CONDITIONAL VARIANCE FUNCTION CHECKING IN HETEROSCEDASTIC REGRESSION MODELS

by

NISHANTHA ANURA SAMARAKOON

B.S., University of Peradeniya, Sri Lanka, 1998
M.S., University of Peradeniya, Sri Lanka, 2002

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2011

Abstract

The regression model has been given a considerable amount of attention and played a significant role in data analysis. The usual assumption in regression analysis is that the variances of the error terms are constant across the data. Occasionally, this assumption of homoscedasticity on the variance is violated; and the data generated from real world applications exhibit heteroscedasticity. The practical importance of detecting heteroscedasticity in regression analysis is widely recognized in many applications because efficient inference for the regression function requires unequal variance to be taken into account. The goal of this thesis is to propose new testing procedures to assess the adequacy of fitting parametric variance function in heteroscedastic regression models.

The proposed tests are established in Chapter 2 using certain minimized L_2-distance between a nonparametric and a parametric variance function estimators. The asymptotic distribution of the test statistics corresponding to the minimum distance estimator under the fixed model and that of the corresponding minimum distance estimators are shown to be normal. These estimators turn out to be $\sqrt{n}-$consistent. The asymptotic power of the proposed test against some local nonparametric alternatives is also investigated. Numerical simulation studies are employed to evaluate the finite sample performance of the test in one dimensional and two dimensional cases.

The minimum distance method in Chapter 2 requires the calculation of the integrals in the test statistics. These integrals usually do not have a tractable form. Therefore, some numerical integration methods are needed to approximate the integrations. Chapter 3 discusses a nonparametric empirical smoothing lack-of-fit test for the functional form of the variance in regression models that do not involve evaluation of integrals. empirical smoothing lack-of-fit test can be treated as a nontrivial modification of Zheng (1996)'s nonparametric smoothing test and Koul and Ni (2004)'s minimum distance test for the mean function in the classic regression models. The asymptotic normality of the proposed test under the null hypothesis is established. Consistency at some fixed alternatives and asymptotic power under some local alternatives are also discussed. Simulation studies are conducted to assess the finite sample performance of the test. The simulation studies show that the proposed empirical smoothing test is more powerful and computationally more efficient than the minimum distance test and Wang and Zhou (2006)'s test.

CONDITIONAL VARIANCE FUNCTION CHECKING IN HETEROSCEDASTIC REGRESSION MODELS

by

NISHANTHA ANURA SAMARAKOON

B. S., University of Peradeniya, Sri Lanka, 1998
M. S., University of Peradeniya, Sri Lanka, 2002

A DISSERTATION

submitted in partial fulfillment of the
requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Statistics
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas
2011

Approved by;

Major Professor
Dr. Weixing Song

Copyright

Abstract

The regression model has been given a considerable amount of attention and played a significant role in data analysis. The usual assumption in regression analysis is that the variances of the error terms are constant across the data. Occasionally, this assumption of homoscedasticity on the variance is violated; and the data generated from real world applications exhibit heteroscedasticity. The practical importance of detecting heteroscedasticity in regression analysis is widely recognized in many applications because efficient inference for the regression function requires unequal variance to be taken into account. The goal of this thesis is to propose new testing procedures to assess the adequacy of fitting parametric variance function in heteroscedastic regression models.

The proposed tests are established in Chapter 2 using certain minimized L_2-distance between a nonparametric and a parametric variance function estimators. The asymptotic distribution of the test statistics corresponding to the minimum distance estimator under the fixed model and that of the corresponding minimum distance estimators are shown to be normal. These estimators turn out to be $\sqrt{n}-$consistent. The asymptotic power of the proposed test against some local nonparametric alternatives is also investigated. Numerical simulation studies are employed to evaluate the finite sample performance of the test in one dimensional and two dimensional cases.

The minimum distance method in Chapter 2 requires the calculation of the integrals in the test statistics. These integrals usually do not have a tractable form. Therefore, some numerical integration methods are needed to approximate the integrations. Chapter 3 discusses a nonparametric empirical smoothing lack-of-fit test for the functional form

of the variance in regression models that do not involve evaluation of integrals. empirical smoothing lack-of-fit test can be treated as a nontrivial modification of Zheng (1996)'s nonparametric smoothing test and Koul and Ni (2004)'s minimum distance test for the mean function in the classic regression models. The asymptotic normality of the proposed test under the null hypothesis is established. Consistency at some fixed alternatives and asymptotic power under some local alternatives are also discussed. Simulation studies are conducted to assess the finite sample performance of the test. The simulation studies show that the proposed empirical smoothing test is more powerful and computationally more efficient than the minimum distance test and Wang and Zhou (2006)'s test.

Table of Contents

List of Figures

List of Tables

Acknowledgments

This thesis came into being in part through years of academic and research that has been accomplished since I came to Kansas State University. I greatly appreciate the faculty, staff, and my colleagues for providing excellent academic environment. Specially, I would like to greatly and sincerely thank Dr. Weixing Song for his precious guidance, understanding, patience, and most importantly, his friendship during my graduate studies at Kansas State University. His mentorship is paramount in providing extraordinary experience towards my long-term career goals. I also greatly acknowledge Dr. James Neill for giving me wonderful help and guidance as the head of the department and a supervisory committee member. I would like to extend my sincere gratitude to Dr. Haiyan Wang, Dr. Jianming Yu, and Dr. Bala Natarajan for working as my doctoral committee, having valuable discussions, and accessibility.

I convey my special acknowledgment to Dr. S. Samita and Dr. W.B. Daundasekara of University of Peradeniya, and Dr. Nirmalie Wickramaratne, Dr. Tilak Hewawasam of Sabaragamuwa University, Sri Lanka for their great support and guidance while I was doing my studies and working as a faculty.

Where would I be without my family? My late parents deserve special mention for their inseparable support and guidance. My father, in the first place is the person who placed the foundation of my learning character, showing me the joy of intellectual pursuit ever since I was a child. My mother is the one who sincerely raised me with caring and gently love. All the brothers and sisters are greatly appreciated for being supportive and encouraging me with their best wishes.

