots,N_g)$ are gridpoints at which we estimate the coefficient functions and the link function. We used $N_\beta=100$ and $N_g=200$.

As shown in Fig. 1.1, the use of a one dimensional smother such as the Nadaraya–Watson estimator in estimating the link function is inefficient. The ASE_g values, not presented here, confirm the need for additional localizing with respect to the index variable 'U'. In Table 1.1 one can clearly see the superior performance of our link function estimator with the additional localizing mechanism over the standard Nadaraya-Watson estimator in estimating the link function.

We compared the performance of our coefficient function estimator with that of Cai *et al.* (2000) using ASE_β values. Table 1.2 summarizes the mean ASE_β for known link and unknown link methods. Our estimator tends to have larger average squared error values compared to the estimator with known link, due to the additional estimation step involved in our method. Both estimators show a decrease in average squared error values as sample size increases.

In order to assess the pointwise variability of the coefficient function estimators

	Link function					
	$g(t) = t$		$g(t) = t^2$		$g(t) = sin(2t)$	
Sample size(n)	$\hat{g}$	$\hat{g}_{NW}$	$\hat{g}$	$\hat{g}_{NW}$	$\hat{g}$	$\hat{g}_{NW}$
100	0·314	158·257	4·694	117·607	1·861	33·458
200	0·200	109·730	2·023	89·854	0·495	18·353
400	0·090	76·070	0·547	89·241	0·187	20·608

Table 1.1: Comparison of average squared error(ASE_g) of the link function estimators based on the proposed method ($\hat{g}$) with the ones that are based on the Nadaraya-Watson estimator ($\hat{g}_{NW}$).

	Link function					
	$g(t) = t$		$g(t) = t^2$		$g(t) = sin(2t)$	
Sample size(n)	Known	Unknown	Known	Unknown	Known	Unknown
100	0·562	0·965	0·701	2·623	1·101	2·460
200	0·219	0·359	0·327	1·385	0·227	0·627
400	0·105	0·177	0·058	0·360	0·077	0·241

Table 1.2: Comparison of average squared error(ASE_β) of the coefficient function estimates from the proposed estimation method with unknown link and the estimation method with the known link.

we plotted the 25th and the 75th pointwise percentiles of the 1000 estimates of the coefficient functions. To save space, we only present the case for $\beta_1(\cdot)$ with quadratic link in Fig. 1.2.

1.3.2 Real Data Example

To illustrate our methodology we analyze the Japanese chemical industry data (Yafeh & Yosha, 2003) which is publicly available at the *Econometric Journal* website at `http://www.res.org.uk`. Data consists of various economic factors collected on 185 chemical firms listed in the Japanese stock market. The dependent variable Y is a measure of expenses on managerial private benefits. Three covariates are considered:

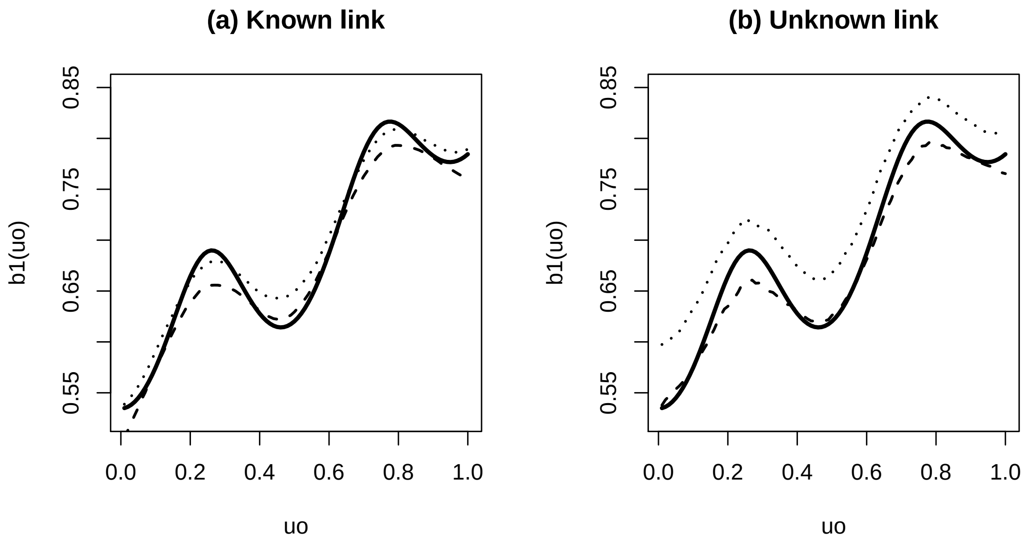

Figure 1.2: Pointwise 25^{th} and 75^{th} percentiles of the estimates of coefficient function $\beta_1(\cdot)$ with link function $g(t) = t^2$ and $n = 200$. True function(solid), 25^{th} percentile(dashed) and 75^{th} percentile(dotted).

the age of the firm; ownership concentration, which gives the percentage of ownership that belongs to the top 10 shareholders; and profit of the firm. They are denoted by AGE, TOP10 and PROFIT respectively. As our effect modifier U, we picked leverage, which is the ratio of debt to debt plus equity . This allows us to examine how the effects of the covariates on the response changes with the firm's debt/equity levels. We used leave-one-out cross-validation to select the two smoothing parameters h_1 and h_2.

Figure 1.3 (a)-(e) shows the estimated coefficient functions and the link function estimate together with 95% bootstrap confidence intervals based on 1000 bootstrap replications. All three confidence intervals of the coefficients exclude zero in most of the support which indicates that all three covariates have a significant impact on the response. Figure 1.3 (b) suggests that ownership concentration is significantly affecting the response in firms that have higher leverage. The link function estimate

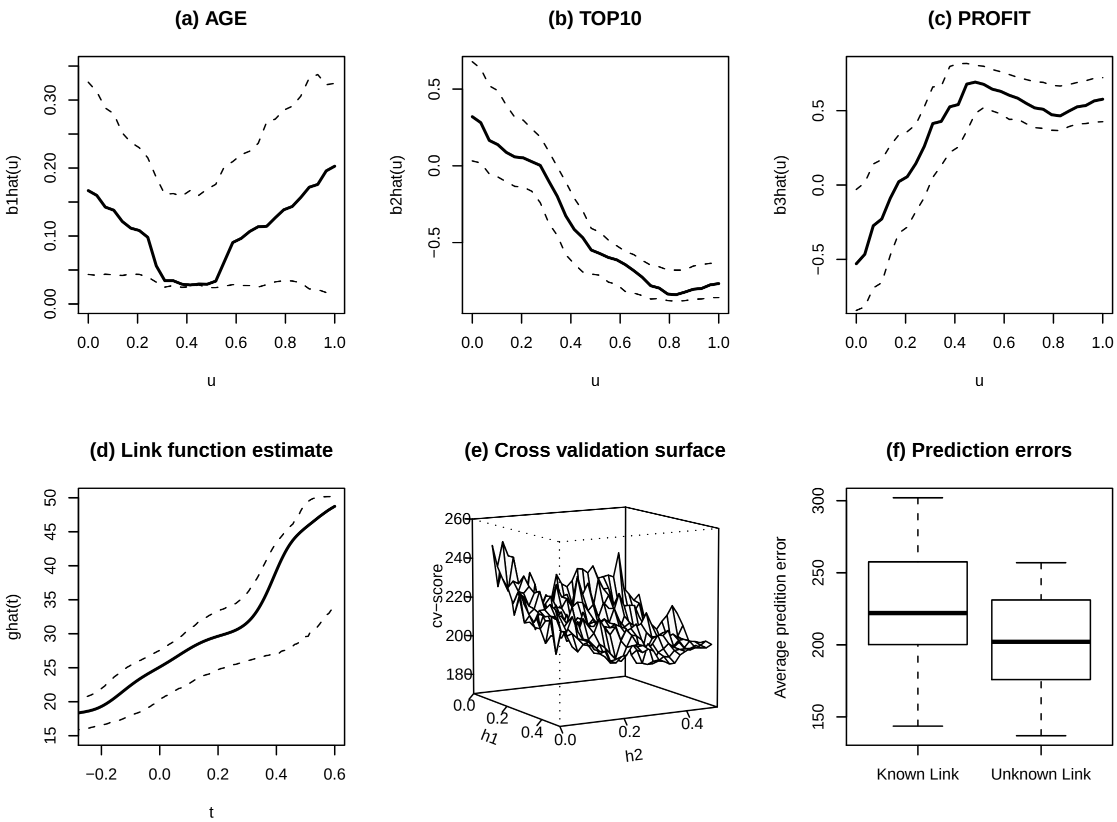

Figure 1.3: Analysis of Japanese chemical industry data. (a)-(c) Estimated coefficient functions , (d) link function estimate , (e) Cross-validation surface, (f) Prediction errors for known link and unknown link methods.

Fig. 1.3 (d) appears to be monotonically increasing for most of the support but nonlinear.

The natural link function for these data is the linear link. In order to assess the performance of our model with that of the known link method, we compared the average prediction errors of the two methods by partitioning the data into a training set of 150 observations and a test set of 35 observations. We performed this for 10 random partitions of the data, with leave-one-out cross validated bandwidth selection on each of the 10 partitions for both known link and unknown link methods. The average prediction errors summarized in Fig. 1.3 (f) show that our method exhibits a lower prediction error compared to the known link method. The confidence bounds for the link function, given in Fig. 1.3 (d), clearly exclude the linear link suggesting that the linear link seems to be an over-simplified assumption that leads to poor prediction in the known link method.

1.4 Concluding Remarks

Data analysis and inference based on an assumed link function, although convenient, could lead to erroneous results. Our proposed methodology can suggest suitable candidates for the link function that can be explored by the data analyst and therefore is useful as a diagnostic tool.

Chapter 2

Hypothesis Testing in Nonparametric Quantile Regression Models

Comparing groups based on independent samples has been a fundamental problem in statistics. In the context of regression analysis, this is typically done by comparing the conditional means of the groups. In parametric regression, comparison of model parameters of a predefined parametric model provides information about the differences among the groups. When a parametric specification is not appropriate, one compares nonparametric mean regression functions to discern any differences between groups (Hall & Hart, 1990; King *et al.*, 1991; Kulasekera, 1995; Neumeyer & Dette, 2003). The classical regression framework within which the above mentioned procedures are developed, requires the errors of the assumed model to have at least finite variance. In practice this assumption may not hold, especially with heavy tailed error distributions. Furthermore, all mean regression procedures are highly sensitive to extreme observations which can lead to spurious results. To overcome these difficulties in mean regression procedures, Koenker & Bassett (1978) in their pioneering work proposed the quantile regression framework. In that, one regresses the conditional quantiles of a response on covariates rather than regressing the conditional mean. In this article we develop a flexible and robust testing procedure to compare groups within the quantile regression framework. Our method uses nonparametric quantile regression functions to construct test statistics that can detect differences at targeted quantiles in two or more conditional distributions.

Suppose we observe data from k independent groups in the form $(X_{ij}, Y_{ij}), i = 1 \ldots k, j = 1 \ldots n_i$ where X is a continuous covariate supported on $[0, 1]$ and Y is a continuous response. For $\tau \in (0, 1)$ let $g_\tau(X)$ denote the τ^{th} conditional quantile function of Y given X (i.e., $P[Y \leq g_\tau(X)|X] = \tau$). We model $g_\tau(\cdot)$ nonparametrically with the following nonparametric quantile regression model:

$$Y_{ij} = g_{\tau,i}(X_{ij}) + \epsilon_{ij}, \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i \tag{2.1}$$

where $g_{\tau,i}$ is the τ^{th} conditional quantile function of Y given X for the i^{th} group and ϵ_{ij} is a sequence of independent random variables assumed to be identically distributed within each group. We further, assume that the τ^{th} conditional quantile of ϵ_{ij} given X_{ij} is zero. The hypothesis of interest is, for fixed $\tau \in (0, 1)$ whether the conditional quantile curves are the same across the k groups or not. That is, we are interested in the testing problem

$$H_0 : g_{\tau,i}, \ldots, g_{\tau,k} = g_\tau \quad \text{vs} \quad H_1 : g_{\tau,i} \neq g_{\tau,i'} \quad \text{for some } i \neq i'$$

over the support of the covariate X.

To our knowledge, the only investigation into comparison of nonparametric quantile regression functions in the form of the above hypothesis is by Sun (2006). Her test is based on an orthogonal moment condition of residuals which holds under the null hypothesis. The resulting test has non-trivial power against local alternatives that converge to the null at a rate $(\sqrt{N}h^{1/4})^{-1}$ with scalar covariate. Here $N = \sum_{i=1}^{k} n_i$ is the total sample size and h is the bandwidth (converging to 0) that is used to estimate the common nonparametric quantile regression function assuming the null hypothesis is true. In contrast, we propose a new test procedure that can detect alternatives converging to the null at $N^{-1/2}$ rate. Our method is based on a marked empirical process of residuals. We note that marked empirical processes based on residuals have been used in the context of testing the equality of mean regression functions by Delgado (1993), Kulasekera (1995) and Neumeyer & Dette (2003).

The remainder of this article is organized as follows. Section 2 describes the test procedure and presents the asymptotic properties. Section 3 reports the results of a simulation study that investigates the finite sample performance of our method together with a real data example. Concluding remarks are given in Section 4. Proofs of theoretical results are given in the Appendix.

2.1 Tests and Properties

For fixed $\tau \in (0, 1)$ consider the nonparametric quantile regression model in (2.1). Let $\eta_{ij} = I\{Y_{ij} < g_{\tau,i}(X_{ij})\} - \tau$ and $U_{ij} = I\{Y_{ij} < g_\tau(X_{ij})\} - \tau$. Here $I(\cdot)$ denotes the indicator function and g_τ is a weighted average of $g_{\tau,i}$s defined as

$$g_\tau(t) = \sum_{i=1}^{k} \lambda_i(t) g_{\tau,i}(t) \tag{2.2}$$

with $\lambda_i(t) \in (0,1)$ and $\sum_{i=1}^k \lambda_i(t) = 1$ (Sun, 2006). Under H_0, we see that $g_\tau \equiv g_{\tau,i}$. Note that U_{ij} can be written as

$$U_{ij} = \eta_{ij} + I\left\{Y_{ij} < g_\tau(X_{ij})\right\} - I\left\{Y_{ij} < g_{\tau,i}(X_{ij})\right\}.$$

Then, under H_0, we see that $U_{ij} = \eta_{ij}$ and are independent mean zero random variables. Now consider the process

$$R_i(t) = \frac{1}{N}\sum_{j=1}^{n_i} U_{ij} I(X_{ij} \leq t) \tag{2.3}$$

for $t \in (0,1)$. Under H_0, $R_i(\cdot)$ is a mean zero random process. If the null hypothesis is false, $R_i(\cdot)$ will have a non-zero mean function which can be used as a basis to detect departures from H_0. To construct test statistics based on process (2.3), we need an efficient estimator of $R_i(t)$ under H_0. Estimation of $R_i(t)$ under H_0 solely depends on estimating the quantile regression functions $g_{\tau,i}(\cdot)$. If H_0 is true, all $g_{\tau,i}(\cdot)$'s are the same across the k groups. Therefore, an efficient nonparametric estimator of the common quantile regression function $g_\tau(\cdot)$ can be constructed by pooling the data from all groups. For independent data, several nonparametric quantile regression function estimators are available in the literature (Yu & Jones, 2003; Dette & Volgushev, 2008; Bondell *et al.*, 2010). In this paper we use the local linear nonparametric quantile regression function estimator proposed by Yu & Jones (2003). Local linear estimators are popular in practice due to good finite sample performance at boundaries of the support (Fan & Gijbels, 1996). The later estimators are proposed to eliminate the 'quantile crossing' problem in estimating quantile regression functions at multiple quantiles. Since our testing problem, hence our estimation, is only for a fixed $\tau \in (0,1)$, this problem does not affect our method.

Let $\hat{g}_\tau(\cdot)$ be the local linear nonparametric quantile regression function estimator of the common quantile regression function $g_\tau(\cdot)$. We define $\hat{g}_\tau(x) = \hat{a}$ where $\hat{a}$ and $\hat{b}$ minimize

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i}\rho_\tau\Big(Y_{ij} - a - b(X_{ij} - x)\Big)K\left(\frac{X_{ij} - x}{h}\right). \tag{2.4}$$

Here $\rho_\tau(u) = u\{\tau - I(u \leq 0)\}$ is called the "check function" and h is a smoothing parameter and $K(\cdot)$ is a symmetric density function on $[-1, 1]$. For $i = 1, \ldots, k$, define the marked empirical process

$$\hat{R}_i(t) = \frac{1}{N}\sum_{j=1}^{n_i}\hat{U}_{ij}I(X_{ij} \leq t) \tag{2.5}$$

where $\hat{U}_{ij} = I[Y_{ij} \leq \hat{g}_\tau(X_{ij})] - \tau$. The process $\hat{R}_i(\cdot)$ in (2.5) is the building block of our test statistics. We will first show that $\hat{R}_i(\cdot)$ converges to a mean zero Gaussian process in the space $D[0, 1]$ under H_0. Then by applying the continuous mapping theorem, one can use functionals of $\hat{R}_i(\cdot)$ to construct test statistics to test the equality of nonparametric quantile regression functions.

THEOREM 3. *Let conditions B1-B5 in appendix B hold. Then under H_0 the marked empirical processes $\sqrt{N}\hat{R}_i(t)$ for $i = 1, \ldots, k$ converge weakly to independent mean zero Gaussian processes with covariance functions $H_i(t, s) = \tau(1-\tau)c_i F_{X_i}(t),\ t < s$ in the space $D[0, 1]$ as $N \to \infty$.*

Proof. The proof is given in Appendix B. □

Based on Theorem 3, many test statistics can be constructed to test the equality of quantile regression curves. In this article we investigate three such test statistics. For $i, j = 1, \ldots, k$, define

1. $T_i = \sup_{0 \le t \le 1} |\sqrt{N}\hat{R}_i(t)|$,

2. $T_{ij} = \sup_{0 \le t \le 1} |\sqrt{N}\hat{R}_i(t) - \sqrt{N}\hat{R}_j(t)| \quad i \neq j$.

3. $T_M = \max\left\{ \sup_{0 \le t \le 1} |\sqrt{N}\hat{R}_1(t)|, \ldots, \sup_{0 \le t \le 1} |\sqrt{N}\hat{R}_k(t)| \right\}$,

For a size α test, we reject H_0 for values of the above statistics exceeding the $(1-\alpha)$ quantiles of their respective null distribution.

2.1.1 Computing Critical values

In order to compute critical values for our tests, we use two methods. The first method uses the asymptotic null distribution of the test statistics and the other uses a bootstrap resampling method. First, we will introduce a theorem that will be used to obtain critical values for tests 1 and 2 based on asymptotic null distributions.

THEOREM 4. *Let conditions B1-B5 in appendix B hold. Then as $N \to \infty$ and under H_0,*

$$\sup_{0 \le t \le 1} |\sqrt{N}\hat{R}_i(t)| \xrightarrow{D} \sup_{0 \le t \le 1} |W(t)|$$

where $W(t)$ is the standard Brownian motion on $[0,1]$.

Proof. The proof is given in Appendix B. □

Theorem 4 allows us to compute critical values for tests 1 and 2 by using the available results on the suprema of standard Brownian motion (Billingsley, 1968). For test 3, in general, it is not possible to use the asymptotic null distribution approach to compute its critical values. To see this, consider the covariance of the process $\sqrt{N}\hat{R}_i(t) - \sqrt{N}\hat{R}_j(t)$ under H_0. Using similar arguments as in Theorem 6, we can show that, under H_0, this process converges to a mean zero Gaussian process with covariance function $H_{ij}(t,s)$,

$$\tau(1-\tau)\{c_i F_{X_i}(t) + c_j F_{X_j}(t)\}, t < s.$$

Therefore, with unknown design densities it is difficult to compute critical values for test 3. However, for the special case of unknown but equal design densities, we can still use arguments in Theorem 4 to obtain critical values for test 3. This can be seen by noting that for equal designs $F_{X_i} = F_{X_j} = F_X$, the above covariance reduces to $\tau(1-\tau)(c_i + c_j)F_X(t)$.

One well-known problem in using asymptotic critical values is poor finite sample performance. Especially with test statistics derived from empirical processes, one can observe poor finite sample behavior due to slow rates in process convergence. To overcome these difficulties we use a modified version of the wild bootstrap proposed by Sun (2006) to compute critical values for our test statistics. This modified version ensures that the τ^{th} quantile of the bootstrap residuals conditional on the covariates is zero as required by our model (2.1).

2.1.2 Local Power

This section establishes the non-trivial power of our test statistics under local alternatives of the form

$$H_1 : g_{\tau,i}(\cdot) = g_\tau(\cdot) + \Delta_i(\cdot)/\sqrt{N}, i = 1, \ldots, k. \tag{2.6}$$

The following theorem shows the consistency of the proposed test statistics.

THEOREM 5. *Let conditions B1-B5 in appendix B hold. Then under local alter-*

natives of the form in (2.6), the marked empirical process $\sqrt{N}\hat{R}_i(\cdot)$ *for* $i = 1, \ldots, k$ *converges weakly to a Gaussian process in the space* $D[0,1]$ *with a mean function*

$$\mu(t) = c_i \int_0^t \lambda_i(x)\Delta_i(x) f_{\epsilon_i|X_i=x}(0) f_{X_i}(x)dx > 0$$

and covariance function $H_i(t,s) = \tau(1-\tau)c_i F_{X_i}(t), t < s$*, where* $\lambda_i(x)$ *is defined in (2.2).*

Proof. The proof is given in Appendix B. □

2.2 Simulation Results and Data Analysis

In this section we discuss the results of our simulation study and the data analysis. The simulation study consists of three parts. First, we investigated the Type 1 error and power properties of our test statistics and compared our method with that of Sun (2006). Next we examined local power properties of our test statistics with that of Sun (2006) which is the only available procedure in the literature in the context of our paper. Finally, we study how our test procedure can be used as an alternative to nonparametric mean regression procedures in comparing groups, especially in the presence of extreme observations. We compared our approach at $\tau = 0.5$(conditional median) with the nonparametric mean regression test proposed by Neumeyer & Dette (2003). Both Type 1 error and power of the two methods were investigated with 100% normal errors and with a mixture of 80% normal and 20% Cauchy errors.

2.2.1 Analysis of Type I Error and Power

We generated data from the following heteroscedastic error model

$$Y_{ij} = g_i(X_{ij}) + \sigma(X_{ij})\epsilon_{ij} \quad i = 1, \ldots, k, \quad j = 1, \ldots, n_i \tag{2.7}$$

with $X \sim U[0, 1]$ and $\epsilon \sim N(0, 1)$. We used a variety of functions for g_i and $\sigma(x)$ in our simulation. It is easy to see that the τ^{th} conditional quantile function for model (2.7) is of the form $g_{\tau,i}(\cdot) = g(\cdot) + \sigma(\cdot)Z_\tau$ where Z_τ is the τ^{th} conditional quantile of ϵ given X. This setup results in quantile curves having various degrees of smoothness. For example, with $g(x) = x$ and $\sigma(x) = cos(2\pi x) + 2$, we get the median being linear and more curvature at extreme quantiles as shown in Figure 2.2.1. We conducted our simulations for the case of comparing three groups ($k = 3$). All simulations were repeated 1000 times and we used equal sample sizes for each group. Three sample sizes $n = 25, 50, 100$ were considered and five quantiles $\tau = .05, .25, .5, .75, .95$ were examined.

As mentioned in section 2.1, we used two methods to compute critical values of our test statistics: Asymptotic null distribution approach and bootstrap approach. We discussed the asymptotic null distribution approach in section 2.1. For the bootstrap method, we followed the procedure outlined in Sun (2006). This method ensures that the τ^{th} conditional quantile of the bootstrap residuals given the data is zero, while the second and the third moments are matched with that of the true residuals. This method is a modified version of the wild bootstrap of Hardle & Mamen (1993). Let $\hat{\epsilon}_{ij} = Y_{ij} - \hat{g}_\tau(X_{ij})$ denote the estimated residuals. Here $\hat{g}_\tau(\cdot)$ is the estimated common quantile regression function using (2.4). Let ϵ^b_{ij} denote the bootstrap resamples of $\hat{\epsilon}_{ij}$ constructed according to the procedure in Sun (2006). Then we create the bootstrap

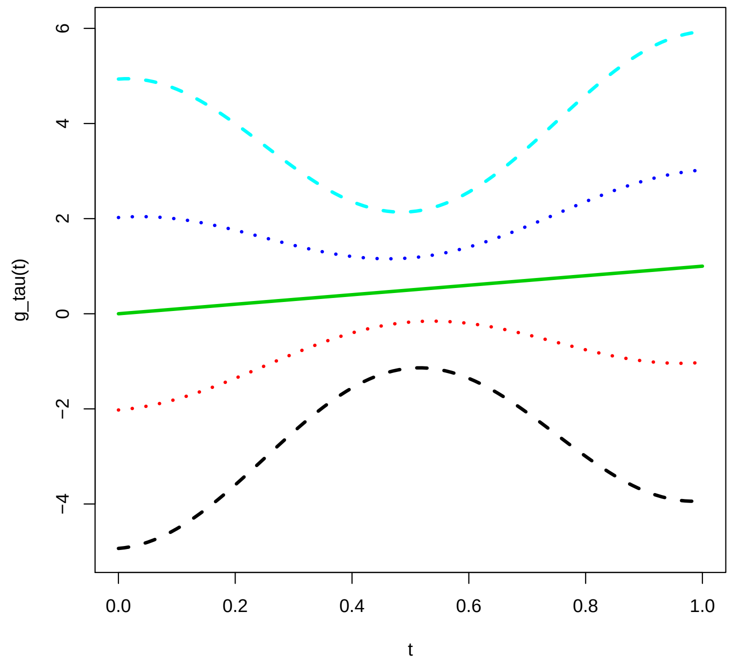

Figure 2.1: Quantile regression curves correspond to the heteroscedastic model in (2.7) with $g(t) = t$ and $\sigma(t) = cos(2\pi t) + 2$. Solid line($\tau = 0.5$), dotted lines($\tau = 0.25, 0.75$) and dashed lines($\tau = 0.05, 0.95$).

samples $\{X_{ij}, Y_{ij}^b\}$, where

$$Y_{ij}^b = \hat{g}_\tau(X_{ij}) + \epsilon_{ij}^b, i = 1, \ldots, k,\ j = 1, \ldots, n_i.$$

Using these bootstrap samples we then construct bootstrap replicates of our test statistics. We conducted B=200 repetitions of bootstrap replications to obtain $(1-\alpha)$ quantiles of the null distributions of the respective test statistics.

Since we are comparing our test procedure with that of Sun (2006), we adopted the bandwidth selection method proposed in Sun (2006) to ensure a better comparison of the two testing approaches. Therefore we used $h = \hat{\sigma}_X N^{-1/5}$ as our smoothing parameter in local linear estimation of the quantile regression functions. Here $\hat{\sigma}_X$ is the sample standard deviation of the covariate in the pooled sample. When constructing bootstrap versions of the test statistics, we need to estimate the quantile regression functions for each of the bootstrap samples. To improve finite sample performance of this bootstrap procedure Sun (2006) suggests to over-smooth the nonparametric estimates of the quantile regression functions constructed using the bootstrap samples. Let h_b denotes the bandwidth used in the estimation of the quantile regression function with bootstrap samples. As suggested in Sun (2006), we used $h_b = \hat{\sigma}_X N^{-1/9}$ in our simulations.

Table 2.1: Type I Error based on 1000 simulated samples

Functions	n	τ	T_M	T_1	T_{12}	SUN
$g_1 = g_2 = g_3 = exp(t)$	25	0.05	0.045	0.033	0.034	0.039
$\sigma_1 = \sigma_2 = \sigma_3 = 1 + t$		0.25	0.042	0.044	0.057	0.042
		0.50	0.044	0.054	0.062	0.048
		0.75	0.038	0.041	0.058	0.040
		0.95	0.033	0.026	0.028	0.042
	50	0.05	0.045	0.042	0.042	0.055
		0.25	0.051	0.050	0.069	0.037
		0.50	0.036	0.034	0.048	0.036
		0.75	0.044	0.039	0.047	0.030
		0.95	0.040	0.017	0.040	0.042
	100	0.05	0.050	0.050	0.043	0.047
		0.25	0.044	0.052	0.064	0.041
		0.50	0.032	0.043	0.053	0.032
		0.75	0.041	0.036	0.057	0.027
		0.95	0.050	0.043	0.051	0.032
$g_1 = g_2 = g_3 = exp(t)$	25	0.05	0.060	0.043	0.042	0.055
$\sigma_1 = \sigma_2 = \sigma_3 = 1 + t + cos(2\pi t)$		0.25	0.049	0.037	0.062	0.037
		0.50	0.032	0.044	0.056	0.048
		0.75	0.030	0.037	0.051	0.027
		0.95	0.032	0.011	0.024	0.041
	50	0.05	0.054	0.031	0.050	0.046
		0.25	0.040	0.033	0.078	0.027
		0.50	0.030	0.033	0.042	0.030
		0.75	0.023	0.022	0.065	0.016
		0.95	0.017	0.009	0.043	0.048
	100	0.05	0.033	0.030	0.063	0.024
		0.25	0.023	0.026	0.067	0.021
		0.50	0.033	0.040	0.055	0.022
		0.75	0.012	0.018	0.052	0.008
		0.95	0.035	0.037	0.067	0.027

Table 2.2: Empirical Power based on 1000 simulated samples

Functions	n	τ	T_M	T_1	T_{12}	T^*_{12}	SUN
$g_1(t) = t$	25	0.05	0.043	0.014	0.010	0.028	0.144
$g_2(t) = exp(t)$		0.25	0.783	0.204	0.496	0.542	0.881
$g_3(t) = sin(2\pi t)$		0.50	0.975	0.398	0.948	0.976	0.979
		0.75	0.937	0.630	0.979	0.981	0.911
		0.95	0.078	0.000	0.101	0.064	0.145
	50	0.05	0.153	0.025	0.022	0.020	0.142
		0.25	0.996	0.396	0.856	0.890	1.000
		0.50	1.000	0.756	1.000	1.000	1.000
		0.75	1.000	0.963	1.000	1.000	1.000
		0.95	0.371	0.002	0.574	0.473	0.289
	100	0.05	0.548	0.037	0.058	0.042	0.473
		0.25	1.000	0.704	0.993	0.998	1.000
		0.50	1.000	0.972	1.000	1.000	1.000
		0.75	1.000	0.999	1.000	1.000	1.000
		0.95	0.847	0.683	0.956	0.953	0.696

NOTE: T^*_{12} uses critical values from the asymptotic distribution

2.2.2 Local Power Properties

In this section we examine the power properties of our test procedure with local alternatives converging to the null at $(\sqrt{N})^{-1}$ rate. We compare our results to that of Sun (2006) . We restricted our simulation to the $k = 2$ case to save computational time. We used model (2.7) with $g_1(t) = e^t, g_2(t) = g_1(t) + \delta/\sqrt{N}$ and $\sigma(t) = c$ (homoscedastic errors). Here δ is a constant that controls the separation of the two functions and we chose $\delta = 0.3$ in our simulation. The sample sizes for each group were chosen large enough to make the testing problem difficult. Three sample sizes $n_1 = n_2 = 500, 2500, 5000$ were considered and three quantiles $\tau = .05, .50, .75$ were examined. All simulations were repeated 1000 times and the results are given in Table 2.3.

Table 2.3: Comparison of Power with Local Alternatives with α=0.05

τ	N	T_M	T_1	T_{12}	T_{12}^*	SUN
0.05	1000	0.244	0.285	0.285	0.265	0.072
	5000	0.258	0.287	0.292	0.272	0.050
	10000	0.273	0.308	0.312	0.296	0.050
0.50	1000	0.431	0.532	0.635	0.624	0.153
	5000	0.301	0.425	0.662	0.647	0.077
	10000	0.242	0.358	0.669	0.667	0.032
0.75	1000	0.417	0.519	0.588	0.556	0.134
	5000	0.299	0.415	0.584	0.560	0.073
	10000	0.266	0.384	0.618	0.591	0.050

NOTE: T_{12}^* uses critical values from the asymptotic distribution

2.2.3 Comparison with Nonparametric Mean Regression

As described in section 1, the mean regression approach, although quite popular in comparing groups, will have severe drawbacks in the presence of extreme observations. In this section we compare our testing procedure at $\tau = 0.5$(conditional median) with the nonparametric mean regression test proposed by Neumeyer & Dette (2003) under two error structures. We used model (2.7) with $\epsilon \sim N(0,1)$ and $\epsilon \sim 0.8N(0,1) + 0.2Cauchy$. The first error structure mimics the classical regression setup and the second error structure creates outliers by having 20% of the errors come from the standard Cauchy distribution. We restricted the study to the $k = 2$ case with $\sigma(t) = 0.5$ to be compatible with the simulation study of Neumeyer & Dette (2003).

Table 2.4: Type I Error and Power with 100% Normal Errors

$Functions$	n	T_M	T_1	T_{12}	T_{12}^*	SUN	ND
$g_1(t) = g_2(t) = exp(t)$	25	0.032	0.036	0.059	0.053	0.028	0.053
	50	0.028	0.034	0.053	0.043	0.022	0.067
	100	0.035	0.044	0.062	0.058	0.033	0.059
$g_1(t) = g_2(t) = sin(2\pi t)$	25	0.029	0.029	0.040	0.039	0.014	0.055
	50	0.028	0.029	0.060	0.038	0.003	0.056
	100	0.013	0.019	0.076	0.058	0.001	0.040
$g_1(t) = exp(t)$, $g_2(t) = g_1(t) + t$	25	0.502	0.557	0.625	0.655	0.500	0.915
	50	0.869	0.902	0.942	0.920	0.852	0.995
	100	0.997	0.990	0.999	0.998	0.999	1.000
$g_1(t) = sin(2\pi t)$, $g_2(t) = g_1(t) + t$	25	0.435	0.433	0.592	0.608	0.350	0.754
	50	0.759	0.721	0.903	0.893	0.558	0.970
	100	0.974	0.952	0.996	0.999	0.814	1.000
$g_1(t) = exp(t)$, $g_2(t) = g_1(t) + sin(2\pi t)$	25	0.156	0.134	0.264	0.271	0.679	0.385
	50	0.473	0.321	0.623	0.597	0.946	0.840
	100	0.928	0.778	0.970	0.974	1.000	0.997
$g_1(t) = sin(2\pi t)$, $g_2(t) = 2sin(2\pi t)$	25	0.212	0.038	0.203	0.195	0.433	0.211
	50	0.533	0.042	0.526	0.501	0.652	0.536
	100	0.878	0.099	0.912	0.953	0.880	0.951

NOTE: T_{12}^* uses critical values from the asymptotic distribution

Table 2.5: Type I Error and Power with 80% normal and 20% Cauchy errors

Functions	n	T_M	T_1	T_{12}	T_{12}^*	SUN	ND
$g_1(t) = g_2(t) = exp(t)$	25	0.030	0.033	0.045	0.054	0.041	0.033
	50	0.035	0.030	0.053	0.039	0.040	0.039
	100	0.032	0.042	0.063	0.052	0.021	0.027
$g_1(t) = g_2(t) = sin(2\pi t)$	25	0.041	0.037	0.050	0.050	0.021	0.030
	50	0.024	0.041	0.056	0.037	0.009	0.033
	100	0.009	0.020	0.057	0.036	0.000	0.031
$g_1(t) = exp(t)\ ,\ g_2(t) = g_1(t) + t$	25	0.385	0.446	0.505	0.542	0.345	0.261
	50	0.753	0.791	0.848	0.826	0.697	0.280
	100	0.971	0.979	0.989	0.992	0.961	0.271
$g_1(t) = sin(2\pi t)\ ,\ g_2(t) = g_1(t) + t$	25	0.343	0.362	0.454	0.480	0.241	0.229
	50	0.651	0.617	0.802	0.778	0.429	0.272
	100	0.927	0.889	0.990	0.989	0.647	0.279
$g_1(t) = exp(t)\quad g_2(t) = g_1(t) + sin(2\pi t)$	25	0.109	0.109	0.198	0.204	0.488	0.084
	50	0.354	0.243	0.473	0.452	0.840	0.112
	100	0.807	0.597	0.881	0.895	0.994	0.135
$g_1(t) = sin(2\pi t)\quad g_2(t) = 2sin(2\pi t)$	25	0.136	0.043	0.177	0.162	0.284	0.061
	50	0.403	0.041	0.411	0.393	0.493	0.092
	100	0.769	0.074	0.811	0.844	0.748	0.120

NOTE: T_{12}^* uses critical values from the asymptotic distribution

2.2.4 Data Analysis

We now present an application of our test procedure to a data example. We analyzed the Japanese chemical industry data (Yafeh and Yosha 2003) which is publicly available at the *Econometric Journal* website at `http://www.res.org.uk`. Data consists of various economic factors collected on 185 chemical firms listed in the Japanese stock market. The dependent variable Y is a measure of expenses on managerial private benefits. The study was focused on investigating the hypothesis that “concentrated shareholding is associated with lower expenditure on managerial private benefits”. Ownership concentration of the firm is measured by the variable TOPTEN, which gives the percentage of ownership that belongs to the top 10 shareholders. Our investigation was to check whether this association is different among firms that have high/low leverage (the ratio of debt to debt plus equity). We divided our data into two groups based on the leverage measurements of the firms. Firms with leverage values above the sample median are categorized into ‘high’ and the rest are into ‘low’ groups.