I want to express my sincere thanks to my wife, Disna Samarakoon for her continuous support and love throughout my period of study. She was always there cheering me up and stood by me through the good and bad times. Finally, I would like to thank everybody who was important to the successful realization of the thesis, as well as expressing my apology that I could not mention each one personally.

Dedication

I would like to dedicate this thesis to My Late Amma (Mother) and Thaththa (Father) to express how much they mean to me; to my mother for her unfailing patience and loving support she has given to me throughout my life, and to my father for his love and encouragement for me to pursue higher education.

Chapter 1

Introduction

It is a common assumption of a regression model that all the random error terms are mutually independent with mean zero and equal variances. Occasionally, this assumption on the variance function (i.e. homoscedasticity) is not satisfied, while the real data generated from the applications often exhibits a certain non-constant variance (i.e. heteroscedasticity) structure. Heteroscedasticity is caused by many things such as data pooling, different levels of determination, different measurements of error, important variables that may be omitted from the model, as well as many others.

It is well known that when the assumptions of the linear regression models are correct, ordinary least squares (OLS) provide unbiased and efficient estimates of the parameters. If the errors are heteroscedastic, the OLS estimators remain unbiased but they are not however the best linear unbiased estimators (BLUE). Also hypothesis testing and confidence intervals which are based on the standard errors will not be correct as their assumptions are violated. The practical importance of detecting non-constant variance in the regression is now widely recognized among researchers and practitioners, in that the efficient statistical inference for the regression analysis should take the heteroscedasticity into account, when the homoscedasticity assumption fails.

The commonly used graphical methods of examining the assumption of homoscedasticity are based on the visual examination of residual plots (i.e. plots of residuals versus either the corresponding fitted values or explanatory variables, etc.) after fitting a parametric or nonparametric model. For example, a fan-shaped or double-bow pattern residual plots indicate non-constant variance.

1.1 Diagnostic Plots Detecting Unequal Error Variances: Blood Pressure Example

The following graph shows the diagnostic plot detecting unequal-variances in the relationship between diastolic blood pressure and age among healthy adult women 20 to 60 years old, collected data on 54 subjects (p.427,*Applied Linear Statistical Models* Kutner, Nachtsheim, Neter, and Li (2004)).

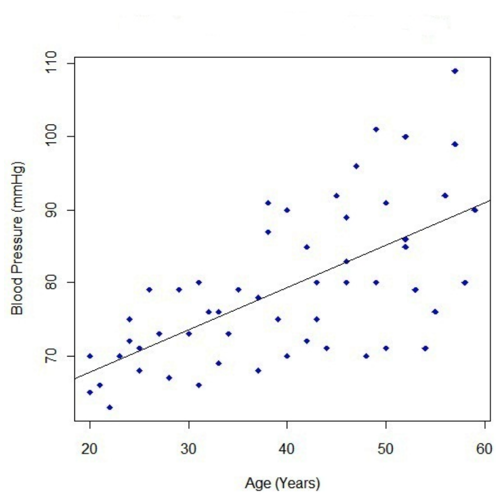

Figure 1.1: *Diagnostic Plots Detecting Unequal Error Variances*

The scatter plot of the data in Figure 1.1 strongly suggests a linear relationship between diastolic blood pressure and age, but also indicates that the error term variance increases with age, i.e. the heteroscedasticity exhibit in this data set is evident. This is a severe model assumption violation in ordinary least squares regression. This encourages us to include a variance function to the regression models and develop a goodness-of-fit testing procedures for checking the adequacy of the variance function.

A significant amount of statistical research has already been conducted in the area of checking for heteroscedasticity in a regression model. Most of the past researches in this area are based on checking whether the variance function is constant or not. However, there is little in the literature where we can find an examination of the adequacy of the variance function. This thesis can provide a contribution to the statistical analysis, namely to regression modeling devoted to the problem of heteroscedasticity in order to obtain efficient and reliable results.

1.2 Literature Review

In literature, a remarkable amount of statistical research has already been carried out for assessing the heteroscedasticity in both parametric and nonparametric regression models. Early works in this area include some graphical procedures and some formal tests, most of which are based on the residuals obtained by fitting a model with a completely specified regression and variance function. Harrison and McCabe (1979) propose a test for heteroscedasticity based on the direct use of ordinary least squares residuals from a single regression on the complete set of observations.

Breusch and Pagan (1979) suggest a simple test for heteroscedasticity in a linear regression model using the Lagrange multiplier test. In addition, White (1980) introduces a natural test for heteroscedasticity by comparing a parameter covariance matrix estimator which is consistent to the usual covariance matrix estimator, even when the errors of a linear regression models are heteroscedastic. Moreover, Koener and Basset (1981) suggest a class of tests for heteroscedasticity in linear regression models based on regression quantile statistics.

A diagnostic test for heteroscedasticity based on the score statistic is presented by Cook and Weisberg (1983) and a graphical procedure is used to implement the test. Most of the above tests have been proposed for checking whether a variance function is constant or not, but do not discuss whether a specific variance function can adequately describe the variability in the data.

Some authors have discussed heteroscedasticity tests in regression models with non parametric variance structures. Diblasi and Bowman (1997) propose a nonparametric test of constant variance for the errors in a linear regression model based on nonparametric smoothing of the residuals. Muller and Zhao (1995) propose a general semi-parametric variance function model in a fixed design regression setting. The regression function is assumed to be smooth and is modeled nonparametrically. The relationship between the mean regression function and the variance function is assumed to follow a generalized linear model. Eubank and Thomas (1993) propose some diagnostic tests and plots for detecting heteroscedasticity in the completely nonparametric regression model. However, this test requires an assumption of the normally distributed errors. When the covariate is one dimensional, Dette and Munk (1998) propose a simple and consistent test for heteroscedasticity in a nonparametric regression setup. The test is based on an estimator for the best L_2- approximation of the variance function by a constant.

Since the problem of testing heteroscedasticity is equivalent to that of pseudoresiduals for a constant mean, Dette (2002) constructs a testing procedure which can detect the alternatives converging to the null at a rate of $(n\sqrt{h})^{-1}$ where n is the sample size and h is the bandwidth in the kernel smoothing. Liero (2003) carries out a nonparametric regression model with random design and derives an asymptotic α-test for the hypothesis that the conditional variance of the observations is constant against that depends on the design. This test is based on the L_2 distance between a nonparametric variance in both null and alternative models. The test by Liero (2003) also can detect local alternatives converging to the null hypothesis as the same rate as in the test of Dette (2002). Classical tests, such as the Wald test, the likelihood ratio test, and the score test may be constructed for assessing the variability of the data, but they require the specification of an alternative model and parametric error distribution.

In the multi-dimensional covariate case, a Cramer-von Mises Type test based on cumulative estimated residuals, is proposed by Zhu et al. (2001). These tests are able to detect the local alternatives converging to the null at the parametric rate of $1/\sqrt{n}$, regardless of the type of regression function and the variance function. The asymptotic distributions of the above test statistics are usually complicated and are not asymptotically distribution free. Some bootstrapping methods are used to find the critical values and p-values. A major shortcoming of these test procedures is that they highly depend on the choice of a smoothing parameter, which can affect the results of the statistical analysis.

According to my knowledge, compared to the research of testing heteroscedasticity, fewer meticulous procedures for testing the adequacy of a given variance function are proposed in the literature. Dette et al. (2007) study the problem of testing the parametric form of the conditional variance in nonparametric regression models. They propose a Kolmogorov-Smirnov and a Cramer-von Mises type tests which are constructed from a stochastic process. These stochastic processes are based on the difference between the empirical processes that are obtained from the standardized nonparametric residuals under the null and alternative hypotheses. They discuss the local behavior and the consistency of a bootstrap approximation. The finite sample properties of the approximation are also investigated by means of a simulation study.

In the multi-dimensional covariate case, Wang and Zhou (2006) present a kernel smoothing based nonparametric test for checking the adequacy of parametric variance function of the covariate or regression mean. It does not specify a parametric distribution for the random errors. Under the null hypothesis, it has an asymptotical normal distribution and is powerful against a large class of alternatives. The test can detect $1/\sqrt{nh^{d/2}}$ local alternative, where n is the sample size, d is the dimension of the covariates, and h is the bandwidth in constructing the test statistic.

In this thesis, a testing procedure is proposed to appraise the adequacy of fitting the variance function with a parametric function in the heteroscedastic regression models. The test is based on certain minimized L_2-distance between a nonparametric variance function estimator and a parametric variance function estimator. The asymptotic normality, consistency and local power of the test are discussed. A simulation study is carried out to check the performance of the test. The proposed test statistic is comprised of some integrations and hence it is computationally not easy. A simple but more powerful new test is proposed using the idea of Zheng (1996).

1.3 The Objective and the Overview of the Thesis

The objective of this thesis is to develop a new testing procedure to assess the adequacy of fitting the variance function with a parametric form in the heteroscedastic regression models. The parametric method is preferred for ease in interpretation, compared with non-parametric or semi-parametric methods, even though these methods are flexible in modeling. The proposed inference procedures are motivated from the minimum distance idea of Wolfowitz (1957) in order to provide strongly consistent estimates and decision rules. Minimum distance method in statistics is a statistical method which can be used to check the goodness of fit statistics in a mathematical model to data. Chi-square test, Cramer-von Mises type test, Kolmogorov-Smirnov test and Anderson-Darling test are some examples of statistical tests that have been used for minimum distance estimation. It is shown that the minimum distance estimates have the invariant property of maximum likelihood estimates (Drossos and Philippou (1980)).

The goal of this study is to propose new testing procedures to assess the adequacy of fitting parametric variance function in heteroscedastic regression models. In contrast to classical methods based on residuals, the proposed tests in Chapter 2 are based on certain minimized L_2-distance between a nonparametric and a parametric variance function estimator. The asymptotic distribution of the test statistic corresponding to the minimum distance estimators under the fixed model is shown to be normal. Also these estimators are $\sqrt{n}-$consistent. The asymptotic power of the proposed test against some local nonparametric alternatives is also investigated. Numerical simulation studies are conducted to evaluate the finite sample performance of the test in one dimensional and two dimensional cases.

The minimum distance method proposed in Chapter 2 requires the calculation of the integrals in the test statistics. These integrals do not have a tractable form. Therefore, some numerical integration methods are needed to approximate the integrations. Chapter 3 of the thesis discusses a nonparametric empirical smoothing lack-of-fit test for the functional form of the variance in regression models that do not involve evaluation of integrals. The proposed test can be treated as a nontrivial modification of Zheng (1996)'s nonparametric smoothing test and Koul and Ni (2004)'s minimum distance test for the mean function in the classic regression models. It establishes the asymptotic normality of the proposed test under the null hypothesis. Consistency at some fixed alternatives and asymptotic power under some local alternatives are also discussed. Simulation studies are conducted to assess the finite sample performance of the proposed empirical smoothing test. The simulation studies show that this test is more powerful and computationally more efficient than some existing tests.

The rest of the thesis is organized as follows: Chapter 2 of the thesis provides the description, the asymptotic normality , consistency, and local power of the proposed minimum distance test against local mis-specifications, a simulation study using the bootstrap method, and some simulation results. The proof of the main results are also included in Chapter 2. Chapter 3 includes the discussion of the nonparametric empirical smoothing lack-of-fit test for the functional form of the variance in regression models. A summary of the thesis and a proposed future works are given in Chapter 4.

Chapter 2

Conditional Variance Function Checking in Heteroscedastic Regression Models

This chapter discusses the test, the test statistic, assumptions, and main results associated with the minimum distance conditional variance function checking in heteroscedastic regression model. Consistency and local power of the minimum distance test are also discussed. The simulation procedure and the simulation results are presented next while the proof of some theorems are given at the end of the chapter.

2.1 Parametric Regression Models

Parametric regression is a form of regression analysis in which the predictors take predetermined form and is constructed according to the information derived from the data. Let Y be a one-dimensional response variable, X be a d-dimensional explanatory variable, β be a p-dimensional unknown parameter vector, and ϵ be a random error. Then, usually a parametric regression model can be written as,

$$Y = m(X; \beta) + \epsilon \tag{2.1}$$

where the function $m(X; \beta) = E(Y|X)$ is the unknown regression function. Including a variance function in the regression model is imperative as it can describe the variability of the model and hence produce efficient parameter estimates.