Figure 2.2 depicts how Y changes with TOPTEN in the two groups. As shown

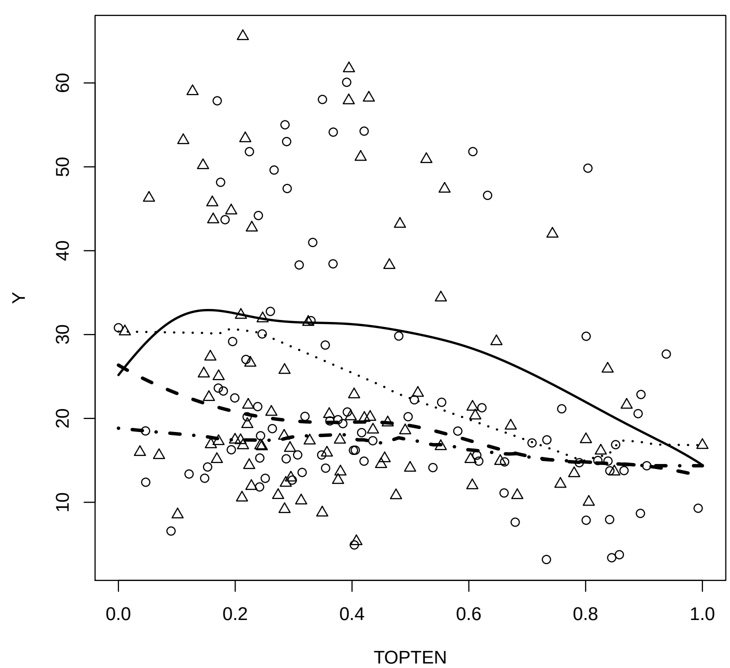

Figure 2.2: Plot of managerial expenses vs shareholder concentration of Japanese Chemical firms. Circles-low leverage, triangles-high leverage. The plotted lines are the nonparametric mean and median regression estimates for the two groups. — mean reg (low), $\cdots$ mean reg (high), - - - q- reg (low) , · - · q-reg (high).

in Figure 2.2, there is a negative association between expenditure on managerial private benefits and the ownership concentration in both groups. The nonparametric mean regression estimates (solid and dotted lines) suggest that the two groups are different. A hypothesis test (Neumeyer & Dette, 2003) revealed that the two mean regression curves are in fact significantly different at .05 level. We then used our test procedure and the one proposed by Sun (2006) at $\tau = 0.5$ to check whether the conditional medians of the two groups are the same or not. All of our test statistics were able to capture a difference between the two conditional medians. However, the test proposed in Sun (2006) did not reject the null hypothesis of equality of conditional median curves. We used the bandwidth selection methods which we adopted in our simulations. For computing bootstrap critical values, we used B=1000 repetitions to increase accuracy. Table 2.6 summarizes the results of our testing at several conditional quantiles. This analysis suggests that the differences between two conditional distributions exists at mid to upper quantiles. No differences were observed at quantiles below the median. It provides a more elaborate picture of the actual differences in the two groups compared to the nonparametric mean regression approach of Neumeyer & Dette (2003).

Table 2.6: Test results of the Japanese chemical industry data. α=0.05

τ	T_M	T_1	T_{12}	SUN
0.05	N	N	N	N
0.25	N	N	N	N
0.50	R	R	R	N
0.75	R	R	R	R
0.95	R	R	R	R

NOTE: R-Rejected , N- did not reject.

Appendices

Appendix A Technical Details for Varying Coefficient Model

Consistency of the Link Function Estimator

In what follows we will establish consistency results regarding this link function estimator and the coefficient function estimator. To facilitate our arguments we impose the following technical conditions.

A1: The link function g and the coefficient functions $\beta_k(\cdot)$ $(k = 1, \ldots, p)$ are three times continuously differentiable and g is non-constant on the support of $\theta^{\mathrm{T}} X$.

A2: The point $\theta_0 = \{\beta_1(u_0), \ldots, \beta_p(u_0)\}^{\mathrm{T}}$ is an interior point of a compact set Θ.

A3: The coefficient functions satisfy the identifiability constraint $\beta_1(u) > 0$ and $\sum_{k=1}^{p} \beta_k^2(u) = 1$ for $0 \leq u \leq 1$.

For each given $U = u_0$, with an unspecified link, model (1.1) is a single index model. This condition is the standard restriction (Ichimura, 1993; Lin & Kulasekera, 2007) . It ensures that our objective function (3) has a well separated minimum (van der Vaart, 1998) in the neighborhood of the true parameter.

A4: There exist a positive definite matrix $M_2(\theta_0)$ such that

$$E\left\{\frac{\partial^2 M_n(\theta)}{\partial\theta\partial\theta^T}\right\}_{\theta=\theta_0} \to M_2(\theta_0), \quad n \to \infty,$$

where $M_2(\theta_0)$ is analogous to the information matrix in classical linear models.

This condition is similar to condition M7 of Chiou & Müller (1998) and condition 3 of Lemma 5.4 in Ichimura (1993).

A5: As $n \to \infty$, $h_1 \sim n^{-\delta_1}$, $h_2 \sim n^{-\delta_2}$ with $0 < \delta_1 \leq 1/5 < \delta_2 < 1$.

A6: The response variable is continuous with $E\left(|Y|^m\right) < \infty$, for some $m > 1+(1+3\delta_1+2\delta_2)/(1-3\delta_1-\delta_2)$. If we set the smoothing parameter $h_1 \sim n^{-1/5}$, which is the optimal order for nonparametric regression function estimators, we then require $m > 6$. We further assume the covariate $X = (x_1, \ldots, x_p)^{\mathrm{T}}$ satisfies $\max_{1\le k\le p} |x_k| \le 1$.

A7: The kernel functions $K_1(\cdot)$, $K_2(\cdot)$ are symmetric densities that are supported on $[-1, 1]$ and are continuously differentiable.

A8: Let $A_n^{(k)}(X, \theta)$ be the kth partial derivative of $A_{n,u_0}(X, \theta)$ with respect to θ and let $A^{(k)}(X, \theta)$ be its probability limit. We assume $\sup_{(X,\theta)\in\mathcal{X}\times\Theta} A^{(k)}(X, \theta) < \infty$ for $k = 0, 1, 2$ and $\inf_{t,\theta} f_{\theta^{\mathrm{T}}X,U}(t, u_0) > 0$.

A9: Let

$$\begin{aligned} \psi(u) &= E\left\{g'(\theta_0^{\mathrm{T}}X)g_1(X, \theta_0) \otimes X^{\mathrm{T}} \mid U = u\right\}, && (8)\\ \rho(u, x) &= E\left[\{Y - g(\theta_0^{\mathrm{T}}X)\}^2 \mid U = u, X = x\right], && (9)\\ \Delta(u) &= E\left\{\rho(U, X)g_1(X, \theta_0)g_1(X, \theta_0)^{\mathrm{T}} \mid U = u\right\}, && (10) \end{aligned}$$

where $\otimes$ denotes the Kronecker product. Here $\hat{g}^{(k)}(X, \theta)$ is the kth partial derivative of $\hat{g}_{u_0}(\theta^{\mathrm{T}}X, \theta)$ with respect to θ and $g_k(X, \theta)$ is its probability limit. We assume $\psi(u)$, $\Delta(u)$, $E(|Y|^3 \mid U = u)$ and the marginal density $f_U(\cdot)$ of U are twice differentiable and $f_U(u_0) > 0$.

First we establish a uniform consistency result for our link function estimator that will be used in proving the asymptotic properties of the estimated varying coefficients.

Lemma 1. *Under assumptions A1-A8, for* $k = 0, 1, 2$

$$\sup_{(X,\theta)\in(\mathcal{X}\times\Theta)} \left|\hat{g}^{(k)}(X, \theta) - g_k(X, \theta)\right| = o_p(1).$$

Proof. First we will show $\sup_{t,\theta} \mid A_{n,u_0}(t,\theta) - A(t,\theta) \mid \xrightarrow{p} 0$ where $A(t,\theta)$ be the probability limit of of $A_{n,u_0}(t,\theta)$. Consider the following.

$$\begin{aligned} \sup_{t,\theta} \mid A_{n,u_0}(t,\theta) - A(t,\theta) \mid \;& \leq \sup_{t,\theta} \mid A_{n,u_0}(t,\theta) - E\{A_{n,u_0}(t,\theta)\} \mid + \\ & \quad \sup_{t,\theta} \mid E\{A_{n,u_0}(t,\theta)\} - A(t,\theta) \mid \\ &= I + II. \end{aligned} \tag{11}$$

For a suitable sequence $a_n \to \infty$ we can write

$$\begin{aligned} A_{n,u_0}(t,\theta) \;&= (nh_1h_2)^{-1} \textstyle\sum_{j=1}^n Y_j K_1 \left(\frac{\theta^T X_j - t}{h_1}\right) K_2 \left(\frac{U_j - u_0}{h_2}\right) \\ &= (nh_1h_2)^{-1} \textstyle\sum_{j=1}^n Y_j I_{[Y_j \notin (-a_n, a_n)]} K_1 \left(\frac{\theta^T X_j - t}{h_1}\right) K_2 \left(\frac{U_j - u_0}{h_2}\right) + \\ & \quad (nh_1h_2)^{-1} \textstyle\sum_{j=1}^n Y_j I_{[Y_j \in (-a_n, a_n)]} K_1 \left(\frac{\theta^T X_j - t}{h_1}\right) K_2 \left(\frac{U_j - u_0}{h_2}\right) \\ &= A_{n,1} + A_{n,2}. \end{aligned}$$

This implies

$$\begin{aligned} \sup_{t,\theta} \mid A_{n,u_0}(t,\theta) - E\{A_{n,u_0}(t,\theta)\} \mid \;& \leq \sup_{t,\theta} \mid A_{n,1}(t,\theta) - E\{A_{n,1}(t,\theta)\} \mid \\ & \quad + \sup_{t,\theta} \mid A_{n,2}(t,\theta) - E\{A_{n,2}(t,\theta)\} \mid \\ &= I_1 + I_2 \end{aligned}$$

Consider $P[I_1 > \epsilon]$. Let $A_{n,1,j}$ be the jth summand of $A_{n,1}$. Then

$$\begin{aligned}
P[I_1 > \epsilon] &= P[\sup_{t,\theta} \mid (nh_1h_2)^{-1} \textstyle\sum_{j=1}^{n}[A_{n,1,j}(t,\theta) - E\{A_{n,1,j}(t,\theta)\}] \mid > \epsilon] \\
&= P[\sup_{t,\theta} \mid \textstyle\sum_{j=1}^{n}[A_{n,1,j}(t,\theta) - E\{A_{n,1,j}(t,\theta)\}] \mid > \epsilon(nh_1h_2)] \\
&= P[\textstyle\sum_{j=1}^{n} \sup_{t,\theta} \mid \ [A_{n,1,j}(t,\theta) - E\{A_{n,1,j}(t,\theta)\}] \mid > \epsilon(nh_1h_2)] \\
&\leq \frac{E\sum_{j=1}^{n} \sup_{t,\theta} \mid \ A_{n,1,j}(t,\theta) - E\{A_{n,1,j}(t,\theta)\} \mid]}{\epsilon n h_1 h_2} \\
&\leq \frac{E[\sum_{j=1}^{n} \sup_{t,\theta} \mid \ A_{n,1,j}(t,\theta)| + \sup_{t,\theta} |E\{A_{n,1,j}(t,\theta)\} \mid]}{\epsilon n h_1 h_2} \\
&= \frac{nE\{\sup_{t,\theta} \mid \ A_{n,1,j}(t,\theta)|\}}{\epsilon n h_1 h_2} + \frac{nE\{\sup_{t,\theta} \mid \ EA_{n,1,j}(t,\theta)|\}}{\epsilon n h_1 h_2} \\
&\leq \frac{E\{\sup_{t,\theta} \mid \ A_{n,1,j}(t,\theta)|\}}{\epsilon h_1 h_2} + \frac{\sup_{t,\theta} E\{| \ A_{n,1,j}(t,\theta)|\}}{\epsilon h_1 h_2} \\
&\leq 2\frac{E\{\sup_{t,\theta} \mid \ A_{n,1,j}(t,\theta)|\}}{\epsilon h_1 h_2} \\
&= 2\frac{E\left\{ \sup_{t,\theta} \left| \ Y_j I_{[Y_j \notin (-a_n, a_n)]} K_1\left(\frac{\theta^T X_j - t}{h_1}\right) K_2\left(\frac{U_j - u_0}{h_2}\right) \right| \right\}}{\epsilon h_1 h_2} \ .
\end{aligned}$$

Since K_1 and K_2 has bounded supports, for suitable constants C_1 and C_2 we get

$$\begin{aligned}
P[I_1 > \epsilon] &\leq \tfrac{2C_1C_2}{\epsilon h_1 h_2} E\left[Y_j I_{\{Y_j \notin (-a_n, a_n)\}}\right] \\
&\leq \tfrac{2C_1C_2}{\epsilon h_1 h_2} \left(E|Y_j|^m\right)^{1/m} \left(\tfrac{E|Y_j|^m}{a_n^m}\right)^{1-1/m} \\
&= \tfrac{2C_1C_2}{\epsilon h_1 h_2} \tfrac{E|Y_j|^m}{a_n^{(m-1)}} \ .
\end{aligned}$$

Therefore $I_1 \xrightarrow{p} 0$ uniformly in (t,θ) if

$$\epsilon h_1 h_2 a_n^{(m-1)} \to \infty \ . \tag{12}$$

Consider $P[I_2 > \epsilon]$. Let $A_{n,2,j}$ be the jth summand of $A_{n,2}$. Then

$$P[I_2 > \epsilon] = P\left[\sup_{t,\theta} \left|(nh_1h_2)^{-1} \sum_{j=1}^{n} [A_{n,2,j}(t,\theta) - E\{A_{n,2,j}(t,\theta)\}] \right| \ > \epsilon\right]$$

Let $\mathcal{X}$ be the set of p-dimensional vectors and Θ be the p-dimensional parameter space. Without loss of generality assume $||x||_\infty \le 1 \ \forall x \in \mathcal{X}$. Following Ichimura (1993), partition Θ into N_1 cubes with length of a side $(h_1h_2)^\nu\delta$ and X into N_2 cubes with length of a side $(h_1h_2)^\nu$, where δ is a small positive number and ν is a large constant. Then, $N_1 = \delta^{-p}(h_1h_2)^{-p\nu}$ and $N_2 = (h_1h_2)^{-p\nu}$ where p is the dimension of the parameter space Θ. Therefore the space $\Theta \times \mathcal{X}$ is partitioned into N cubes with $N = N_1 \times N_2$. Each cube is $p \times p$ dimensional and becomes smaller as $n \to \infty$. Let $B_k^N, k = 1, \ldots, n$ denote all these $p \times p$ cubes. Pick a point (t_k^N, θ_k^N) from each B_k^N. Then

$$\begin{aligned} P[I_2 > \epsilon] &= P\left[\bigcup_{k=1}^{N} \sup_{(t,\theta)\in B_k^N} \left|(nh_1h_2)^{-1}\sum_{j=1}^{n}[A_{n,2,j}(t,\theta) - E\{A_{n,2,j}(t,\theta)\}]\right| > \epsilon\right] \\ &\le \sum_{k=1}^{N} P\left[\sup_{(t,\theta)\in B_k^N} \left|(h_1h_2)^{-1}\sum_{j=1}^{n}[A_{n,2,j}(t,\theta) - E\{A_{n,2,j}(t,\theta)\}]\right| > n\epsilon\right] \end{aligned} \quad (13)$$

Each summand of the outside sum of (44) can be decomposed into three parts by adding and subtracting $A_{n,2,j}(t_k^N, \theta_k^N)$ and $E\{A_{n,2,j}(t_k^N, \theta_k^N)\}$ within the inside sum. That is

$$P\left[\sup_{(t,\theta)\in B_k^N} \left|(h_1h_2)^{-1}\sum_{j=1}^{n}[A_{n,2,j}(t,\theta) - E\{A_{n,2,j}(t,\theta)\}]\right| > n\epsilon\right] \le P[I_{21k}] + P[I_{22k}] + P[I_{23k}]$$

where

$$\begin{aligned} P[I_{21k}] &= P\left[\left|(h_1h_2)^{-1}\textstyle\sum_{j=1}^{n}[A_{n,2,j}(t_k^N, \theta_k^N) - E\{A_{n,2,j}(t_k^N, \theta_k^N)\}]\right| > \tfrac{n\epsilon}{2}\right] \\ P[I_{22k}] &= P\left[\sup_{(t,\theta)\in B_k^N} \left|(h_1h_2)^{-1}\textstyle\sum_{j=1}^{n}[A_{n,2,j}(t,\theta) - A_{n,2,j}(t_k^N, \theta_k^N)]\right| > \tfrac{n\epsilon}{4}\right] \\ P[I_{23k}] &= P\left[\sup_{(t,\theta)\in B_k^N} \left|(h_1h_2)^{-1}\textstyle\sum_{j=1}^{n}[E\{A_{n,2,j}(t_k^N, \theta_k^N)\} - E\{A_{n,2,j}(t,\theta)\}]\right| > \tfrac{n\epsilon}{4}\right]. \end{aligned}$$

Then

$$\begin{aligned} P[I_2 > \epsilon] &\leq \textstyle\sum_{k=1}^{N} P[I_{21k}] + \sum_{k=1}^{N} P[I_{22k}] + \sum_{k=1}^{N} P[I_{23k}] \\ &= I_2^1 + I_2^2 + I_2^3 \end{aligned}$$

Now we will show that these 3 terms converge in probability to 0 as $n \to \infty$. Consider I_2^1.

$$P[I_{21k}] = P\left[|\sum_{j=1}^{n} W_{jn}| > \frac{n\epsilon h_1 h_2}{2} \right]$$

where $W_{jn} = [A_{n,2,j}(t_k^N, \theta_k^N) - EA_{n,2,j}(t_k^N, \theta_k^N)]$. Using the definition of $A_{n,2}$ and boundedness of the two kernels, we have $|W_{in}| \leq 2a_n C_1 C_2$ and

$$\begin{aligned} \text{var}[W_{jn}] &\leq E[A_{n,2,j}^2] \\ &\leq E\left[|Y|^2 K_1^2 \left(\tfrac{\theta^T X - t_k^N}{h_1} \right) K_2^2 \left(\tfrac{U-u_0}{h_2} \right) \right] \\ &= \int\int \psi(s,u) K_1^2 \left(\frac{s - t_k^N}{h_1} \right) K_2^2 \left(\frac{u-u_0}{h_2} \right) f_{\theta^T X, U}(s,u) ds du \end{aligned}$$

where $\psi(s,u) = E\left[|Y|^2 \mid \theta^T X = s, U = u\right]$. Assuming $\psi(\cdot,\cdot)$ and $f_{\theta^T X,U}(\cdot,\cdot)$ are differentiable with bounded derivatives, standard arguments yield $\text{var}[W_{jn}] = O(h_1 h_2)$. Applying Bernstein's inequality to $P[I_{21k}]$ we get $P[I_{21k}] \leq exp(-d_{n1}/2)$ where

$$d_{n1} = \frac{n\epsilon^2 h_1 h_2 C}{[1 + a_n \epsilon C]} \ .$$

Here and what follows we use C to denote a generic positive constant. Therefore

$$\begin{aligned} \textstyle\sum_{k=1}^{N} P[I_{21k}] &\leq N exp(-d_{n1}) \\ &= \left[\delta^{-p}(h_1 h_2)^{-2p\nu}\right]^2 exp(-d_{n1}) \\ &= \delta^{2p} exp\left[-d_{n1} \left\{ 1 + 4p\nu \tfrac{ln(h_1 h_2)}{d_{n1}} \right\} \right] \end{aligned}$$

Note that $-ln(h_1h_2)/d_{n1} \to 0 \Rightarrow d_{n1} \to \infty$ faster than $-ln(h_1h_2)$. Therefore if

$$-ln(h_1h_2)/d_{n1} \to 0 \tag{14}$$

then $\sum_{k=1}^{N} P[I_{21k}] \to 0$ as $n \to \infty$ which implies that $I_2^1 = O(c_{1n})$ where

$$c_{1n} = exp\left[-d_{n1}\left\{1 + C\frac{ln(h_1h_2)}{d_{n1}}\right\}\right] .$$