2.2 Heteroscedastic Regression Models

The goal of this report is to consider the heteroscedastic parametric regression models which means the variance of the disturbances are not constant across the data. Consider the regression model,

$$Y = m(X; \beta) + \sqrt{v(X)}\epsilon \tag{2.2}$$

where v is the conditional variance of Y given X and

$$E(\epsilon|X) = 0, \qquad E(\epsilon^2|X) = 1 \tag{2.3}$$

From the assumptions in (2.3), we have

$$E[(Y - m(X; \beta))^2|X = x] = v(x). \tag{2.4}$$

Let $v(X; \beta, \theta)$ be a given parametric variance function, where $(\beta, \theta) \in \Gamma \times \Theta$, and Θ is a compact subset of $\mathbb{R}^q$.

Then the hypothesis of interest, can be written as,

$$\begin{aligned} H_0 : v(X) &= v(X;\beta_0,\theta_0) \text{ for some } (\beta_0,\theta_0) \in \Gamma \times \Theta \qquad (2.5) \\ H_1 : v(X) &\neq v(X;\beta,\theta) \text{ for all } (\beta,\theta) \in \Gamma \times \Theta. \end{aligned}$$

The above hypothesis tests whether the variance function $v(x)$ can be modeled parametrically. Additionally, assuming that the given variance function holds, we are interested in finding the parameters, (β_0, θ_0) in the given family that best fits the data.

In real applications, β is usually unknown, but a natural way to proceed is to replace β with an estimator $\hat{\beta}$. There are many estimating procedures which can provide an estimator of β, say $\hat{\beta}_n$, such that $\sqrt{n}(\hat{\beta}_n - \beta_0) \to N(0, \Sigma_{\beta_0,\theta_0})$ in distribution, where $\Sigma_{\beta_0,\theta_0}$ is a $p \times p$ positive definite matrix defined on the true parameters β_0 and θ_0. In the case of known β, the hypothesis test (2.5) is equivalent to the testing of the regression function in the model

$$(Y - m(X;\beta))^2 = v(X) + \xi. \qquad (2.6)$$

In the equation (2.6), $(Y - m(X;\beta))^2$ can be viewed as the new response variable and $\xi = (Y - m(X;\beta))^2 - E[(Y - m(X;\beta))^2|X]$, which is the error term, is uncorrelated with X (i.e. $E(\xi|X) = 0$).

2.2.1 Minimum Distance Method: Test Statistic

The proposed inference procedures are motivated from the minimum distance method which is developed by Wolfowitz (1953), Wolfowitz (1954), and Wolfowitz (1957) for estimating parameters or function of distributions. The test statistic of (2.7) is constructed in a similar way that Koul and Ni (2004) use in minimum distance model checking procedure. It is based on the L_2-distance between a nonparametric variance function estimator and a parametric variance function estimator. The test statistic is of the form,

$$T_n^\star(\beta,\theta) = \int_C \left[\frac{\sum_{i=1}^n K_h(x-X_i)(Y_i - m(X_i;\beta))^2}{\sum_{i=1}^n K_w(x-X_i)} - v(x;\beta,\theta) \right]^2 dG(x), \tag{2.7}$$

where C is a compact set in $\mathbb{R}^d$, G is a weighting measure with c which is a compact subset of its support, K is a kernel function $K_h(.) = h^{-d}K(./h)$, and h is the bandwidth. Note that the first term of the square, inside the integrand is a Nadaraya-Watson kernel regression estimator.

The corresponding minimum distance estimator is

$$\theta_n^\star = \underset{\theta\in\Theta}{\text{argmin}} T_n^\star(\hat{\beta},\theta) \tag{2.8}$$

Since the integrand inside the square of $T_n^\star$ is not centered, and because of the non negligible asymptotic bias in the nonparametric estimator, which is the first term inside the integrand, the statistic $T_n^\star(\hat{\beta}_n,\theta_n^\star)$ does not have desirable asymptotic properties under the null hypothesis. In addition, the estimator, $\theta_n^\star$ is consistent but not normally distributed. To overcome this difficulty, another form of L_2-distance, $T_n(\hat{\beta}_n,\hat{\theta}_n)$, is used to construct the test statistic, where

$$T_n(\beta,\theta) = \int_C \left[\frac{\sum_{i=1}^n K_h(x-X_i)[(Y_i - m(X_i;\beta))^2 - v(X_i;\beta,\theta)]}{\sum_{i=1}^n K_w(x-X_i)} \right]^2 dG(x) \tag{2.9}$$

and the corresponding estimator of θ is

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\text{argmin}} T_n(\hat{\beta}_n,\theta). \tag{2.10}$$

Under the null hypothesis H_0, the i^{th} summand inside the squared integral of $T_n(\beta, \theta)$ is now conditionally centered, given the i^{th} explanatory variable, $1 \leq i \leq n$. But the asymptotic bias in $n^{1/2}(\hat{\theta}_n - \theta_0)$ and $T_n(\beta, \hat{\theta}_n)$ caused by the nonparametric estimator $\hat{f}_h$ of f, in the denominator of $T_n(\beta, \theta)$. According to Koul and Ni (2004), these asymptotic biases can be made negligible if we use an optimal window width (w) for the estimation of the density f different from h and possibly a different kernel to estimate f.

2.3 Required Assumptions

Here we shall state the following assumptions for the results and proof in our procedures. Let $\dot{m}(x;\beta)$ as the derivative of m with respect to β, $\dot{v}_\beta(x;\beta,\theta)$ as the derivative of v with respect to β, and $\dot{v}_\theta(x;\beta,\theta)$ as the derivative of v with respect to θ.

(e_1). The random variables $\{(X_i, Y_i) : X_i \in \mathbb{R}^d, Y_i \in \mathbb{R}, i = 1, 2, \cdots, n\}$ are i.i.d. with respect to regression function $E(Y|X = x) = m(x;\beta)$ and $E((Y - m(x;\beta))^2|X = x) = v(x)$ satisfying $\int v^2(x)dG(x) < \infty$, where G is a σ-finite measure on $\mathbb{R}^d$.

(e_2). $E\{((Y - m(X;\beta))^2 - v(X))^2\} < \infty$, and
the function $\tau(x) = E\{((Y - m(X;\beta))^2 - v(X))^2|X = x\}$ is a.s. (G) continuous on C.

(e_3). $E\{(Y - m(X;\beta))^2 - v(X)\}^{2+\delta} < \infty$ for some $\delta > 0$.

(e_4). $E\{(Y - m(X;\beta))^2 - v(X)\}^4 < \infty$.

(f_1). X has a uniformly continuous density f, that is bounded from below on C.

(f_2). The density function f, is twice continuously differentiable with a compact support.

(g). G has a continuous density function g.

(k). The kernel function K, is positive symmetric square integrable densities on $[-1,1]^d$. In addition, it satisfies the Lipschitz condition.

(h_1). $h, w \to 0, nh^{2d}, nw^{2d} \to \infty$ as $n \to \infty$.

(h_2). $w \sim n^{-a}$, where $a < min(1/2d, 4/d(d+4))$.

(m_1). For any fixed $x, m(x,\beta)$ is differentiable with respect to β and its derivative is square integrable, that is $E\|\dot{m}(X;\beta)\|^2 < \infty$ and $\int_C \|\dot{m}(x;\beta)\|^4 dG(x) < \infty$.

(m_2). For any $\sqrt{n}$ consistent estimator of β_0,

$$\sqrt{n} \sup_{1 \le i \le n} |m(X_i;\hat{\beta}_n) - m(X_i;\beta_0) - (\hat{\beta}_n - \beta_0)^{'} \dot{m}(X_i;\beta_0)| = o_p(1).$$

(v_1). For all β and θ, $v(x;\beta,\theta)$, $\dot{v}_\beta(x;\beta,\theta)$, and $\dot{v}_\theta(x;\beta,\theta)$ are a.s. continuous in x with respect to integrating measure G.

(v_2). The parametric family of variance function $v(x;\beta_0,\theta)$ is identifiable with respect to θ, that is if $v(x;\beta_0,\theta_1) = v(x;\beta_0,\theta_2)$, for almost all $x(G)$, then $\theta_1 = \theta_2$.

(v_3). $v(x;\beta,\theta)$ is Lipschitz continuous with respect to β and θ. That is, for some positive continuous function l on C, and for any $\alpha > 0$,

$$|v(x;\beta_1,\theta_1) - v(x;\beta_2,\theta_2)| \le l(x)[\|\beta_1 - \beta_2\|^\alpha + \|\theta_1 - \theta_2\|^\alpha]$$

holds for all $\beta_1, \beta_2, \theta_1$, and θ_2.

(v_4).

$$\lim_{n\to\infty} \sup P\left(\sup \frac{|v(X_i;\beta,\theta) - v(X_i;\beta,\theta_0) - \dot{v}_\theta^{'}(X_i;\beta,\theta_0)(\theta-\theta_0)|}{\|\theta-\theta_0\|} \ge \epsilon \right) = 0$$

where the supremum is taking over the set $\{1 \le i \le n; \beta \in \Gamma; \sqrt{nh^d}\|\theta - \theta_0\| \le k\}$ for any $k > 0$, and

$$\lim_{n\to\infty} \sup P\left(\sup \frac{|v(X_i;\beta,\theta) - v(X_i;\beta_0,\theta) - \dot{v}_\beta^{'}(X_i;\beta_0,\theta)(\beta-\beta_0)|}{\|\beta-\beta_0\|} \ge \epsilon\right) = 0$$

where the supremum is taking over the set $\{1 \le i \le n; \theta \in \Theta; \sqrt{nh^d}\|\beta - \beta_0\| \le k\}$ for any $k > 0$.

(v_5).

$$\lim_{n\to\infty} \sup P\left(\sup h^{-d/2}\|\dot{v}_\theta(X_i;\beta,\theta) - \dot{v}_\theta(X_i;\beta,\theta_0)\| \ge \epsilon\right) = 0$$

where the supremum is taking over the set $\{1 \le i \le n; \beta \in \Gamma; \sqrt{nh^d}\|\theta - \theta_0\| \le k\}$ for any $k > 0$, and

$$\lim_{n\to\infty} \sup P\left(\sup h^{-d/2}\|\dot{v}_\beta(X_i;\beta,\theta) - \dot{v}_\beta(X_i;\beta_0,\theta)\| \ge \epsilon\right) = 0$$

where the supremum is taking over the set $\{1 \le i \le n; \theta \in \Theta; \sqrt{nh^d}\|\beta - \beta_0\| \le k\}$ for any $k > 0$.

(v_6). For any β, there exist a function $k(x)$, such that $\int_C k^2(x)dG(x) < \infty$, and

$$\sup_{\theta\in\Theta}|v(x;\beta,\theta) - v(x;\beta_0,\theta)| + \sup_{\theta\in\Theta}|\dot{v}_\theta(x;\beta,\theta) - \dot{v}_\theta(x;\beta_0,\theta)| < k(x)\|\beta - \beta_0\|.$$

(v_7). $\sup_{\theta\in\Theta}\int_C v^2(x;\beta_0,\theta)dG(x) < \infty, \int_C \|\dot{v}_\theta(x;\beta_0,\theta_0)\|^2 dG(x) < \infty.$

Under the conditions, $(f_1), (k), (h_1), (h_2)$, it is well-known that the followings: (See Mack and Siverman (1982)),

$$\sup_{x\in C}\left|\hat{f}_h(x) - f(x)\right| = o_p(1), \qquad \sup_{x\in c}\left|\hat{f}_w(x) - f(x)\right| = o_p(1), \tag{2.11}$$

$$\sup_{x \in C} \left| \frac{f(x)}{\hat{f}_w(x)} - 1 \right| = o_p(1).$$

2.4 Main Results

Theorem 2.4.1. *Assume that the conditions* $(e_1), (e_2), (f_1), (h_1), (h_2), (k), (m_1), (m_2)$, *and* $(v_1) - (v_3)$ *hold, then under* H_0 *in (2.5),* $\theta_n^\star \to \theta_0, \hat{\theta}_n \to \theta_0$ *in probability.*

To show the asymptotic normality of the minimum distance estimator $\hat{\theta}_n$, we shall assume that $\hat{\beta}_n$ has the following approximate linear expression,

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} L(Y_i, X_i; \beta_0, \theta_0) + o_p(1) \tag{2.12}$$

with $EL(Y, X; \beta_0, \theta_0) = 0, \Sigma_L = EL(Y, X; \beta_0, \theta_0)L^{'}(Y, X; \beta_0, \theta_0) > 0$ and

$$E\|L(Y, X; \beta_0, \theta_0)\|^{2+\delta} < \infty. \tag{2.13}$$

In literature, there are some standard estimation procedures to find $\hat{\beta}_n$, namely, the least squares, weighted least squares, quasi-likelihood procedures, etc. Hence the above is a convenience assumption. We make the following additional assumption on L.