Now consider I_2^2. Note that

$$P[I_{22k}] \le P\left[(h_1h_2)^{-1}\sum_{j=1}^{n}\sup_{(t,\theta)\in B_k^N}\left|A_{n,2,j}(t,\theta) - A_{n,2,j}(t_k^N,\theta_k^N)\right| > \frac{n\epsilon}{4}\right] .$$

Denoting $T_{nk} = \sum_{j=1}^{n}\sup_{(t,\theta)\in B_k^N}\left|A_{n,2,j}(t,\theta) - A_{n,2,j}(t_k^N,\theta_k^N)\right|$ we get

$$P[I_{22k}] \le \left[|T_{nk} - E(T_{nk}) + E(T_{nk})| > \frac{n\epsilon(h_1h_2)}{4}\right] = T_{nk1} + T_{nk2}$$

where

$$T_{nk1} = P\left[|T_{nk} - E(T_{nk})| > \frac{n\epsilon(h_1h_2)}{8}\right]; \quad T_{nk2} = P\left[E(T_{nk}) > \frac{n\epsilon(h_1h_2)}{8}\right].$$

First note that

$$\begin{aligned}
&\sup_{(t,\theta)\in B_k^N}\left|A_{n,2,j}(t,\theta) - A_{n,2,j}(t_k^N,\theta_k^N)\right| \\
&= \sup_{(t,\theta)\in B_k^N}\Bigg|Y_jI_{[Y_j\in(-a_n,a_n)]}K_2\left(\frac{U_j-u_0}{h_2}\right)\left\{K_1\left(\frac{\theta^TX_j-t}{h_1}\right)\right. \\
&\left.-K_1\left(\tfrac{(\theta_k^N)^TX_j-t_k^N}{h_1}\right)\right\}\Bigg|
\end{aligned}$$

Assuming $K_1(\cdot)$ is differentiable, for a suitable mean value $\bar{c}$, we get

$$\begin{aligned}
&\sup_{(t,\theta)\in B_k^N}\left|K_1\left(\frac{\theta^T X_j - t}{h_1}\right) - K_1\left(\frac{(\theta_k^N)^T X_j - t_k^N}{h_1}\right)\right| \\
&= \sup_{(t,\theta)\in B_k^N}\left|\frac{(\theta^T X_j - t) - \left\{(\theta_k^N)^T X_j - t_k^N\right\}}{h_1} K_1'(\bar{c})\right| \\
&\leq \frac{C}{h_1}\sup_{(t,\theta)\in B_k^N}\left|(\theta^T X_j - t) - \left\{(\theta_k^N)^T X_j - t_k^N\right\}\right| .
\end{aligned}$$

Note that t_k^N and t are points from B_k^N and therefore, $(t_k^N - t)$ can be written as $(\theta_k^N)^T x^*$ for some suitable $x^* \in \mathcal{X}$. Also note that $||\theta - \theta_k^N||_2 \leq \sqrt{p(h_1h_2)^{2\nu}\delta^2} = C(h_1h_2)^\nu$. Then from the boundedness of the X vectors and the compactness of the parameter space Θ, we get

$$\sup_{(t,\theta)\in B_k^N}\left|A_{n,2,j}(t,\theta) - A_{n,2,j}(t_k^N,\theta_k^N)\right| \leq \frac{Ca_n(h_1h_2)^\nu}{h_1} . \tag{15}$$

Now consider the expectation of the left hand side of (15).

$$\begin{aligned}
&E\left[\sup_{(t,\theta)\in B_k^N}\left|A_{n,2,j}(t,\theta) - A_{n,2,j}(t_k^N,\theta_k^N)\right|\right] \\
&\leq \tfrac{C(h_1h_2)^\nu}{h_1} E\left[\sup_{(t,\theta)\in B_k^N}\left|Y_j I_{[Y_j\in(-a_n,a_n)]} K_2\left(\frac{U_j - u_0}{h_2}\right)\right|\right] \\
&= \frac{(h_1h_2)^\nu}{h_1}\int \phi_1(u) K_2\left(\frac{u-u_0}{h_2}\right) f_U(u)du
\end{aligned}$$

where $\phi_1(u) = E[|Y| \mid U = u]$ and $f_U(u)$ is the density of U. Assuming ϕ and f_U are differentiable with bounded derivatives, standard arguments yield

$$E\left[\sup_{(t,\theta)\in B_k^N}\left|A_{n,2,j}(t,\theta) - A_{n,2,j}(t_k^N,\theta_k^N)\right|\right] = O\left(h_1^{\nu-1}h_2^{\nu+1}\right) . \tag{16}$$

Now consider T_{nk1}. To apply Bernstein's inequality, let $W_{jn} = T_{nk1} - E(T_{nk1})$. Using (15) and (16) we get $|W_{jn}| \leq |T_{nk1}| + |E(T_{nk1})| \leq C a_n h_1^{\nu-2} h_2^{\nu-1}$. Also

$$\begin{aligned}
\text{var}[W_{jn}] &= \text{var}[T_{nk1}] \\
&\leq E[T_{nk1}^2] \\
&\leq \frac{C(h_1 h_2)^{2\nu}}{h_1^2} E\left[Y^2 K_2^2\left(\frac{U-u_0}{h_2}\right)\right] \\
&= \frac{C(h_1 h_2)^{2\nu}}{h_1^2} \int \phi_2(u) K_2^2\left(\frac{u-u_0}{h_2}\right) f_U(u) du
\end{aligned}$$

where $\phi_2(u) = E[Y^2 \mid U = u]$. As before, assuming ϕ_2 and f_U is differentiable with bounded derivatives, we get $\text{var}[W_{jn}] = O(h_1^{2\nu-4} h_2^{2\nu-1})$. Now applying Bernstein's inequality to T_{n1} we get $T_{nk1} \leq exp(-d_{n2}/2)$ where

$$d_{n2} = \frac{n\epsilon^2 h_1 h_2 C}{h_1^{2\nu-5} h_2^{2\nu-4} + \epsilon a_n h_1^{\nu-2} h_2^{\nu-1} C} .$$

If $\nu \geq 3$ then $d_{n2} > d_{n1}$ for large enough n. If (14) is satisfied then $d_{n1} \to \infty \Rightarrow d_{n2} \to \infty$ as $n \to \infty$ which implies $\sum_{k=1}^N T_{nk1} \leq N exp(-d_{n2})$. Therefore if (14) is satisfied then $\sum_{k=1}^N T_{nk1} \to 0$ as $n \to \infty$.

Now consider T_{nk2}.

$$\begin{aligned}
T_{nk2} &= P\left[E(T_{nk}) > \tfrac{n\epsilon(h_1 h_2)}{8}\right] \\
&= P\left[nE\big\{\sup_{(t,\theta)\in B_k^N} \left|A_{n,2,j}(t,\theta) - A_{n,2,j}(t_k^N, \theta_k^N)\right|\big\} > \tfrac{n\epsilon(h_1 h_2)}{8}\right] \\
&\leq P\left[O\left(h_1^{\nu-1} h_2^{\nu+1}\right) > \tfrac{\epsilon(h_1 h_2)}{8}\right]
\end{aligned}$$

where the last inequality follows from (16). For $\nu \geq 2$ we get $T_{nk2} \to 0$ as $n \to \infty$ which implies $\sum_{k=1}^N T_{nk2} \to 0$. Therefore we have $\sum_{k=1}^N P[I_{22k}] \to 0$ which yields $I_2^2 \to 0$ as $n \to \infty$. Moreover we can choose $\nu \geq 3$ so that the rate is faster than c_{1n}

because for $\nu \geq 3$, $d_{n2} > d_{n1}$. Here N is the number of cubes and is fixed for all n.

Now consider I_2^3.

$$\begin{aligned} P[I_{23k}] &\leq P\left[E\Big\{\sup_{(t,\theta)\in B_k^N}\left|(h_1h_2)^{-1}\textstyle\sum_{j=1}^n A_{n,2,j}(t_k^N,\theta_k^N) - A_{n,2,j}(t,\theta)\right|\Big\} > \tfrac{n\epsilon}{4}\right] \\ &\leq P\left[E\Big\{\textstyle\sum_{j=1}^n \sup_{(t,\theta)\in B_k^N}\left|A_{n,2,j}(t_k^N,\theta_k^N) - A_{n,2,j}(t,\theta)\right|\Big\} > \tfrac{n\epsilon(h_1h_2)}{4}\right] \\ &= P\left[nE\Big\{\sup_{(t,\theta)\in B_k^N}\left|A_{n,2,1}(t_k^N,\theta_k^N) - A_{n,2,1}(t,\theta)\right|\Big\} > \tfrac{n\epsilon(h_1h_2)}{4}\right] \\ &\leq P\left\{O\left(h_1^{\nu-1}h_2^{\nu+1}\right) > \tfrac{\epsilon(h_1h_2)}{4}\right\} \end{aligned}$$

where the last inequality follows from (16). As before, for $\nu \geq 2$ we have $\sum_{k=1}^N P[I_{23k}] \to 0$ which yields $I_2^3 \to 0$ as $n \to \infty$ as desired. Therefore if (14) is satisfied we have $I \xrightarrow{p} 0$ as $n \to \infty$.

Now consider II of (11).

$$\begin{aligned} E\left\{A_n(t,\theta)\right\} &= E\left\{\frac{1}{nh_1h_2}\sum_{j=1}^n Y_jK_1\left(\frac{\theta^TX_j - t}{h_1}\right)K_2\left(\frac{U_j-u_0}{h_2}\right)\right\} \\ &= \frac{1}{h_1h_2}\int\int \psi(s,u)K_1\left(\frac{s-t}{h_1}\right)K_2\left(\frac{u-u_0}{h_2}\right)f_{\theta^TX,U}(s,u)dsdu \end{aligned}$$

where $\psi(s,u) = E(Y|\,\theta^TX = s, U = u)$ and $f_{\theta^TX,U}(s,u)$ is the joint density of θ^TX, U. Let $\psi^* = \psi f$ and using a change of variables we can rewrite the above expression as

$$\frac{1}{h_1h_2}\int\int \psi^*(t+h_1v, u_0+h_2w)K_1(v)\,K_2(w)\,h_1h_2dvdw\ .$$

Assuming ψ^* is differentiable and uniformly bounded in (t,θ), a Taylor series expansion yields

$$E\left[A_n(t,\theta)\right] = \psi^*(t,u_0) + O(h_1^2) + O(h_2^2) + O(h_1h_2)\ .$$

Denoting $\psi^*(t, u_0) = A(t, \theta)$ we get

$$E\left[A_n(t, \theta)\right] - A(t, \theta) = O(h_1^2) + O(h_2^2) + O(h_1 h_2) \ .$$

Therefore term II of (11) converge to zero in probability as $n \to \infty$.

We have now shown that $\sup_{t,\theta} \mid A_{n,u_0}(t, \theta) - A(t, \theta) \mid \xrightarrow{p} 0$ as $n \to \infty$. Let the rate of convergence be a_{n0}. Similarly, we can show that

$$\sup_{t,\theta} \mid B_{n,u_0}(t, \theta) - B(t, \theta) \mid \xrightarrow{p} 0$$

as $n \to \infty$ with convergence rate b_{n0}. To avoid $\inf_{t,\theta} |B_{n,u_0}(t, \theta)| = 0$ in technical arguments we will add a $c_n > 0 \to 0$ to B_{n,u_0} and pick c_n appropriately so that the estimator is uniformly convergent. Note that $A(t, \theta) = E[Y|\theta^T X = t, U = u_0] f_{\theta^T X, U}(t, u_0)$ and $B(t, \theta) = f_{\theta^T X, U}(t, u_0)$. We will assume $\inf_{t,\theta} |B(t, \theta)| > 0$. We now show that

$$\sup_{t,\theta} \left| \frac{A_n(t, \theta)}{B_n(t, \theta)} - \frac{A(t, \theta)}{B(t, \theta)} \right| \xrightarrow{p} 0$$

as $n \to \infty$. Suppressing the argument (t, θ), we can write

$$\begin{aligned} \left| \frac{A_{n,u_0}}{B_{n,u_0}} - \frac{A}{B} \right| &= \left| \frac{A_{n,u_0}}{B_{n,u_0}} - \frac{A}{B_{n,u_0}} + \frac{A}{B_{n,u_0}} - \frac{A}{B} \right| \\ &\leq \frac{|A_{n,u_0} - A|}{|B_{n,u_0}|} + \frac{|A||B - B_{n,u_0}|}{|B_{n,u_0} B|} \\ &\leq \frac{\sup_{t,\theta} |A_{n,u_0}(t, \theta) - A(t, \theta)|}{\inf_{t,\theta} |B_{n,u_0}(t, \theta)|} + \frac{\sup_{t,\theta} |A(t, \theta)||B(t, \theta) - B_{n,u_0}(t, \theta)|}{\inf_{t,\theta} |B_{n,u_0}(t, \theta) B(t, \theta)|} \ . \end{aligned}$$

Note that $\inf_{t,\theta} |B_{n,u_0}(t, \theta)| = c_n$ and using condition A8 we have

$$\sup_{t,\theta} \left| \frac{A_{n,u_0}(t, \theta)}{B_{n,u_0}(t, \theta)} - \frac{A(t, \theta)}{B(t, \theta)} \right| = O\left(a_{n0}/c_n\right) + O\left(b_{n0}/c_n\right) \ .$$

This implies $\sup_{t,\theta} |\hat{g}_{u_0}(t,\theta) - g_0(t,\theta)| = o_p(1)$ where, $g_0(t,\theta) = A(t,\theta)/B(t,\theta)$. □

THEOREM 6. *Under assumptions A1-A8 in appendix A, the minimizer $\hat{\theta}$ of (3) is a consistent estimator of θ_0.*

Proof. By definition of $\hat{\theta}$, $P\left[M_n^{1/2}(\hat{\theta}) \leq M_n^{1/2}(\theta_0)\right] = 1$. Let $B_r(\theta_0)$ denote the open ball of radius $r > 0$ centered at θ_0. Let A be the event $M_n^{1/2}(\hat{\theta}) \leq M_n^{1/2}(\theta_0)$. Then

$$
\begin{aligned}
P\left[M_n^{1/2}(\hat{\theta}) \leq M_n^{1/2}(\theta_0)\right] &= P\left[A\ ,\ \hat{\theta} \in B_r(\theta_0)\ \bigcup\ A\ ,\ \hat{\theta} \notin B_r(\theta_0)\right] \\
&\leq P\left[A\ ,\ \hat{\theta} \in B_r(\theta_0)\right] + P\left[A\ ,\ \hat{\theta} \notin B_r(\theta_0)\right] \\
&\leq P\left[\hat{\theta} \in B_r(\theta_0)\right] + P\left[A\ ,\ \hat{\theta} \in \Theta \setminus B_r(\theta_0)\right] \\
&\leq P\left[\hat{\theta} \in B_r(\theta_0)\right] + P\left[\inf_{\theta \in \Theta \setminus B_r(\theta_0)} M_n^{1/2}(\theta) \leq M_n^{1/2}(\theta_0)\right] \\
&= I + II\ .
\end{aligned}
$$

If $II \to 0$, then we have for any $r > 0$ with probability tending to one, $\hat{\theta} \in B_r(\theta_0)$ as $n \to \infty$ which completes the proof. Therefore, we will show that $II \to 0$ as $n \to \infty$.

Note that $M(\theta) = E\Big[\{Y_i - g_0(X_i,\theta)\}^2 | U = u_0\Big]$ is the probability limit of (3). By adding and subtracting $M(\theta)$ and $M(\theta_0)$ into II we get

$$
II \leq P\Big[A1 + A2 \geq \inf_{\theta \in \Theta \setminus B_r(\theta_0)} M^{1/2}(\theta) - M^{1/2}(\theta_0)\Big]
$$

where

$$
A1 = \sup_{\theta \in \Theta} |M_n^{1/2}(\theta) - M^{1/2}(\theta)|; \quad A2 = |M_n^{1/2}(\theta_0) - M^{1/2}(\theta_0)|.
$$

Similar to (Ichimura (1993),pg89) by condition A3 we have

$$
\inf_{\theta \in \Theta \setminus B_r(\theta_0)} M^{1/2}(\theta) - M^{1/2}(\theta_0) > \epsilon
$$

Therefore the proof reduces to showing that $P[A1 > \epsilon/2]$ and $P[A2 > \epsilon/2]$ converging to zero as $n \to \infty$. Note that the first convergence will imply the second. Consider $A1$. Let

$$\tilde{M}_n(\theta) = (nh_2)^{-1} \sum_{i=1}^{n} \{Y_i - g_0(X_i, \theta)\}^2 K_2\left(\frac{U_i - u_0}{h_2}\right)$$

and $M^*(\theta) = E\left[\tilde{M}_n(\theta)\right]$. Then we have

$$\begin{aligned} A1 &\leq \sup_{\theta \in \Theta} |M_n^{1/2}(\theta) - \tilde{M}_n^{1/2}(\theta)| + \sup_{\theta \in \Theta} |\tilde{M}_n^{1/2}(\theta) - M^{*1/2}(\theta)| + \sup_{\theta \in \Theta} |M^{*1/2}(\theta) - M^{1/2}(\theta)| \\ &= A1_1 + A1_2 + A1_3. \end{aligned}$$

Consider $A1_1$. Using the general result

$$\left| \left(\sum_{i=1}^{n} w_i a_i^2\right)^{1/2} - \left(\sum_{i=1}^{n} w_i b_i^2\right)^{1/2} \right|^2 \leq \sum_{i=1}^{n} w_i (a_i - b_i)^2$$

and letting

$$w_i = \frac{1}{nh_2} K_2\left(\frac{U_i - u_0}{h_2}\right), \quad a_i = Y_i - \hat{g}_{u_0}(\theta^T X_i, \theta), \quad b_i = Y_i - g_0(X_i, \theta)$$

we get

$$\begin{aligned} \left| M_n^{1/2}(\theta) - \tilde{M}_n^{1/2}(\theta) \right|^2 &\leq \sum_{i=1}^{n} \frac{1}{nh_2} K_2\left(\frac{U_i - u_0}{h_2}\right) \left[g_0(X_i, \theta) - \hat{g}_{u_0}(\theta^T X_i, \theta)\right]^2 \\ &\leq \sup_{t,\theta} \left| g_0(X_i, \theta) - \hat{g}_{u_0}(\theta^T X_i, \theta) \right|^2 \frac{1}{nh_2} K_2\left(\frac{U_i - u_0}{h_2}\right). \end{aligned}$$

If $nh_2 \to \infty$ as $n \to \infty$ then by the consistency result of $\hat{g}_{u_0}(t, \theta)$ in Lemma 1 gives us

$$\sup_{t,\theta} \left| M_n^{1/2}(\theta) - \tilde{M}_n^{1/2}(\theta) \right| = o_p(1).$$

Now consider $A1_2$. Note that

$$|\tilde{M}_n^{1/2}(\theta) - M^{*1/2}(\theta)|^2 \le \left| \tilde{M}_n(\theta) - M^*(\theta) \right|.$$

Let

$$W_i(\theta) = \{Y_i - g_0(X_i, \theta)\}^2 K_2\left(\frac{U_i - u_0}{h_2}\right).$$

Now we need to show that

$$\sup_{\theta \in \Theta} \left| \frac{1}{nh_2} \sum_{i=1}^n W_i(\theta) - E\{W_i(\theta)\} \right| \xrightarrow{p} 0.$$

This type of convergence results are established in Andrews (1987). By verifying assumption A1,B1,B2 and A4 of Andrews (1987) we see that $A1_2 \to 0$ in probability.

Finally, consider $A1_3$. A simple calculation yields $M^*(\theta) = M(\theta) + O(h^2)$. By our assumptions in A9 we see that the order term is uniformly bounded in θ which shows that $A1_3 \to 0$ in probability. □

THEOREM 7. *Under assumptions A1-A9 in appendix A, the minimizer $\hat{\theta}$ of (1.3) converges in distribution to a normal random variable with mean vector θ_0 and co-variance matrix Σ_{u_0} where $\Sigma_{u_0} = \{M_2(\theta_0)\}^{-1} f_U(u_0)\nu_0\Delta(u_0)$ and $\nu_j = \int s^j K_2^2(s)ds$.*

Proof. A Taylor series expansion of (1.3) yields

$$0 = M_n^{(1)}(\theta_0) + M_n^{(2)}(\overline{\theta})(\hat{\theta} - \theta_0), \tag{17}$$

where $\overline{\theta}$ between $\hat{\theta}$ and θ_0 and $M_n^{(k)}(\theta^*)$ is the kth partial derivative of M_n with respect to θ evaluated at $\theta = \theta^*$. Using $M_2(\theta_0)$ defined in A4 and for a normalizing sequence

$(nh_2)^{1/2}$, we can write (17) as

$$\begin{aligned} M_2(\theta_0)(nh_2)^{1/2}(\hat{\theta}-\theta_0) &= -(nh_2)^{1/2}M_n^{(1)}(\theta_0) \\ &\quad + \left\{M_2(\theta_0) - M_n^{(2)}(\overline{\theta})\right\}(nh_2)^{1/2}(\hat{\theta}-\theta_0) \\ &= \textstyle\sum_{k=1}^4 T_{1,k} + T_2 \end{aligned} \tag{18}$$

where

$$\begin{aligned} T_{1,1} &= d_n \textstyle\sum_{i=1}^n \{Y_i - g_0(X_i,\theta_0)\} g_1(X_i,\theta_0) K_2\left(\frac{U_i-u_0}{h_2}\right), \\ T_{1,2} &= d_n \textstyle\sum_{i=1}^n \{Y_i - g_0(X_i,\theta_0)\} \left\{\hat{g}^{(1)}(X_i,\theta_0) - g_1(X_i,\theta_0)\right\} K_2\left(\frac{U_i-u_0}{h_2}\right), \\ T_{1,3} &= d_n \textstyle\sum_{i=1}^n \{g_0(X_i,\theta_0) - \hat{g}_{u_0}(\theta_0^{\mathrm{T}}X_i,\theta_0)\} g_1(X_i,\theta_0) K_2\left(\frac{U_i-u_0}{h_2}\right), \\ T_{1,4} &= d_n \textstyle\sum_{i=1}^n \{g_0(X_i,\theta_0) - \hat{g}_{u_0}(\theta_0^{\mathrm{T}}X_i,\theta_0)\} \left\{\hat{g}^{(1)}(X_i,\theta_0) - g_1(X_i,\theta_0)\right\} K_2\left(\frac{U_i-u_0}{h_2}\right), \end{aligned}$$

with $d_n = 2(nh_2)^{-1/2}$. Using Lemma 1, for suitably chosen bandwidth sequences h_1 and h_2 that satisfy condition A5 and $nh_2 \to \infty$, we can easily show that $T_{1,k}$ ($k = 2,3,4$) converges in probability to zero as $n \to \infty$. It remains to show $T_{1,1}$ is asymptotically normal. Since $T_{1,1}$ is a sum of independent random vectors, asymptotic normality of $T_{1,1}$ follows from the Cramer–Wold device if we show that for any unit vector a, $a^{\mathrm{T}}T_{1,1}$ converges to a univariate normal random variable. Hence we will find the mean and the covariance matrix of $T_{1,1}$, and verify Lyapounov's condition for the sequence $a^{\mathrm{T}}T_{1,1}$.

Let $\mu_j = \int s^j K_2(s)ds$, and

$$\beta^{(1)}(u_0) = \left\{\beta_1'(u_0),\ldots,\beta_p'(u_0)\right\}^{\mathrm{T}}, \beta^{(2)}(u_0) = \left\{\beta_1''(u_0),\ldots,beta_p''(u_0)\right\}^{\mathrm{T}}.$$

Also let $\psi^{(1)}(u_0)$ be the $p \times p$ matrix of first derivatives of (8) with respect to u

evaluated at u_0. Set

$$b(u_0) = f_U(u_0)\left\{\psi^{(1)}(u_0)\beta^{(1)}(u_0) + \frac{1}{2}\psi(u_0)\beta^{(2)}(u_0)\right\} + f'_U(u_0)\left\{\psi(u_0)\beta^{(1)}(u_0)\right\} \quad (19)$$

Using (8), (10) and (19) together with condition A8 and A9, standard calculations show that

$$E\left(T_{1,1}\right) = (nh_2)^{1/2}\left\{2\mu_2 h_2^2 b(u_0) + O(h_2^3)\right\} \quad (20)$$

and

$$\text{var}\left(T_{1,1}\right) = 4f_U(u_0)\nu_0\Delta(u_0) + o(1). \quad (21)$$

Now for any unit vector a, we will verify Lyapounov's condition for the sequence $a^{\mathrm{T}}T_{1,1}$. To this end, write $a^{\mathrm{T}}T_{1,1} = \sum_{i=1}^{n} d_n W_i^{(n)} a^{\mathrm{T}} g_1(X_i, \theta_0)$ where

$$W_i^{(n)} = \left\{Y_i - g_0(X_i, \theta_0)\right\} K_2\left(\frac{U_i - u_0}{h_2}\right).$$

Let $s_n^2 = \text{var}(a^{\mathrm{T}}T_{1,1})$. Since $g_0(\cdot, \theta_0) = g(\cdot)$, from (21) we have

$$s_n^2 = 4f_U(u_0)\nu_0 a^{\mathrm{T}}\Delta(u_0)a + o(1)$$

and hence Lyapounov's condition holds for $a^{\mathrm{T}}T_{1,1}$, if

$$\sum_{i=1}^{n} E\left|d_n W_i^{(n)} a^{\mathrm{T}} g_1(X_i, \theta_0)\right|^{2+\delta} \to 0$$

for some $\delta > 0$. From A8 and A9, we have $E\left(|W^{(n)}a^{\mathrm{T}}g_1(X, \theta_0)|^3\right) = O(h_2)$ which leads to

$$d_n^3 \sum_{i=1}^{n} E\left\{\left|W_i^{(n)} a^{\mathrm{T}} g_1(X_i, \theta_0)\right|^3\right\} = O\left(n^{-1/2}h_2^{-1/2}\right).$$

Therefore, if $nh_2 \to \infty$ with $nh_2^5 \to 0$, we have $T_{1,1}$ converges in distribution to a multivariate normal random variable by the Cramer–Wold device.

It remains to show that T_2 converges in probability to zero. From Lemma 1, we can show that

$$\left\{M_n^{(2)}(\overline{\theta}) - M_2(\theta_0)\right\} \to 0 \tag{22}$$

in probability for suitably chosen bandwidth sequences satisfying condition A5. Now we will show $(nh_2)^{1/2}(\hat{\theta} - \theta_0) = O_p(1)$. For a p-dimensional vector x, let $\|x\|_\infty = max_{1\le k\le p}|x_k|$. Since $M_n(\hat{\theta}) \le M_n(\theta_0)$, a Taylor series expansion, Theorem 1 and (22) gives

$$(\hat{\theta} - \theta_0)^{\mathrm{T}} M_n^{(1)}(\theta_0) + 1/2(\hat{\theta} - \theta_0)^{\mathrm{T}} M_2(\theta_0)(\hat{\theta} - \theta_0) + o_p(||\hat{\theta} - \theta_0||_\infty^2) \le 0. \tag{23}$$

Following Ichimura (1993), multiplying both sides of (23) by

$$nh_2 \left\{1 + (nh_2)^{1/2}||\hat{\theta} - \theta_0||_\infty\right\}^{-2}$$

and letting $c_n(\theta) = (nh_2)^{1/2}(\theta - \theta_0)\left\{1 + (nh_2)^{1/2}||\theta - \theta_0||_\infty\right\}^{-1}$ yields

$$\frac{c_n(\hat{\theta})^{\mathrm{T}}(nh_2)^{1/2} M_n^{(1)}(\theta_0)}{\left\{1 + (nh_2)^{1/2}||\hat{\theta} - \theta_0||_\infty\right\}} + \frac{1}{2} c_n(\hat{\theta})^{\mathrm{T}} M_2(\theta_0) c_n(\hat{\theta}) + \frac{o_p(1) nh_2 ||\hat{\theta} - \theta_0||_\infty^2}{\left\{1 + (nh_2)^{1/2}||\hat{\theta} - \theta_0||_\infty\right\}^2} \le 0\ . \tag{24}$$

If $(nh_2)^{1/2}||(\hat{\theta} - \theta_0)||_\infty \to \infty$ in probability then we have $c_n(\hat{\theta})$ converging in probability to a finite vector with at least one of the entries being equal to one. We have already shown that $(nh_2)^{1/2} M_n^{(1)}(\theta_0)$ converges in distribution to a finite random variable. Therefore as $n \to \infty$ from (24) we get

$$(1/2) c_n(\hat{\theta})^{\mathrm{T}} M_2(\theta_0) c_n(\hat{\theta}) \le o_p(1)\ . \tag{25}$$

Since $M_2(\theta_0)$ is positive definite (25) leads to $c_n(\hat{\theta})$ converging to the zero vector in probability which is a contradiction. Hence we must have $(nh_2)^{1/2}||\hat{\theta}-\theta_0||_\infty = O_p(1)$ which implies that $T_2 \to 0$ in probability as $n \to \infty$. Therefore as $n \to \infty$ from (18) we have $M_2(\theta_0)(nh_2)^{1/2}(\hat{\theta}-\theta_0)$ converging in distribution to a multivariate normal random variable. □

A.1 Additional details of the proof of Theorem 2

Lemma 2. *Under conditions A1-A9,*

$$\sup_{(X,\theta)\in(\mathcal{X}\times\Theta)} \left| \frac{\partial A_{n,u_0}(\theta^T X, \theta)}{\partial \theta} - A^{(1)}(X,\theta) \right| = O_p(a_{n1})$$

Lemma 3. *Under conditions A1-A9,*

$$\sup_{(X,\theta)\in(\mathcal{X}\times\Theta)} \left| \frac{\partial^2 A_{n,u_0}(\theta^T X, \theta)}{\partial \theta \partial \theta^T} - A^{(2)}(X,\theta) \right| = O_p(a_{n2})$$

The proofs of these two lemmas follow the same lines of arguments and hence only the proof of Lemma 3 will be given. For convenience we will supress the subscript u_0.

Proof. As in Lemma 1, we can decompose

$$\left| \frac{\partial^2 A_n(\theta^T X, \theta)}{\partial \theta \partial \theta^T} - A^{(2)}(X,\theta) \right|$$

into two terms as

$$\begin{aligned}\left|\tfrac{\partial^2 A_n(\theta^T X,\theta)}{\partial\theta\partial\theta^T} - A^{(2)}(X,\theta)\right| &\leq \left|\tfrac{\partial^2 A_n(\theta^T X,\theta)}{\partial\theta\partial\theta^T} - E\tfrac{\partial^2 A_n(\theta^T X,\theta)}{\partial\theta\partial\theta^T}\right| + \left|E\tfrac{\partial^2 A_n(\theta^T X,\theta)}{\partial\theta\partial\theta^T} - A^{(2)}(X,\theta)\right| \\ &= I + II\ .\end{aligned}$$

Term I can be analyzed using the techniques given in Lemma 1 term I with an additional h_2^2 in the denominators of the order terms. Therefore we get $I = a_{n2,1}$ where

$$a_{n2,1} = O\left\{\frac{1}{h_1^3 h_2 a_n^{(m-1)}}\right\} + O\left[exp\left\{-d''_{n1}\left(1+\frac{lnh_1h_2}{d''_{n1}}\right)\right\}\right]$$

where

$$d''_{n1} = \frac{Cnh_1^5 h_2}{1+Ca_n h_1^2}$$

and a_n is a sequence that satisfy $a_n \to \infty$ as $n \to \infty$ whose rate will be determined later.

Now consider the term II.

$$\frac{\partial^2 A_n(\theta^T X,\theta)}{\partial\theta\partial\theta^T} = \frac{1}{nh_1^3 h_2}\sum_{j=1}^{n} Y_j K_1''\left(\frac{\theta^T X_j - \theta^T X}{h_1}\right) K_2\left(\frac{U_j - u_0}{h_2}\right)(X_j - X)(X_j - X)^T\ .$$

Note that this is a $p \times p$ matrix and we will analyze only the (1,1) element. First we need to compute the expectation of this element.

$$\begin{aligned}E\left[\left\{\tfrac{\partial^2 A_n(\theta^T X,\theta)}{\partial\theta\partial\theta^T}\right\}_{(1,1)}\right] &= E\left[E\left(\cdot \mid\ \theta^T X\right)\right] \\ &= \int \phi_2^*(s) f_{\theta^T X}(s) ds\end{aligned}$$

where

$$
\begin{aligned}
\phi_2^*(t) &= E\left[\tfrac{1}{h_1^3 h_2} Y_j K_1''\left(\tfrac{\theta^T X_j - \theta^T X}{h_1}\right) K_2\left(\tfrac{U_j - u_0}{h_2}\right) [(X_j - X_1)(X_j - X)^T]_{1,1} \mid \ \theta^T X = t\right] \\
&= E[E(\ \cdot \mid \theta^T X_j = s, U = u, \theta^T X = t) \mid \ \theta^T X = t] \\
&= \frac{1}{h_1^3 h_2} \int\int \psi(s,u,t) K_1\left(\frac{s-t}{h_1}\right) K_2\left(\frac{u-u_0}{h_2}\right) f_{\theta^T X_1, U|\theta^T X = t}(s,u|t) ds du \\
&= \frac{1}{h_1^3 h_2} \int\int \psi^*(t + h_1 v, u_0 + h_2 w, t) K_1''(v) K_2(w) h_1 dv h_2 dw \\
&= T_1 + \ldots + T_3 + \tfrac{1}{2}(T_4 + \ldots + T_6) + \tfrac{1}{6}\textstyle\sum_{k=1}^{9} T_{7,k}
\end{aligned}
$$

where $\psi^* = \psi f_{\theta^T X_1, U|\theta^T X = t}$ and

$$
\psi(s,u,t) = E\left[Y \left\{(X_j - X)_{1,1}\right\}^2 |\theta^T X_j = s, U = u, \theta^T X = t\right].
$$

We will analyze these terms next. We need further notation and conditions to analyze these terms. Let $\phi^*_{x^j y^k}(\cdot,\cdot,\cdot) = \frac{\partial^{j+k}\phi^*(x,y,t)}{\partial x^j \partial y^k}$. Using our assumptions on the two kernels in condition A7 , we see that all odd moments of K_1 and K_2 are zero and all even moments are non zero. It is also easy to see that all odd moments of K_1'' and K_2'' are zero and all even moments are non zero. In addition to these we require $\int K_1''(s)ds = 0$. These conditions are satisfied by most kernel that are used in practice. For example $K(x) = C(1-x^2)^2 I_{[-1,1]}(x)$ is one such kernel. Under these conditions we have

$$
\begin{aligned}
T_1 &= \frac{1}{h_1^2}\phi^*(t,u_0,t) \int K_1''(v)dv \int K_2(w)dw = 0 \\
T_2 &= \frac{1}{h_1^2}\phi^*_x(t,u_0,t) \int\int h_1 v K_1''(v) K_2(w) dv dw = 0 \\
T_3 &= \frac{1}{h_1^2}\phi^*_y(t,u_0,t) \int\int K_1''(v) h_2 w K_2(w) dv dw = 0 \\
T_4 &= \frac{1}{h_1^2}\phi^*_{x^2}(t,u_0,t) \int\int h_1^2 v^2 K_1''(v) K_2(w) dv dw = \phi^*_{xx}(t,u_0,t) \\
T_5 &= \frac{1}{h_1^2}\phi^*_{y^2}(t,u_0,t) \int\int K_1''(v) h_2^2 w^2 K_2(w) dv dw = 0 \\
T_6 &= \frac{1}{h_1^2}\phi^*_{xy}(t,u_0,t) \int\int h_1 v K_1''(v) h_2 w K_2(w) dv dw = 0\ .
\end{aligned}
$$

Now we will analyze T_7.

$$
\begin{aligned}
T_{7,1} &= \frac{1}{h_1^2}\phi^*_{x^3}(t,u_0,t)\int\int h_1^3v^3K_1''(v)K_2(w)dvdw = 0\\
T_{7,2} &= \frac{1}{h_1^2}\phi^*_{y^3}(t,u_0,t)\int\int K_1''(v)h_2^3w^3K_2(w)dvdw = 0\\
T_{7,3} &= \frac{1}{h_1^2}\phi^*_{x^2y}(t,u_0,t)\int\int h_1^2v^2K_1''(v)h_2wK_2(w)dvdw = 0\\
T_{7,4} &= \frac{1}{h_1^2}\phi^*_{xy^2}(t,u_0,t)\int\int h_1vK_1''(v)h_2^2w^2K_2(w)dvdw = 0\\
T_{7,5} &= \frac{1}{h_1^2}\int\int \phi^*_{x^4}(\bar{c_1},\bar{c_2},t)h_1^4v^4K_1''(v)K_2(w)dvdw\\
T_{7,6} &= \frac{1}{h_1^2}\int\int \phi^*_{y^4}(\bar{c_3},\bar{c_4},t)K_1''(v)h_2^4w^4K_2(w)dvdw\\
T_{7,7} &= \frac{1}{h_1^2}\int\int \phi^*_{x^3y}(\bar{c_5},\bar{c_6},t)h_1^3v^3K_1''(v)h_2wK_2(w)dvdw\\
T_{7,8} &= \frac{1}{h_1^2}\int\int \phi^*_{x^2y^2}(\bar{c_7},\bar{c_8},t)h_1^2v^2K_1''(v)h_2^2w^2K_2(w)dvdw\\
T_{7,9} &= \frac{1}{h_1^2}\int\int \phi^*_{xy^3}(\bar{c_9},\bar{c}_{10},t)h_1vK_1''(v)h_2^3w^3K_2(w)dvdw
\end{aligned}
$$

where $\bar{c_1},\ldots,\bar{c}_{10}$ are the corresponding in between values of the Taylors expansion of ϕ^* . Note that $||X||_\infty \le 1$ and θ is in a compact set and hence θ^TX will be in a bounded interval. Also recall that $U \in [0,1]$. Therefore $\phi^*(s,u,t) = \psi(s,u,t)f_{\theta^TX_1,U|\theta^TX=t}(s,u|t)$ is defined on compact set. So if we assume $\phi(s,u,t)$ is continuous in t, then all the order terms in $T_{7,5},\ldots,T_{7,9}$ are free of t and their magnitudes are listed below.

$$T_{7,5} = O(h_1^2),\quad T_{7,6} = O(h_1^{-2}h_2^4),\quad T_{7,7} = O(h_1h_2),\quad T_{7,8} = O(h_2^2),\quad T_{7,9} = O(h_1^{-1}h_2^3).$$

Therefore we get $\phi_2^*(t) = \phi^*_{xx}(t,u_0,t) + \sum_{k=5}^9 T_{7,k}$ and hence

$$E\left[\left\{\frac{\partial^2 A_n(\theta^TX,\theta)}{\partial\theta\partial\theta^T}\right\}_{(1,1)}\right] = \int \phi^*_{xx}(s,u_0,s)f_{\theta^TX}(s)ds + \sum_{k=5}^{9} T_{7,k}$$

which yields $II = a_{n2,2}$ where

$$a_{n2,2} = O(h_1^2) + O(h_2^2) + O(h_1 h_2) + O(h_1^{-1} h_2^3) + O(h_1^{-2} h_2^4) \ .$$

Finally we get the convergence rate $a_{n2} = a_{n2,1} + a_{n2,2}$ which completes the proof. $\square$

Appendix B Technical Details for Quantile Regression Model

Test Statistic

Let

$$\hat{R}_i(t) = \frac{1}{N}\sum_{j=1}^{n_i} \hat{U}_{ij} I(X_{ij} \le t) \tag{26}$$

where $N = n_1 + n_2$ and $\hat{g}_\tau(\cdot)$ being the pooled sample local linear nonparametric quantile regression function estimator. We will show that the marked empirical process defined in (26) converges weakly to a Gaussian process in the space $D[0,1]$ with proper normalization. To facilitate our arguments we assume the following.