(l). $\rho(x) = E[(\epsilon^2 - 1)L(Y, X; \beta_0, \theta_0)|X = x]$ is a.s. continuous in x with respect to integrating measure G.

We define the following terms for easy use of the future procedures.

$$\begin{aligned}
\Pi &= \int_C v_\theta(x;\beta_0,\theta_0)v_\beta^{'}(x;\beta_0,\theta_0)dG(x), && (2.14)\\
\Sigma_0 &= \int_C \dot{v}_\theta(x;\beta_0,\theta_0)\dot{v}_\theta^{'}(x;\beta_0,\theta_0)dG(x), && (2.15)\\
\Omega &= \int \frac{\tau(x)v^2(x;\beta_0,\theta_0)\dot{v}_\theta(x;\beta_0,\theta_0)\dot{v}_\theta^{'}(x;\beta_0,\theta_0)g^2(x)}{f^2(x)}dx, && (2.16)\\
M &= \int \rho(x)v(x;\beta_0,\theta_0)\dot{v}_\theta^{'}(x;\beta_0,\theta_0)g(x)dx. && (2.17)
\end{aligned}$$

Theorem 2.4.2. *Assume that the conditions* $(e_1)-(e_3), (f_1), (f_2), (g), (h_2), (l), (m_1), (m_2)$, *and* $(v_1)-(v_5)$ *hold. Then under* H_0 *in (2.5),*

$$\sqrt{n}(\hat{\theta}_n-\theta_0) \overset{d}{\Rightarrow} N(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1}) \tag{2.18}$$

where Σ_0 *as in 2.15, and* $\Sigma = \Omega + \Pi\Sigma_L\Pi + \Pi M + M^{'}\Pi$.

If $\rho(x) = 0$ in (l), then $M = 0$, and the asymptotic variance of $\hat{\theta}_n$ is simply $\Omega + \Pi\Sigma_L\Pi$. Again, we define the following terms which will use for the asymptotic normality of the minimum distance statistic $T_n(\hat{\beta},\hat{\theta})$,

$$\begin{aligned}
C_n(\beta,\theta) &= \frac{1}{n^2}\sum_{i=1}^n \int_C K_h^2(x-X_i)[(Y_i-m(X_i;\beta))^2 - v(X_i;\beta,\theta)]^2 d\hat{\psi}_w(x) && (2.19)\\
\Gamma_n(\beta,\theta) &= \frac{2h^d}{n^2}\sum_{i\neq j}\left(\int_C K_h(x-X_i)K_h(x-X_j)\xi_i(\beta;\theta)\xi_j(\beta;\theta)d\hat{\psi}_w(x)\right)^2 && (2.20)\\
\Gamma &= 2\int_C \frac{\tau^2(x)g^2(x)}{f^2(x)}dx.\int\left(\int K(u)K(u+v)du\right)^2 dv, && (2.21)
\end{aligned}$$

where

$$\xi_i(\beta;\theta) = (Y_i - m(X_i;\beta))^2 - v(X_i;\beta,\theta), \tag{2.22}$$

and $d\hat{\psi}_w(x) = dG(x)/\hat{f}_w^2(x)$.

Theorem 2.4.3. *Assume that the conditions* $(e_1), (e_2), (e_4), (f_1), (f_2), (g), (h_2), (l), (m_1), (m_2)$, *and* $(v_1) - (v_5)$ *hold. Then under* H_0 *in (2.5),*

$$nh^{d/2}\Gamma_n^{-1/2}(\hat{\beta}, \hat{\theta})(T_n(\hat{\beta}, \hat{\theta}) - C_n(\hat{\beta}, \hat{\theta})) \overset{d}{\Rightarrow} N(0, 1). \tag{2.23}$$

Thus, the test that rejects H_0 whenever $|nh^{d/2}\Gamma_n^{-1/2}(\hat{\beta}, \hat{\theta})(T_n(\hat{\beta}, \hat{\theta}) - C_n(\hat{\beta}, \hat{\theta}))| \geq z_{\alpha/2}$ is of the asymptotic size α, where $z_{\alpha/2}$ is the $100(1-\alpha)th$ percentile of the standard normal distribution.

2.5 Consistency and Local Power of the Minimum Distance Test

In this section, we shall show that, under some regularity conditions, the minimum distance test is consistent for certain fixed alternatives, and has non-trivial asymptotic power against a large class of $1/\sqrt{nh^{d/2}}$ local nonparametric alternatives.

2.5.1 Consistency

Suppose $v_1(x)$ to be a known positive and real-valued function such that $v_1(x) \notin \{v(x; \beta, \theta) : \beta \in \Gamma, \theta \in \Theta\}$. Consider the alternative hypothesis $H_a : v(x) = v_1(x)$, for all $x \in \mathbb{R}^d$. Suppose the true value of β under H_a is still β_0, the estimator $\hat{\beta}_n$ is usually not a consistent estimator for β_0. But under some regularity conditions, Jennrich (1969)'s Theorem 6 implies for any n, there exists a least square estimator which is consistent of some other value, say β_a, and also it is asymptotically normal. The minimum distance estimator $\hat{\theta}_n$ has the same property. So without loss of generality, we assume now the estimators $\hat{\beta}_n$ and the minimum distance estimator $\hat{\theta}_n$ in (2.10) satisfy

$$
\begin{aligned}
\sqrt{n}(\hat{\beta}_n - \beta_a) &= O_p(1) \quad \text{and} &(2.24)\\
\sqrt{n}(\hat{\theta}_n - \theta_a) &= O_p(1) &(2.25)
\end{aligned}
$$

for some $\beta_a \in \Gamma, \theta_a \in \Theta$.