B1 : The quantile regression functions are twice continuously differentiable with bounded derivatives.

B2 : The design densities are supported on $[0,1]$ and are denoted by $f_{X_i}(\cdot)$ for $i = 1, 2$.

B3 : The conditional density of ϵ_i given the covariates denoted by $f_{\epsilon_i|x_i}(\cdot)$ is twice differentiable with uniformly bounded derivatives with $f_{\epsilon_i|x_i}(0) > 0$. Moreover we assume $f_{\epsilon_i^2|x_i}(\cdot)$ is uniformly bounded and $f_{\epsilon_i^2|x_i}(0) > 0$.

B4 : The bandwidth h used in the local linear estimation satisfy as $N \to \infty$

$$h \to 0, \;\; Nh^8 \to 0, Nh^2 \to \infty \; .$$

B5 : The individual sample sizes satisfy $\frac{n_i}{N} = c_i + O(\frac{1}{N})$ as $N \to \infty$, for $i = 1, 2$.

B.1 Limiting Process under H_0

We will show that $\sqrt{N}\hat{R}_i(t)$ converges weakly to a Gaussian process with mean $\mu(t) = 0$ and covariance function $H_i(t, s)$.

THEOREM 8. *Let conditions B1-B5 hold. Then under H_0 the marked empirical processes $\sqrt{N}\hat{R}_i(t)$ for $i = 1, \ldots, k$ converge weakly to independent mean zero Gaussian processes with covariance functions $H_i(t, s) = \tau(1-\tau)c_i F_{X_i}(t),\ t < s$ in the space $D[0, 1]$ as $N \to \infty$.*

PROOF OUTLINE

1. $\sqrt{N}R_i(t)$ converges weakly to a Gaussian process where $\sqrt{N}R_i(t)$ is the same process as in (26) with $\hat{U}_{ij}$ replaced by the true U_{ij}.

2. $\sup_{0 \le t \le 1} |\sqrt{N}\hat{R}_i(t) - \sqrt{N}R_i(t)| = o_p(1)$

PROOF:

First we will show the weak convergence of

$$\sqrt{N}R_i(t) = \frac{1}{\sqrt{N}} \sum_{j=1}^{n_i} U_{ij} I(X_{ij} \le t) \ .$$

Recall that under H_0, $U_{ij} = \eta_{ij}$ and it is easy to see $E(\eta_{ij}|X_{ij}) = 0$ for $i = 1, \ldots, k$. Therefore we have $E\left\{\sqrt{N}R_i(t)\right\} = 0$ for $t \in [0,1]$ under H_0.

To find the covariance of $\sqrt{N}R_i(t)$ we proceed as follows. Since $E\left\{\sqrt{N}R_i(t)\right\} = 0$ we have

$$Cov\left\{\sqrt{N}R_i(t), \sqrt{N}R_i(s)\right\} = E\left\{\sqrt{N}R_i(t).\sqrt{N}R_i(s)\right\} \ .$$

Then we get

$$\begin{aligned}
E\left\{\sqrt{N}R_i(t).\sqrt{N}R_i(s)\right\} &= \frac{1}{N}\sum_{j=1}^{n_i}\eta_{ij}I(X_{ij} \le t).\sum_{j=1}^{n_i}\eta_{ij}I(X_{ij} \le s) \\
&= \frac{1}{N}\Bigg[E\left\{\eta_{i1}^2 I(X_{i1} \le t)I(X_{i1} \le s) + \cdots + \eta_{in_i}^2 I(X_{in_i} \le t)I(X_{in_i} \le s)\right\} \\
&\qquad + E\left(cross\ products\right)\Bigg] \\
&= \frac{1}{N}\left[n_i E\left\{\eta_{i1}^2 I(X_{i1} \le t)I(X_{i1} \le s)\right\} + 0\right] \\
&= \frac{n_i}{N}E\Bigg\{E(\cdot|X_{i1})\Bigg\} \\
&= \frac{n_i}{N}E\Bigg(I(X_{i1} \le t)I(X_{i1} \le s)E\left[\left\{I(\epsilon_{i1} \le 0) - \tau\right\}^2 |X_{i1}\right]\Bigg) \\
&= \frac{n_i}{N}E\Bigg\{I(X_{i1} \le t)I(X_{i1} \le s)\tau(1-\tau)\Bigg\} \\
&= \frac{n_i}{N}\tau(1-\tau)F_{X_i}(t \wedge s) \ .
\end{aligned}$$

Then we get

$$Cov\left\{\sqrt{N}R_i(t), \sqrt{N}R_i(s)\right\} \quad = \tau(1-\tau)c_i F_{X_i}(t \wedge s) + o(1) \ .$$

This implies that

$$\left\{\sqrt{N}R_i(t_1), \cdots, \sqrt{N}R_i(t_k)\right\} \xrightarrow{D} N\Bigg(0, V_\tau(t_1, \cdots, t_k)\Bigg)$$

by the central limit theorem . Weak convergence of $\sqrt{N}R_i(t)$ now follows if we can show the following moment condition (Billingsley, 1968).

$$N^2 E\left[\{R_i(w) - R_i(v)\}^2 \{R_i(v) - R_i(u)\}^2\right] \leq C(w-u)^2 \ , \forall \ 0 \leq u \leq v \leq w \leq 1. \quad (27)$$

$$\begin{aligned}
E\left[\{R_i(w) - R_i(v)\}^2 \{R_i(v) - R_i(u)\}^2\right] \quad &= E\left\{\left(\sum_{j=1}^{n_i} \alpha_j\right)^2 \left(\sum_{j=1}^{n_i} \beta_j\right)^2\right\} \\
&= n_i E(\alpha_1^2 \beta_1^2) + n_i(n_i - 1)E(\alpha_1^2 \beta_2^2) \\
&\quad +2n_i(n_i - 1)E(\alpha_1 \beta_1 \alpha_2 \beta_2) \\
&= n_i E(\alpha_1^2 \beta_1^2) + n_i(n_i - 1)E(\alpha_1^2)E(\beta_2^2) \\
&\quad +2n_i(n_i - 1)E(\alpha_1 \beta_1)E(\alpha_2 \beta_2)
\end{aligned}$$

where the last two equalities follow by noting that

$$\alpha_j = \tfrac{1}{N} U_{ij} \left\{ I\left(X_{ij} \leq w\right) - I\left(X_{ij} \leq v\right) \right\}$$

$$\beta_j = \tfrac{1}{N} U_{ij} \left\{ I\left(X_{ij} \leq v\right) - I\left(X_{ij} \leq u\right) \right\}$$

are α_j and β_j are i.i.d. and α_j and β_k are independent for $j \neq k$. Now consider $E(\alpha_1^2 \beta_1^2)$.

$$\begin{aligned} E(\alpha_1^2 \beta_1^2) &= \tfrac{1}{N^4} E\left[U_{i1} \left\{ I\left(X_{i1} \leq w\right) - I\left(X_{i1} \leq v\right) \right\}^2 U_{i1} \left\{ I\left(X_{i1} \leq v\right) - I\left(X_{i1} \leq u\right) \right\}^2 \right] \\ &= 0 \end{aligned}$$

where the last equality follows by recalling that $0 \leq u \leq v \leq w \leq 1$. Similarly we get $E(\alpha_1 \beta_1) = 0$ and $E(\alpha_2 \beta_2) = 0$. Now consider $E(\alpha_1^2)$.

$$\begin{aligned} E(\alpha_1^2) &= E\left[\frac{1}{N^2} U_{i1} \left\{ I\left(X_{i1} \leq w\right) - I\left(X_{i1} \leq v\right) \right\}^2 \right] \\ &= \frac{1}{N^2} E \Big\{ E(\cdot | X_{i1}) \Big\} \\ &= \frac{1}{N^2} E\left[\left\{ I\left(X_{i1} \leq w\right) - I\left(X_{i1} \leq v\right) \right\}^2 E(U_{i1} | X_{i1}) \right] . \end{aligned} \tag{28}$$

Recall that under H_0 , $U_{ij} = \eta_{ij}$ and therefore

$$E(U_{i1} | X_{i1}) = E(\eta_{i1}^2 | X_{i1}) = \tau(1 - \tau) . \tag{29}$$

Substituting (29) in (28) we get

$$
\begin{aligned}
E(\alpha_1^2) &= \tfrac{1}{N^2} E\left[\{I(X_{i1} \le w) - I(X_{i1} \le v)\}^2 \tau(1-\tau)\right] \\
&= \frac{1}{N^2}\tau(1-\tau)\int_0^1 \{I(x \le w) - I(x \le v)\}^2 \, dF_{X_i}(x) \\
&= \frac{1}{N^2}\tau(1-\tau)\int_v^w dF_{X_i}(x) \\
&\le \tfrac{1}{N^2}\tau(1-\tau)O(w-v) \ .
\end{aligned}
$$

Similarly we get

$$E(\beta_1^2) \ \le \tfrac{1}{N^2}\tau(1-\tau)O(v-u) \ .$$

Then we have

$$
\begin{aligned}
E\left[\{R_i(w) - R_i(v)\}^2 \{R_i(v) - R_i(u)\}^2\right] &\le \frac{\tau^2(1-\tau)^2}{N^4}\left[n_i(n_i-1)O(w-v)O(v-u)\right] \\
&= O\left(\frac{1}{N^2}\right)(w-u)^2
\end{aligned}
$$

which establishes (27) and thereby show that $\sqrt{N}R_i(t)$ converges weakly to a mean zero Gaussian process on $D[0,1]$. To complete the proof it remains to show that

$$\sup_{0\le t\le 1} |\sqrt{N}\hat{R}_i(t) - \sqrt{N}R_i(t)| = o_p(1) \ .$$

To this end consider the following decomposition.

$$\sqrt{N}\hat{R}_i(t) = \sqrt{N}R_i(t) + W_i(t)$$

where for $i = 1, 2$

$$W_i(t) = \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}(\hat{U}_{ij} - U_{ij})I(X_{ij} \leq t)\ .$$

We will show $\sup_{0\leq t\leq 1}|W_i(t)| = o_p(1)$. Recall that

$$\hat{U}_{ij} \quad = I\{Y_{ij} \leq \hat{g}_\tau(X_{ij})\} - \tau$$

$$U_{ij} \quad = I\{Y_{ij} \leq g_\tau(X_{ij})\} - \tau$$

and $U_{ij} = \eta_{ij}$ under H_0. To shorten our proof, in what follows we assume that $\hat{g}_\tau$ is the leave-one-out pooled sample quantile function estimator. In Lemma 1 we will show that the limiting behavior of the test statistic if we had used the full sample quantile function estimator. i.e. We will show

$$\sup_{0\leq t\leq 1}|\sqrt{N}R_i^{\hat{g}_\tau^{-1}}(t) - \sqrt{N}R_i^{\hat{g}_\tau}(t)| = o_p(1)$$

where $\hat{g}_\tau^{-1}$ is the leave-one-out pooled sample quantile function estimator and $\hat{g}_\tau$ is the quantile function estimator based on the full sample. As stated in what follows

$\hat{g}_\tau = \hat{g}_\tau^{-1}$.

$$\begin{aligned}
\sup_{0\le t\le 1}|W_i(t)| &\le \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I(\hat{\epsilon}_{ij}\le 0)-\tau-I(\epsilon_{ij}\le 0)+\tau\right| \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I(\hat{\epsilon}_{ij}\le 0)-I(\epsilon_{ij}\le 0)\right| \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I\left\{\epsilon_{ij}\le \hat{g}_\tau(X_{ij})-g_\tau(X_{ij})\right\}-I(\epsilon_{ij}\le 0)\right| \\
&\le \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i} I\left(|\epsilon_{ij}|\le |\hat{g}_\tau(X_{ij})-g_\tau(X_{ij})|\right)
\end{aligned} \tag{30}$$

where the last inequality follows by the identity

$$|I(X\le Y)-I(X\le 0)|\le I(|X|\le |Y|)\ . \tag{31}$$

Now we will find a probability bound for each term of the sum given in (30). For $0\le\delta\le 1$,

$$\begin{aligned}
P[I(|\epsilon_{ij}|\le|\hat{g}_\tau(X_{ij})-g_\tau(X_{ij})|>\delta] &\le \frac{1}{\delta}E\left\{I(|\epsilon_{ij}|\le|\hat{g}_\tau(X_{ij})-g_\tau(X_{ij})|\right\} \\
&= \frac{1}{\delta}E\left\{I(|\epsilon_{ij}|^2\le|\hat{g}_\tau(X_{ij})-g_\tau(X_{ij})|^2\right\} \\
&= \frac{1}{\delta}E_{\mathbb{X}}\left(P\left[\epsilon_{ij}^2\le\{\hat{g}_\tau(X_{ij})-g_\tau(X_{ij})\}^2\,|\mathbb{X}\right]\right) \\
&= \frac{1}{\delta}E_{\mathbb{X}}\left[\int_{-\infty}^{\infty}\int_0^{\{s-g_\tau(x_{ij})\}^2} f_{\epsilon^2|X}(u)du f_{\hat{g}_\tau|X}(s)ds\right]\ .
\end{aligned}$$

Using the boundedness condition of $f_{\epsilon^2|X}(\cdot)$ (condition B3), we get

$$
\begin{aligned}
P[I(|\epsilon_{ij}| \leq |\hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})| > \delta] &\leq \frac{M}{\delta} E_{\mathbb{X}} \left[\int_{-\infty}^{\infty} \{s - g_\tau(x_{ij})\}^2 f_{\hat{g}_\tau|X}(s) ds \right] \\
&= \frac{M}{\delta} E_{\mathbb{X}} \left(E\left[\{\hat{g}_\tau(x_{ij}) - g_\tau(x_{ij})\}^2 | \mathbb{X} \right] \right) \\
&\leq \frac{M}{\delta} E_{\mathbb{X}} \left\{ O_p(h^4) + O_p\left(\frac{1}{Nh} \right) \right\} \\
&= O(h^4) + O\left(\frac{1}{Nh} \right)
\end{aligned}
$$

where the last inequality follows by results of Yu & Jones (2003) on local linear quantile regression estimation and the boundedness of the predictor variables (condition B2). Now using condition B4 and B5 we get for $i = 1, 2$,

$$
\begin{aligned}
\sup_{0 \leq t \leq 1} |W_i(t)| &\leq \frac{1}{\sqrt{N}} \sum_{j=1}^{n_i} O(h^4) + O\left(\frac{1}{Nh} \right) \\
&= \frac{n_i}{\sqrt{N}} \left[O(h^4) + O\left(\frac{1}{Nh} \right) \right] \\
&= o_p(1) \ . \qquad \square
\end{aligned}
$$

B.2 Local Power

This section establishes the non-trivial power of our test statistics under local alternatives of the form

$$H_1 : g_{\tau,i}(\cdot) = g_{\tau,1}(\cdot) + \Delta_i(\cdot)/\sqrt{N}, i = 1, \ldots, k. \tag{32}$$

The following theorem shows the consistency of the proposed test statistics.

THEOREM 9. *Let Assumption 1-5 hold. Then under local alternatives of the form in (32), the marked empirical process* $\sqrt{N}\hat{R}_i(\cdot)$ *converges weakly to a Gaussian process in the space* $D[0,1]$ *with a mean function*

$$\mu(t) = c_i \int_0^t \lambda_i(x)\Delta_i(x) f_{\epsilon_i|X_i=x}(0) f_{X_i}(x)dx > 0$$

and covariance function $H_i(t,s) = \tau(1-\tau)c_i F_{X_i}(t), t < s$*, where* $\lambda_i(x)$ *is defined in (2.2).*

Proof. Following the same proof outline as in theorem 8, first we will show the weak convergence of $\sqrt{N}R(t)$ to a Gaussian process with a non zero mean function. First we will find the $\sqrt{N}R_i(t)$ under H_1.

$$\begin{aligned} E\left\{V_i(t)\right\} &= \frac{1}{\sqrt{N}} E \sum_{j=1}^{n_i} U_{ij} I(X_{ij} \leq t) \\ &= \tfrac{n_i}{\sqrt{N}} E\Big\{ U_{ij} I(X_{ij} \leq t) \Big\} \\ &= \tfrac{n_i}{\sqrt{N}} E\Big\{ E(\cdot|X_{ij}) \Big\} \\ &= \tfrac{n_i}{\sqrt{N}} E\Big\{ I(X_{ij} \leq t) E(U_{ij}|X_{ij}) \Big\} . \end{aligned}$$

Now consider $E(U_{ij}|X_{ij})$ under H_1. Recall that, under H_1,

$$U_{ij} = \eta_{ij} + I\left\{Y_{ij} < g_\tau(X_{ij})\right\} - I\left\{Y_{ij} < g_{\tau,i}(X_{ij})\right\} \; .$$

Therefore

$$\begin{aligned} E(U_{ij}|X_{ij}) &= 0 + E[I\left\{Y_{ij} \leq g_\tau(X_{ij})\right\}|X_{ij}] - \tau \\ &= E[I\left\{\epsilon_{ij} \leq g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\right\}|X_{ij}] - \tau \\ &= P\left\{\epsilon_{ij} \leq g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})|X_{ij}\right\} - \tau \\ &= F_{\epsilon|X}(0) + \left\{g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\right\} f_{\epsilon|X} + \left\{g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\right\}^2 f'_{\epsilon|X}(\bar{g}) - \tau \end{aligned}$$

where $\bar{g}$ is between 0 and $\{g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\}$. Under H_1 and from (2.2) we have

$$g_\tau(t) - g_{\tau,i}(t) = \frac{\lambda_i(t)\Delta(t)}{\sqrt{N}} \ . \tag{33}$$

Using (33) we get

$$\begin{aligned} E\sqrt{N}R_i(t) &= \frac{n_i}{\sqrt{N}} E\left(\left[f_{\epsilon_i|X_i}(0)\frac{\lambda_i(X_{ij})\Delta(X_{ij})}{\sqrt{N}} + \{g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\}^2 f'_{\epsilon_i|X_i}(\bar{g})\right] I(X_{ij} \leq t)\right) \\ &= \frac{n_i}{N}\int_0^t \lambda_i(x)\Delta(x)f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx + O\left(\frac{1}{\sqrt{N}}\right) \\ &= c_i\int_0^t \lambda_i(x)\Delta(x)f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx + O\left(\frac{1}{\sqrt{N}}\right) \end{aligned}$$

where we have used condition B3 to bound $f'_{\epsilon|X}(\bar{g})$ which yields

$$E\sqrt{N}R_i(t) = c_i\int_0^t \lambda_i(x)\Delta(x)f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx + O\left(\frac{1}{\sqrt{N}}\right)$$

for $t \in [0,1]$. To have non-trivial power for our test we need ensure that the above expectation is non zero for some t. To this end, note that $\lambda_i(t) \in (0,1)$ and therefore the above integral will be zero for all $t \in (0,1)$ if and only if H_0 is true. Therefore under H_1 we will have non-trivial power in our test.

To find the covariance of $\sqrt{N}R_i(t)$ under H_1 we note from our previous calculation that

$$\begin{aligned} E\sqrt{N}R_i(t)E\sqrt{N}R_i(s) &= c_i^2\left[\int_0^t \lambda_i(x)\Delta(x)f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx \int_0^s \lambda_i(x)\Delta(x)f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx \right. \\ &\quad \left. +O\left(\frac{1}{\sqrt{N}}\right)\right] \ . \end{aligned} \tag{34}$$

To find $E\left\{\sqrt{N}R_i(t)\sqrt{N}R_i(s)\right\}$ under H_1 consider the decomposition of $E\left\{\sqrt{N}R_i(t)\sqrt{N}R_i(s)\right\}$ as in theorem 6.

$$E\left\{\sqrt{N}R_i(t)\sqrt{N}R_i(s)\right\}$$

$$= \frac{1}{N}\sum_{j=1}^{n_i} U_{ij} I(X_{ij} \leq t) \sum_{j=1}^{n_i} U_{ij} I(X_{ij} \leq s)$$

$$= \frac{1}{N}\Bigg[E\Bigg\{U_{i1}^2 I(X_{i1} \leq t) I(X_{i1} \leq s) + \cdots + U_{in_i}^2 I(X_{in_i} \leq t) I(X_{in_i} \leq s)$$

$$+E(cross\ products)\Bigg] \tag{35}$$

$$= \frac{1}{N}\left[n_i E\left\{U_{i1}^2 I(X_{i1} \leq t) I(X_{i1} \leq s)\right\} + n_i(n_i-1)E(cross\ product)\right]$$

$$= \frac{1}{N}\Bigg\{n_i T_1 + n_i(n_i-1)T_2\Bigg\} \ . \tag{36}$$

Consider T_1.

$$\begin{aligned} E\left\{U_{i1}^2 I(X_{i1} \leq t) I(X_{i1} \leq s)\right\} &= E\Bigg\{E(\cdot|X_{i1})\Bigg\} \\ &= E\Bigg\{I(X_{i1} \leq t) I(X_{i1} \leq s) E(U_{i1}^2|X_{i1})\Bigg\} \end{aligned} \tag{37}$$

and

$$\begin{aligned} E(U_{i1}^2|X_{i1}) &= E\left(\left[I\left\{Y_{i1} \leq g_\tau(X_{i1})\right\} - \tau\right]^2 \Bigg| X_{i1}\right) \\ &= E\left(I^2 + \tau^2 - 2\tau I \Bigg| X_{i1}\right) \\ &= E\left(I + \tau^2 - 2\tau I \Bigg| X_{i1}\right) \\ &= \tau^2 + (1-2\tau) E\left(I|X_{i1}\right) \ . \end{aligned} \tag{38}$$

Note that

$$\begin{aligned}
E\left(I|X_{i1}\right) &= E\left[I\left\{\epsilon_{i1} \leq g_\tau(X_{i1}) - g_{\tau,i}(X_{i1})\right\}\Big|X_{i1}\right] \\
&= E\left(I\left[\epsilon_{i1} \leq \frac{\Delta(X_{i1})\lambda_i(X_{i1})}{\sqrt{N}}\right]\Big|X_{i1}\right) \\
&= F_{\epsilon|X}\left[\frac{\Delta(X_{i1})\lambda_i(X_{i1})}{\sqrt{N}}\right] \\
&= F_{\epsilon|X}(0) + \frac{\Delta(X_{i1})\lambda_i(X_{i1})}{\sqrt{N}} f_{\epsilon|X}(0) + \frac{\left[\Delta(X_{i1})\lambda_i(X_{i1})\right]^2}{N} f'_{\epsilon|X}(\bar{c}) \qquad (39)
\end{aligned}$$

where $\bar{c}$ is between 0 and $\frac{\Delta(X_{i1})\lambda_i(X_{i1})}{\sqrt{N}}$. Substituting (39) in (38) we get

$$\begin{aligned}
E(U_{i1}^2|X_{i1}) &= \left[\tau(1-\tau) + (1-2\tau)\left\{\frac{\Delta(X_{i1})\lambda_i(X_{i1})}{\sqrt{N}} f_{\epsilon_1|X_1}(0)\right.\right. \\
&\quad \left.\left. + \frac{\left[\Delta(X_{i1})\lambda_i(X_{i1})\right]^2}{N} f'_{\epsilon_i|X_i}(\bar{c})\right\}\right] \qquad (40)
\end{aligned}$$

and substituting this in (37) we get

$$\begin{aligned}
T_1 &= E\Bigg(I(X_{i1} \leq t)I(X_{i1} \leq s)\left[\tau(1-\tau) + (1-2\tau)\left\{\tfrac{\Delta(X_{i1})\lambda_i(X_{i1})}{\sqrt{N}} f_{\epsilon_i|X_i}(0)\right.\right. \\
&\quad \left.\left. + \frac{\left[\Delta(X_{i1})\lambda_i(X_{i1})\right]^2}{N} f'_{\epsilon_i|X_i}(\bar{c})\right\}\right] \\
&= \tau(1-\tau)F_{X_1}(t \wedge s) + \frac{(1-2\tau)}{\sqrt{N}} \int_0^{t\wedge s} \Delta(x)\left\{\lambda(x) - 1\right\} f_{\epsilon_i|X_i}(0) f_{X_i}(x)dx \\
&\quad + \frac{(1-2\tau)^2}{N} \int_0^{t\wedge s} \left[\Delta(x)\lambda_i(x)\right]^2 f_{\epsilon_i|X_i}(0) f'_{\epsilon_i|X_i}(\bar{c}) f_{X_i}(x)dx \\
&= \tau(1-\tau)F_{X_i}(t \wedge s) + O\left(\frac{1}{\sqrt{N}}\right).
\end{aligned}$$

Now consider T_2.

$$
\begin{aligned}
T_2 &= E\Bigg\{U_{i1}I(X_{i1} \leq t)U_{i2}I(X_{i2} \leq s)\Bigg\} \\
&= E\Bigg\{U_{i1}I(X_{i1} \leq t)\Bigg\}E\Bigg\{U_{i2}I(X_{i2} \leq s)\Bigg\}
\end{aligned}
$$

where the last equality follows by independence. Now consider the first term in the above equation.

$$
\begin{aligned}
E\Bigg\{U_{i1}I(X_{i1} \leq t)\Bigg\} &= E\Bigg\{E(\cdot|X_{i1})\Bigg\} \\
&= E\Bigg\{I(X_{i1} \leq t)E(U_{i1}|X_{i1})\Bigg\}
\end{aligned}
$$

where

$$
\begin{aligned}
E\Bigg(U_{i1}\Bigg|X_{i1}\Bigg) &= E\Bigg[I\{\epsilon_{i1} \leq g_\tau(X_{i1}) - g_{\tau,i}(X_{i1})\} - \tau\Bigg|X_{i1}\Bigg] \\
&= E(I|X_{i1}) - \tau \\
&= F_{\epsilon_i|X_i}(0) + \frac{\Delta(X_{i1})\lambda_i(X_{i1})}{\sqrt{N}}f_{\epsilon_i|X_i}(0) + \frac{[\Delta(X_{i1})\lambda_i(X_{i1})]^2}{N}f'_{\epsilon_i|X_i}(\bar{c}) - \tau
\end{aligned}
$$

where the last equality follows by substituting the value of $E(I|X_{i1})$ from equation

(39). Then we have

$$E\Big\{U_{i1}I(X_{i1}\le t)\Big\} = \frac{1}{\sqrt{N}}\int_0^t \Delta(x)\lambda_i(x)f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx + O\bigg(\frac{1}{N}\bigg)\ .$$

Similarly we get

$$E\Big\{U_{i2}I(X_{i2}\le s)\Big\} = \frac{1}{\sqrt{N}}\int_0^s \Delta(x)\lambda_i(x)f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx + O\bigg(\frac{1}{N}\bigg)\ .$$

Therefore we have

$$T_2 = \frac{1}{N}\int_0^t \Delta(x)\lambda_i(x)f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx\int_0^s \Delta(x)\lambda_i(x)f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx + o\bigg(\frac{1}{N}\bigg)\ .$$

Now substituting back the values of T_1 and T_2 in (36) we get

$$\begin{aligned} E\Big\{\sqrt{N}R_i(t)\sqrt{N}R_i(s)\Big\} &= c_i\tau(1-\tau)F_{X_i}(t\wedge s) \\ &\quad + c_i^2\int_0^t \Delta(x)\lambda(x)_i f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx\int_0^s \Delta(x)\lambda_i(x)f_{\epsilon_i|X_i}(0)f_{X_i}(x)dx \\ &\quad + O\bigg(\frac{1}{\sqrt{N}}\bigg)\ . \end{aligned} \tag{41}$$

Therefore

$$\begin{aligned} Cov\Big\{\sqrt{N}R_i(t),\sqrt{N}R_i(s)\Big\} &= E\Big\{\sqrt{N}R_i(t)\sqrt{N}R_i(s)\Big\} - E\Big\{\sqrt{N}R_i(t)\Big\}E\Big\{\sqrt{N}R_i(s)\Big\} \\ &= (41)-(34) \\ &= c_i\tau(1-\tau)F_{X_i}(t\wedge s) + o(1)\ . \end{aligned}$$

This implies that

$$\left\{\sqrt{N}R_i(t_1),\cdots,\sqrt{N}R_i(t_k)\right\} \xrightarrow{D} N\left\{\mu_i(t), V_\tau(t_1,\cdots,t_k)\right\}$$

by the central limit theorem. Weak convergence of $\sqrt{N}R_i(t)$ now follows if we establish the moment condition given in (27) under H_1. Following the setup in theorem 6, consider the expression in equation (28). Note that $E(U_{11}^{*^2}|X_{i1}) = (n_2^*)^2 E(U_{11}|X_{i1})$ and $E(U_{11}|X_{i1})$ under H_1 is given in (38). Substituting this in (28) we get

$$\begin{aligned} E(\alpha^2) &= \left(\tfrac{i}{N}\right)^2 E\left[\tau(1-\tau) + (1-2\tau)\left\{\tfrac{\Delta(X_{i1})\lambda_i(X_{i1})}{\sqrt{N}} f_{\epsilon_i|X_i}(0)\right.\right. \\ &\quad \left.\left. + \frac{[\Delta(X_{i1})\lambda_i(X_{i1})]^2}{N} f'_{\epsilon_i|X_i}(\bar{c})\right\}\right] \\ &= \frac{1}{N^2}\tau(1-\tau)\int_v^w dF_{X_i}(x) + O\left(\frac{1}{N^{5/2}}\right) \\ &= \tfrac{1}{N^2}\tau(1-\tau)O(w-v) + O\left(\tfrac{1}{N^{5/2}}\right). \end{aligned}$$

Similarly we get

$$E(\beta_1^2) = \tfrac{1}{N^2}\tau(1-\tau)O(v-u) + O\left(\tfrac{1}{N^{5/2}}\right)$$

As in theorem 6, $E(\alpha_1^2\beta_1^2) = E(\alpha_1\beta_1) = 0$ regardless of H_0 and H_1. This establishes (27) and thereby show that the process $\sqrt{N}R_i(t)$ converges weakly to a Gaussian process with mean $\mu_i(t) \neq 0$ on $D[0,1]$ under H_1. To complete the proof it remains to show that

$$\sup_{0\le t\le 1} |\sqrt{N}\hat{R}_i(t) - \sqrt{N}R_i(t)| = o_p(1)$$

under H_1. To this end consider the following decomposition as in theorem 6.

$$\sqrt{N}\hat{R}_i(t) = \sqrt{N}R_i(t) + W_i(t)$$

where for $i = 1, 2$

$$W_i(t) = \frac{1}{\sqrt{N}} \sum_{j=1}^{n_i} (\hat{U}_{ij} - U_{ij}) I(X_{ij} \le t) \ .$$

We will show $\sup_{0\le t\le 1} |W_i(t)| = o_p(1)$. Recall that

$$\begin{aligned}\hat{U}_{ij} &= n_{3-i}\Big[I\left\{Y_{ij} \le \hat{g}_\tau(X_{ij})\right\} - \tau\Big] \\ U_{ij} &= n_{3-i}\Big[I\left\{Y_{ij} \le g_\tau(X_{ij})\right\} - \tau\Big] \ .\end{aligned}$$

As in theorem 8, we assume that $\hat{g}_\tau$ is the leave-one-out pooled sample quantile function estimator.

$$\begin{aligned}
\sup_{0\le t\le 1}|W_i(t)| &\le \frac{n_{3-i}}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I\left\{Y_{ij}\le \hat{g}_\tau(X_{ij})\right\}-\tau-I\left\{Y_{ij}\le g_\tau(X_{ij})\right\}+\tau\right|\\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I\left\{\epsilon_{ij}\le \hat{g}_\tau(X_{ij})-g_{\tau,i}(X_{ij})\right\}-I\left\{\epsilon_{ij}\le g_\tau(X_{ij})-g_{\tau,i}(X_{ij})\right\}\right|\\
&\le \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I\left\{\epsilon_{ij}\le \hat{g}_\tau(X_{ij})-g_{\tau,i}(X_{ij})\right\}-I\Big(\epsilon_{ij}\le 0\Big)\right|\\
&\quad +\frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I\left\{\epsilon_{ij}\le g_\tau(X_{ij})-g_{\tau,i}(X_{ij})\right\}-I\Big(\epsilon_{ij}\le 0\Big)\right|\\
&= W_{i1}(t)+W_{i2}(t)\ .
\end{aligned}$$

We will analyze the two $W_{i1}(t)$ and $W_{i2}(t)$ for $i=1$. The calculations for $i=2$ use same arguments. First consider $W_{11}(t)$.

$$\begin{aligned}
W_{11}(t) &= \frac{n_2}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I\left\{\epsilon_{1j}\le \hat{g}_\tau(X_{ij})-g_{\tau,1}(X_{ij})\right\}-I\Big(\epsilon_{1j}\le 0\Big)\right|\\
&\le \frac{n_2}{\sqrt{N}}\sum_{j=1}^{n_i} I\Big(|\epsilon_{1j}|\le|\hat{g}_\tau(X_{ij})-g_{\tau,1}(X_{ij})|\Big)
\end{aligned} \tag{42}$$

where the last inequality follows by the identity (31). Now we will find a probability bound for each term of the sum given in (42). For $0 \leq \delta \leq 1$,

$$\begin{aligned}
P\left[I(|\epsilon_{1j}| \leq |\hat{g}_\tau(X_{ij}) - g_{\tau,1}(X_{ij})| > \delta\right] &\leq \frac{1}{\delta} E\left\{I(|\epsilon_{1j}| \leq |\hat{g}_\tau(X_{ij}) - g_{\tau,1}(X_{ij})|\right\} \\
&= \frac{1}{\delta} E\left\{I(|\epsilon_{1j}|^2 \leq |\hat{g}_\tau(X_{ij}) - g_{\tau,i}(X_{ij})|^2\right\} \\
&= \frac{1}{\delta} E_{\mathbb{X}}\left(P\left[\epsilon_{1j}^2 \leq \{\hat{g}_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\}^2 |\mathbb{X}\right]\right) \\
&= \frac{1}{\delta} E_{\mathbb{X}}\left[\int_{-\infty}^{\infty} \int_0^{\{s - g_{\tau,i}(X_{ij})\}^2} f_{\epsilon^2|X}(u) du f_{\hat{g}_\tau|X}(s) ds\right] .
\end{aligned}$$

Using the boundedness condition of $f_{\epsilon^2|X}(\cdot)$ (condition B3), we get

$$
\begin{aligned}
&P[I(|\epsilon_{1j}| \leq |\hat{g}_\tau(X_{ij}) - g_{\tau,i}(X_{ij})| > \delta] \\
&\leq \frac{M}{\delta} E_{\mathbb{X}} \left[\int_{-\infty}^{\infty} \{s - g_{\tau,i}(X_{ij})\}^2 f_{\hat{g}_\tau|X}(s) ds \right] \\
&= \frac{M}{\delta} E_{\mathbb{X}} \left(E\left[\{\hat{g}_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\}^2 | \mathbb{X} \right] \right) \\
&= \frac{M}{\delta} E_{\mathbb{X}} \left(E\left[\{\hat{g}_\tau(X_{ij}) - g_\tau(X_{ij}) + g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\}^2 | \mathbb{X} \right] \right) \\
&= \frac{M}{\delta} E_{\mathbb{X}} \Bigg(E\left[\{\hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})\}^2 \middle| \mathbb{X} \right] \\
&\quad + E\left[\{g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\}^2 \middle| \mathbb{X} \right] \\
&\quad + 2E\left[\{\hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})\}\{g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\} \middle| \mathbb{X} \right] \Bigg) \\
&\leq \frac{M}{\delta} E_{\mathbb{X}} \Bigg[O_p(h^4) + O_p\left(\frac{1}{Nh} \right) + \frac{\{\lambda(X_{ij}) - 1\}^2 \Delta^2(X_{ij})}{N} \\
&\quad + \frac{\{\lambda(X_{ij}) - 1\} \Delta(X_{ij})}{\sqrt{N}} O\left(h^2 \right) \Bigg]
\end{aligned}
$$

where the last inequality follows by results of Yu & Jones (2003) on local linear quantile regression estimation and equation (33). Now using boundedness of the predictor variables (condition B2) and condition B4 and B5 we get

$$W_{i1}(t) = O\left(\sqrt{N}h^4\right) + O\left(\frac{1}{\sqrt{N}h}\right) + O\left(\frac{1}{\sqrt{N}}\right) + O\left(h^2\right) = o(1) .$$

Now consider $W_{i2}(t)$.

$$\begin{aligned} W_{i2}(t) &= \frac{n_2}{\sqrt{N}} \sum_{j=1}^{n_i} \left| I\left\{\epsilon_{ij} \le g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\right\} - I\left(\epsilon_{ij} \le 0\right) \right| \\ &\le \frac{n_2}{\sqrt{N}} \sum_{j=1}^{n_i} I\left(|\epsilon_{1j}| \le |g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})| \right) \end{aligned} \tag{43}$$

where the last inequality follows by the identity (31). Now we will find a probability bound for each term of the sum given in (43). For $0 \le \delta \le 1$,

$$\begin{aligned} P\left[I(|\epsilon_{ij}| \le |g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})| > \delta \right] &\le \frac{1}{\delta} E\left\{ I(|\epsilon_{ij}| \le |g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})|\right\} \\ &= \frac{1}{\delta} E\left\{ I(|\epsilon_{ij}|^2 \le |\hat{g}_\tau(X_{ij}) - g_{\tau,i}(X_{ij})|^2 \right\} \\ &= \frac{1}{\delta} E_{\mathbb{X}} \left(P\left[\epsilon_{ij}^2 \le \{g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\}^2 \, |\mathbb{X} \right] \right) \\ &= \frac{1}{\delta} E_{\mathbb{X}} \left[\int_0^{\{g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\}^2} f_{\epsilon^2|X}(u) du \right] . \end{aligned}$$

Using the boundedness condition of $f_{\epsilon^2|X}(\cdot)$ (condition B3), we get

$$
\begin{aligned}
P[I(|\epsilon_{ij}| \leq |g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})| > \delta] &\leq \frac{M}{\delta} E_{\mathbb{X}} \left[\{g_\tau(X_{ij}) - g_{\tau,i}(X_{ij})\}^2 \right] \\
&\leq \frac{M}{\delta} E_{\mathbb{X}} \left[\frac{\lambda_i(X_{ij})^2 \Delta^2(X_{ij})}{N} \right]
\end{aligned}
$$

where the last inequality follows equation (33). Now using boundedness of the predictor variables (condition B2) we get

$$
W_{i2}(t) = O\left(\frac{1}{\sqrt{N}}\right) = o(1)
$$

which implies that $\sup_{0 \leq t \leq 1} |W_i(t)| = o_p(1)$ which completes the proof of theorem 7. □

So far we have assumed $\hat{g}_\tau = \hat{g}_\tau^{-1}$. To complete our arguments we will show

$$
\sup_{0 \leq t \leq 1} |\sqrt{N} R_i^{\hat{g}_\tau^{-1}}(t) - \sqrt{N} R_i^{\hat{g}_\tau}(t)| = o_p(1)
$$

where $\hat{g}_\tau^{-1}$ is the leave-one-out pooled sample quantile function estimator and $\hat{g}_\tau$ is the quantile function estimator based on the full sample.