The following notations are used to express the consistency of the minimum distance test procedure. Let $m_0(x) = m(x;\beta_0), m_a(x) = m(x;\beta_a),$ and $v_a(x) = v(x;\beta_a,\theta_a)$ and then

$$
\Delta = \int_C [(m_0(x) - m_a(x))^2 + (v_1(x) - v_a(x))]^2 dG(x)
$$

Theorem 2.5.1. *Suppose the conditions for Theorem 2.4.3 hold with β_0, θ_0 being replaced by β_a, θ_a. Then under H_a, if (2.24), and (2.25) hold with $\Delta > 0$, for $0 < \alpha < 1$, the test that rejects H_0 whenever $|nh^{d/2}\Gamma_n^{-1/2}(\hat{\beta},\hat{\theta})(T_n(\hat{\beta},\hat{\theta}) - C_n(\hat{\beta},\hat{\theta}))| \geq z_{\alpha/2}$ is consistent for H_a.*

2.5.2 Local Power

Let $\delta(x)$ be a positive real valued function such that $\int_C \delta^2(x)dG(x) < \infty$. Here we shall study the asymptotic power of the proposed minimum distance test against the local alternatives

$$
H_{LOC} : v(x) = v(x;\beta_0,\theta_0) + c_n\delta(x), \quad \text{for all} x \in \mathbb{R}^d. \tag{2.26}
$$

Under H_{LOC}, the regression model is of the form

$$
Y = m(X;\beta_0) + \sqrt{v(X;\beta_0,\theta_0) + c_n\delta(X)}\epsilon.
$$

We shall assume that the estimators $\hat{\beta}, \hat{\theta}$ used in the test statistic have the same asymptotic properties as in the null case. Then we have the following theorem.

Theorem 2.5.2. *Suppose the conditions in Theorem 2.4.3 hold and* $c_n = 1/\sqrt{nh^{d/2}}$. *Then under the local alternative* H_{LOC},

$$nh^{d/2}\Gamma_n^{-1/2}(\hat{\beta},\hat{\theta})(T_n(\hat{\beta},\hat{\theta}) - C_n(\hat{\beta},\hat{\theta})) \overset{d}{\Rightarrow} N\left(\Gamma^{-1/2}\int_C \delta^2(x)dG(x), 1\right).$$

2.6 Simulation Study

We have conducted out a simulation study to investigate the performance of the test proposed in finite sample situations. There are several purposes for conducting the simulation study. Namely, investigating the validity of the test procedure, checking the influence of the bandwidth choice, and error distribution on the validity and power of the test. The test statistic has a relatively complicated form, which makes the implementation of the test procedure difficult. In particular, the integration usually has no tractable expressions. So that a Reimman-sum approximation is necessary. But the test statistic can be simplified by choosing proper weighting measure G, and using an approximately equivalent expression for $\hat{\Gamma}_n$. For example, choose $dG(x) = g(x)dx = \hat{f}_w^2(x)dx$, then $T_n(\hat{\beta}_n,\hat{\theta}_n)$ and $C_n(\hat{\beta}_n,\hat{\theta}_n)$ in (2.9) and (2.19) respectively, can be simplified as

$$\begin{aligned}
T_n(\hat{\beta}_n,\hat{\theta}_n) &= \int_C \left[\frac{1}{n}\sum_{i=1}^n K_h(x - X_i)\xi_i(\hat{\beta}_n,\hat{\theta}_n)\right]^2 dx, \\
C_n(\hat{\beta}_n,\hat{\theta}_n) &= \frac{1}{n^2}\sum_{i=1}^n \int_C K_h^2(x - X_i)\xi_i^2(\hat{\beta}_n,\hat{\theta}_n)\, dx,
\end{aligned}$$

where $\xi_i(\beta;\theta) = (Y_i - m(X_i;\beta))^2 - v(X_i;\beta,\theta)$. With the definition of $\tau^2(x)$ in (e_2), and $g(x) = \hat{f}_w(x)^2$, a simpler consistent estimator of Γ in (2.21) can be written as

$$\hat{\Gamma}_n = 2\int_C \left[\sum_{i=1}^{n} K_h(x - X_i)\xi_i^2(\hat{\beta}_n, \hat{\theta}_n)\right]^2 dx. \int \left(\int K(u)K(u+v)du\right)^2 dv.$$

2.6.1 A Bootstrap Algorithm

It is observed by the authors Hardle and Mamman (1993) that in testing parametric assumptions regarding the regression function, the asymptotic calculation of the level by approximations (similar as in Theorems 2.4.2, 2.4.3 and 2.5.1) is inappropriate for realistic sample sizes. It is well known that the bootstrap procedure usually provides better performance for small to moderate sample sizes in the nonparametric smoothing tests.

To investigate the finite sample performance of the minimum distance test procedure, we generate the samples from the following models:

$$\begin{aligned}
Model\,0 : Y_i &= \beta_1 + \beta_2 X_i + \sqrt{\theta_1 + \theta_2 X_i}\epsilon_i, \\
Model\,1 : Y_i &= \beta_1 + \beta_2 X_i + \sqrt{\theta_1 + \theta_2 X_i + 0.5X_i^2}\epsilon_i, \\
Model\,2 : Y_i &= \beta_1 + \beta_2 X_i + \sqrt{\theta_1 + \theta_2 X_i + 0.8X_i^2}\epsilon_i, \\
Model\,3 : Y_i &= \beta_1 + \beta_2 X_i + \sqrt{\theta_1 + \theta_2 X_i + X_i^2}\epsilon_i,
\end{aligned}$$

for $i = 1, 2, \cdots, n$.

The data from model 0 are used to study the empirical level, while data from models $1 - 3$ are used to study the empirical power of the test.

In the simulation, we generate $X_i \sim N(0, 1)$, for $i = 1, 2, \cdots, n$, with $\beta_1 = 1, \beta_2 = 2, \theta_1 = 2$ and $\theta_2 = 0.1$. Two types of error distributions considered are:

(1). $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ are independently drawn from the standard normal distribution $N(0, 1)$;

(2). $\epsilon_1, \epsilon_2, \cdots, \epsilon_n$ are independently drawn from the uniform distribution $U(-\sqrt{3}, \sqrt{3})$.