Lemma 4. *Let $\hat{g}_\tau^{-1}$ be the leave-one-out pooled sample quantile function estimator and $\hat{g}_\tau$ is the quantile function estimator based on the full sample. Under conditions B1-B5 and assuming the conditional density of Y given $X = x$ is bounded away from zero such that $\inf_s f_{Y|X=x}(s) = c^*$ $\forall x$ we have as $N \to \infty$*

$$
\sup_{0 \leq t \leq 1} |\sqrt{N} R_i^{\hat{g}_\tau^{-1}}(t) - \sqrt{N} R_i^{\hat{g}_\tau}(t)| = o_p(1) \ .
$$

Proof.

$$\sup_{0\le t\le 1}|\sqrt{N}R_i^{\hat{g}_\tau^{-1}}(t)-\sqrt{N}R_i^{\hat{g}_\tau}(t)|=\sup_{0\le t\le 1}|T_i^{\hat{g}_\tau^{-1}}-T_1^{\hat{g}_\tau}|$$

where

$$T_1^{\hat{g}_\tau^{-1}} = \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left[I\left\{Y_{ij}\le \hat{g}_\tau^{-1}(X_{ij})\right\}-\tau\right]I\left(X_{ij}\le t\right)$$

$$T_i^{\hat{g}_\tau} = \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left[I\left\{Y_{ij}\le \hat{g}_\tau(X_{ij})\right\}-\tau\right]I\left(X_{ij}\le t\right)$$

and now we get

$$\begin{aligned}
&\sup_{0\le t\le 1}|\sqrt{N}R_i^{\hat{g}_\tau^{-1}}(t)-\sqrt{N}R_i^{\hat{g}_\tau}(t)| \\
&\le \sup_{0\le t\le 1}|T_i^{\hat{g}_\tau^{-1}}-T_i^{\hat{g}_\tau}| \\
&\le \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I\left\{Y_{ij}\le \hat{g}_\tau^{-1}(X_{ij})\right\}-I\left\{Y_{ij}\le \hat{g}_\tau(X_{ij})\right\}\right| \\
&= \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I\left\{\epsilon_{ij}\le \hat{g}_\tau^{-1}(X_{ij})-g_\tau(X_{ij})\right\}-I\left\{\epsilon_{ij}\le \hat{g}_\tau(X_{ij})-g_\tau(X_{ij})\right\}\right| \\
&\le \frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I\left\{\epsilon_{ij}\le \hat{g}_\tau^{-1}(X_{ij})-g_\tau(X_{ij})\right\}-I\left(\epsilon_{ij}\le 0\right)\right| \\
&\quad +\frac{1}{\sqrt{N}}\sum_{j=1}^{n_i}\left|I\left\{\epsilon_{ij}\le \hat{g}_\tau(X_{ij})-g_\tau(X_{ij})\right\}-I\left(\epsilon_{ij}\le 0\right)\right| \\
&= I_1+I_2\ .
\end{aligned}$$

We have already shown that $I_1 = o_p(1)$ (see equation (6) and what follows it) and therefore consider I_2. We will try to bound each term in I_2 in probability. Using the identity (31) we have

$$\begin{aligned} |I\{\epsilon_{ij} \le \hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})\} - I(\epsilon_{ij} \le 0)| &\le I(|\epsilon_{ij}| \le |\hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})|) \\ &= I(|\epsilon_{ij}|^2 \le |\hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})|^2) \end{aligned}$$

therefore

$$\begin{aligned} P\{|I\{\epsilon_{ij} \le \hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})\} - I(\epsilon_{ij} \le 0)| > \delta\} &\le P\{I(|\epsilon_{ij}|^2 \le |\hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})|^2) > \delta\} \\ &\le \frac{E\{I(|\epsilon_{ij}|^2 \le |\hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})|^2)\}}{\delta} \\ &= \frac{P\left[\epsilon_{ij}^2 \le \{\hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})\}^2\right]}{\delta} \quad . \end{aligned}$$

Now consider the numerator of the above equation.

$$\begin{aligned} P\left[\epsilon_{ij}^2 \le \{\hat{g}_\tau(X_{ij}) - g_\tau(X_{ij})\}^2\right] &= P\left[\epsilon_{ij}^2 \le \{\hat{g}_\tau(X_{ij}) - \hat{g}_\tau^{-1}(X_{ij}) + \hat{g}_\tau^{-1}(X_{ij}) - g_\tau(X_{ij})\}^2\right] \\ &= P\left(\epsilon_{ij}^2 \le 2\left[\{\hat{g}_\tau(X_{ij}) - \hat{g}_\tau^{-1}(X_{ij})\}^2 + \{\hat{g}_\tau^{-1}(X_{ij}) - g_\tau(X_{ij})\}^2\right]\right) \\ &\le P\left[\frac{\epsilon_{ij}^2}{2} \le 2\{\hat{g}_\tau(X_{ij}) - \hat{g}_\tau^{-1}(X_{ij})\}^2\right] \\ &\quad + P\left[\frac{\epsilon_{ij}^2}{2} \le 2\{\hat{g}_\tau^{-1}(X_{ij}) - g_\tau(X_{ij})\}^2\right] \\ &= T_1 + T_2 \quad . \end{aligned}$$

We have shown that $T_2 = O(h^4) + O(\frac{1}{Nh})$. Consider T_1. In order to analyze it we need a result which we call proposition 1.

Proposition 1. *For a random sample of size n from a distribution F_Y with density $f_Y(\cdot)$ such that $inf_t\ f_Y(t) = c$, and for a sequence of real number Δ_n that goes to zero as $n \to \infty$ we have*

$$P\left[|Q - Q^*| > \Delta_n\right] \leq (1 - \Delta_n)^n$$

where Q is the q^{th} quantile of the full sample and Q^* is the q^{th} quantile of the one removed sample.

Proof. We only give the proof for n being even and for $q = .5$ as the other cases follow similarly. Let M be the median of the full sample and M^* be the median of the one removed sample. Since removing one observation will only change the position of the median by one we have $|M - M^*| \leq Y_{(\frac{n}{2})+1} - Y_{(\frac{n}{2})}$ where $Y_{(k)}$ denote the k^{th} order statistic of the full sample. From the results of **?** on spacings we know for a uniform random sample,

$$P\left[U_{(\frac{n}{2})+1} - U_{(\frac{n}{2})} \leq t\right] = 1 - (1 - t)^n \quad ,0 < t < 1\ .$$

Therefore,

$$\begin{aligned} P\left[|M - M^*| > \Delta_n\right] &\leq P\left[Y_{(\frac{n}{2})+1} - Y_{(\frac{n}{2})} > \Delta_n\right] \\ &= P\left[F_Y^{-1}\left\{U_{(\frac{n}{2})+1}\right\} - F_Y^{-1}\left\{U_{(\frac{n}{2})}\right\} > \Delta_n\right] \\ &= P\left[\left\{U_{(\frac{n}{2})+1} - U_{(\frac{n}{2})}\right\} F_Y^{-1'}(\bar{c}) > \Delta_n\right] \end{aligned}$$

Note that F is increasing and hence F^{-1} is increasing and therefore $F^{-1'} \geq 0$. Also note that

$$F_Y^{-1'}(t) = \frac{1}{f_Y\left(F_Y^{-1}(t)\right)}$$

and using our assumption $\inf_t f_Y(t) = c$ we get

$$\begin{aligned} P\left[|M - M^*| > \Delta_n\right] &\leq P\left[\left\{U_{(\frac{n}{2})+1} - U_{(\frac{n}{2})}\right\} > c\Delta_n\right] \\ &= 1 - \left[1 - (1 - \Delta_n)^n\right] \\ &= (1 - \Delta_n)^n \ . \quad \square \end{aligned}$$

$\square$

Since $\hat{g}_\tau(t)$ is the τ^{th} quantile of Nh observations from the conditional distribution $F_{Y|X=t}$, using proposition 1 we get

$$P\left[\left|\hat{g}_\tau(t) - \hat{g}_\tau^{-1}(t)\right| > \Delta_n\right] \leq (1 - \Delta_n^{Nh}) \ . \tag{44}$$

Now consider T_1. For a sequence $\Delta_n \to 0$, we have

$$P\left[\epsilon_{ij}^2 \le 4\left\{\hat{g}_\tau(X_{ij}) - \hat{g}_\tau^{-1}(X_{ij})\right\}^2\right]$$
$$= P\left[\epsilon_{ij}^2 \le 4\left\{\hat{g}_\tau(X_{ij}) - \hat{g}_\tau^{-1}(X_{ij})\right\}^2, \left|\hat{g}_\tau(t) - \hat{g}_\tau^{-1}(t)\right| > \Delta_n\right]$$

$$+P\left[\epsilon_{ij}^2 \le 4\left\{\hat{g}_\tau(X_{1j}) - \hat{g}_\tau^{-1}(X_{1j})\right\}^2, \left|\hat{g}_\tau(t) - \hat{g}_\tau^{-1}(t)\right| \le \Delta_n\right]$$

$$\le P\left[\left|\hat{g}_\tau(t) - \hat{g}_\tau^{-1}(t)\right| > \Delta_n\right] + P\left[\epsilon_{ij}^2 \le 4\Delta_n^2\right]$$

$$\le (1-\Delta_n)^{Nh} + F_{\epsilon^2}(4\Delta_n^2)$$

$$= (1-\Delta_n)^{Nh} + 4\Delta_n^2 f_{\epsilon^2}(\bar{c})$$

where $\bar{c}$ is a number between 0 and $4\Delta_n^2$. Using the bounded condition on f_{ϵ^2} (B3) we require

$$\Delta_n \to 0 \tag{45}$$

$$\sqrt{N}(1-\Delta_n)^{Nh} \to 0 \tag{46}$$

$$\sqrt{N}\Delta_n^2 \to 0 \tag{47}$$

in order for $\sqrt{N}T_1 \to 0$. We can now pick our sequence Δ_n accordingly. For example if we use the mean square error optimal bandwidth $N^{-1/5}$ as our h and let $\Delta_n = N^{-\delta}$, then we would require $\delta > 1/4$ and this completes the proof of lemma (4). □

B.3 Test Statistic and Computation of Critical Values

For a size α test of $H_0 : g_{\tau,1} = g_{\tau,2}$ we propose to reject the null hypothesis if

$$\max_{1\leq i\leq 2}\left\{\sup_{0\leq t\leq 1}|T_i(t)|\right\} > C_\alpha$$

where

$$T_i(t) = \frac{|\sqrt{N}R_i(t)|}{\sqrt{\tau(1-\tau)(\frac{n_i}{N})}}$$

In order to compute the critical values for our test we use the following theorem.

THEOREM 10. *Let assumption B1-B5 hold. Then as $N \to \infty$ and under H_0,*

$$\sup_{0\leq t\leq 1}|\sqrt{N}\hat{R}_i(t)| \xrightarrow{D} \sup_{0\leq t\leq 1}|W(t)|$$

where $W(t)$ is the standard Brownian motion on $[0,1]$.

Proof. If $F_{X_i}(x) \equiv x$ (uniform designs)

$$\hat{R}_i^*(t) = \frac{1}{N}\sum_{j=1}^{n_i}\hat{U}_{ij}I(\xi_{ij} \leq t) \tag{48}$$

where ξ_{ij} denote the uniform X's, then from theorem (6) and continuous mapping theorem we know $\sup_t|T_i(t)| \xrightarrow{D} \sup_t|W(t)|$. Let ξ be i.i.d uniform random variables on $[0,1]$. Then

$$R_i(t) = \frac{1}{N}\sum_{j=1}^{n_i}\hat{U}_{ij}I\{F_{X_i}^{-1}(\xi_{ij}) \leq t\}$$

and $R_i(t) = R_i^*\{F_{X_i}(t)\}$. Following the proof of theorem 3.1 in Billingsley (1968) define $\psi : D \to D$ by $\psi x(t) = x(F_{X_i}(t))$. Then we see that ψ is continuous and therefore by the continuous mapping theorem and the fact that $R_i^* \Rightarrow W$ we have $\psi(R_i^*) \Rightarrow \psi(W)$. Again applying the continuous mapping theorem we see that

$\sup_t |T_i(t)| \xrightarrow{D} \sup_{0\le t\le 1} |W(F_{X_i}(t))|$. It remains to show that $\sup_{0\le t\le 1} |W_{F_{X_i}}(t)|$ and $\sup_t |W(t)|$ has the same distributions. Let $F_{X_i}(t) = s$ then

$$\begin{aligned}
\sup_{0\le t\le 1} |W(F_{X_i}(t))| &= \sup_{0\le F_{X_i}^{-1}(s)\le 1} |W(s)| \\
&= \sup_{F_{X_i}^{-1}(0)\le s\le F_{X_i}^{-1}(1)} |W(s)| \\
&= \sup_{0\le s\le 1} |W(s)|
\end{aligned}$$

where the last equality follows by noting that X_i's are on $[0, 1]$. □

Bibliography

Andrews, D. 1987. Consistency in non linear econometric models. *Econometrica*, **55**, 1465–471.

Billingsley, P. 1968. *Convergence of probability measures*. New York: John Wiley.

Bondell, H.D., Reich, B., & Wang, H. 2010. Non-crossing quantile regression curve estimation. *Biometrika*, **97(4)**, 825–838.

Cai, Z., & Xu, X. 2009. Nonparametric quantile estimations for dynamic smooth coefficient models. *J. am. statist. assoc.*, **104**, 371–383.

Cai, Z., Fan, J., & Li, R. 2000. Efficient estimation and inferences for varying-coefficient models. *J. am. statist. assoc.*, **95**, 888–902.

Cai, Z., Fan, J., Jiang, J., & Zhou, H. 2008. Partially linear hazard regression with varying coefficients for multivariate survival data. *J. r. statist. soc. b*, **70**, 141–158.

Chiou, J. M., & Müller, H. G. 1998. Quasi-likelihood regression with unknown link and variance functions. *J. am. statist. assoc.*, **93**, 1376–387.

Delgado, M.A. 1993. Testing the equality of nonparametric regression curves. *Statistics and probability letters*, **17**, 199–204.

Dette, H., & Volgushev, S. 2008. Non-crossing non-parametric estimates of quantile curves. *J. r. statist. soc. b*, **70(3)**, 609–627.

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

Fan, J., Huang, T., & Li, R. 2007. Analysis of longitudinal data with semiparametric estimation of covariance function. *J. am. statist. assoc.*, **35**, 632–641.

Hall, P., & Hart, J. 1990. Bootstrap test for difference between means in nonparametric regression. *J. am. statist. assoc.*, **85**, 1039–1049.

Hardle, W., & Mamen, E. 1993. Comparing nonparametric versus parametric regression fits. *Ann. statist.*, **21**, 1926–1947.

Hastie, T. J., & Tibshirani, R. J. 1993. Varying-coefficient models (with discussion). *J. r. statist. soc. b*, **55**, 757–96.

Horowitz, J. L. 2001. Nonparametric estimation of a generalized additive model with an unknown link function. *Econometrica*, **69**, 499–513.

Huang, J.Z., & Shen, H. 2004. Functional coefficient regression models for non-linear time series: A polynomial spline approach. *Scand. j. statist*, **31**, 809–22.

Ichimura, H. 1993. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. economet*, **58**, 71–120.

King, E.C., Hart, J.D., & Wehrly, T.E. 1991. Testing the equality of two regression curves using linear smoothers. *Statistics and probability letters*, **12**, 239–247.

Koenker, R., & Bassett, G. 1978. Regression quantiles. *Econometrica*, **46**, 33–50.

Kulasekera, K.B. 1995. Comparison of regression curves using quasi-residuals. *J. am. statist. assoc.*, **90**, 1085–1093.

Li, K. C., & Duan, N. 1989. Regression analysis under link violation. *Ann. statist.*, **17**, 1009–052.

Lin, W., & Kulasekera, K.B. 2007. Identifiability of single-index models and additive-index models. *Biometrika*, **94**, 496–501.

Neumeyer, N., & Dette, H. 2003. Nonparametric comparison of regression curves: An empirical process approach. *Ann. statist.*, **31**, 880–920.

Ramsay, J.O., & Silverman, B.W. 1998. *The Analysis of Functional Data*. Berlin:Springer-Verlag.

Sun, Y. 2006. A consistent nonparametric equality test of conditional quantile functions. *Econometric theory*, **22**, 614–632.

van der Vaart, A. W. 1998. *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Weisberg, S., & Welsh, A. H. 1994. Adapting for the missing link. *Ann. statist.*, **22**, 1674–700.

Yafeh, Y., & Yosha, O. 2003. Large shareholders and banks: Who monitors and how ? *Economic journal*, **113**, 128–46.

Yu, K., & Jones, M.C. 2003. Local linear quantile regression. *J. am. statist. assoc.*, **93**, 228–289.