The normality of the minimum distance test statistic that is proved here allows one to use bootstrap methodology. The following is a simple bootstrap algorithm to implement the minimum distance test procedure which consists of six steps.

Step 1. For a given random sample of observations, obtain a $\sqrt{n}$- consistent estimator $\hat{\beta}_n$ of β under the null hypothesis. Such estimator can be found by using least squares procedures, weighted least squares, pseudo-likelihood procedures, etc..

Step 2. Obtain the minimum distance estimator $\hat{\theta}_n$ of θ by minimizing $T_n(\hat{\beta}_n, \theta)$ under the null hypothesis.

Step 3. Define $\hat{\epsilon}_i = [Y_i - m(X_i; \hat{\beta}_n)]/\sqrt{v(X_i; \hat{\beta}_n, \hat{\theta}_n)}$ for $i = 1, 2, \cdots, n$.. Center and standardize $\hat{\epsilon}_1, \hat{\epsilon}_2, \cdots, \hat{\epsilon}_n$ such that they have means of zero and variances of one.

Step 4. Obtain a bootstrap sample from the standardized residuals in step 3; denote them as $\hat{\epsilon}_i^\star$ for $i = 1, 2, \cdots, n$ and define $Y_i^\star = m(X_i; \hat{\beta}_n) + \sqrt{v(X_i; \hat{\beta}_n, \hat{\theta}_n)}\hat{\epsilon}_i^\star$ for $i = 1, 2, \cdots, n$.

Step 5. For the bootstrap sample $(X_i, Y_i^\star)$ for $i = 1, 2, \cdots, n$, calculate the estimator $\hat{\beta}_n^\star$ as in step 1 and the minimum distance estimator $\hat{\theta}_n^\star$ as in Step 2 under the null hypothesis. Let $\xi_i^\star(\hat{\beta}_n^\star, \hat{\theta}_n^\star) = (Y_i^\star - m(X_i; \hat{\beta}_n^\star))^2 - v(X_i; \hat{\beta}_n^\star, \hat{\theta}_n^\star)$. Then the bootstrap version of the test statistic is

$$T_n^\star(\hat{\beta}_n^\star, \hat{\theta}_n^\star) = \int_C \left[\frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \xi_i^\star(\hat{\beta}_n^\star, \hat{\theta}_n^\star) \right]^2 dx.$$

Step 6. Repeat Steps 4 and 5 a sufficiently large number of times. For a specified significance level of the test, the critical value is then determined as the appropriate quantile of the bootstrap distribution of the test statistic.

The kernel function K is chosen to be the *Epanechnikov kernel* function which is of the form,

$$K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1),$$

which is used throughout the simulation. The integration $\int[\int K(u)K(u+v)du]^2 dv = 0.4338$. The bandwidth h is chosen to be $an^{-1/3}$, where a is some positive constant, and the sample sizes are taken to be $n = 100, 200, 300, 400, 500, 800$, and 1000. The compact set C is chosen to be $[-3, 3]$ and the integration is approximated by a Riemman sum with $[-3, 3]$ being equally divided into 300 subintervals. The test is calculated with 500 simulation runs with the nominal level $\alpha = 0.05$. Thus, the simulated level has a Monte Carlo Error of $\sqrt{0.05 \times 0.95/500} \approx 1\%$. We use 400 samples per run to obtain the critical value $c^\star_\alpha$. The empirical size and power are computed by using the relative frequency of the event $\#\{T_n(\hat{\beta}_n, \hat{\theta}_n) \geq c^\star_\alpha\}/500$. The simulation is done using the R statistical software.

2.6.2 Simulation Results

For $a = 1$, Table 2.1 and Table 2.2 report the minimum distance estimator of θ_1 and θ_2 for different sample sizes.

Table 2.1: *Mean and MSE of* $\hat{\theta}_1$

	100	200	300	400	500	800	1000
Mean	1.9747	1.9759	1.9778	1.9929	1.9981	1.9932	1.9914
MSE	0.0970	0.0443	0.0316	0.0259	0.0181	0.0123	0.0091

The mean of the minimum distance estimator of θ_1 is around the true value of 2, and the mean square error of it decreases when the sample size gets bigger. The situation is the same for the estimator of θ_2 in the following Table 2.2.

Table 2.2: *Mean and MSE of* $\hat{\theta_2}$

	100	200	300	400	500	800	1000
Mean	0.1002	0.0941	0.1055	0.1037	0.0968	0.0968	0.0986
MSE	0.1151	0.0567	0.0397	0.0291	0.0263	0.0136	0.0116

Table 2.3: *Empirical size and power for* $h = n^{-1/3}$

		100	200	300	400	500	800	1000
$\epsilon \sim N(0,1)$	Model 0	0.044	0.038	0.042	0.040	0.044	0.045	0.046
	Model 1	0.130	0.170	0.252	0.312	0.358	0.526	0.610
	Model 2	0.210	0.330	0.452	0.556	0.600	0.846	0.912
	Model 3	0.226	0.386	0.552	0.636	0.764	0.934	0.936
$\epsilon \sim U(-\sqrt{3}, \sqrt{3})$	Model 0	0.042	0.034	0.048	0.052	0.072	0.042	0.052
	Model 1	0.210	0.384	0.534	0.648	0.780	0.948	0.990
	Model 2	0.364	0.644	0.834	0.918	0.954	1.000	1.000
	Model 3	0.476	0.732	0.874	0.954	0.980	1.000	1.000

Table 2.3 shows the empirical size and power, which are the frequencies of rejecting the corresponding null hypothesis under the significance level $\alpha = 0.05$ of the minimum distance test for two different error distributions with $a = 1$.

We see from the tables with $a = 1$, that under H_0 (that is, model 0 in the variance function) the empirical levels are slightly less than the nominal level $\alpha = 0.05$, regardless of the selection of bandwidth for all the chosen sample sizes. Thus the proposed test is conservative for all chosen sample sizes, which is clear from the power curve in Figure 2.1. The empirical powers against all alternative models get larger when the sample sizes get larger. For fixed sample size, the alternative model 1 has smaller powers. Then the power becomes bigger when the alternative model is further apart from the null model (as the coefficient of x^2 changes from 0.5 to 1). In considering the influence of the error distribution on the performance of the test, it is clear that the empirical levels with the uniformly distributed

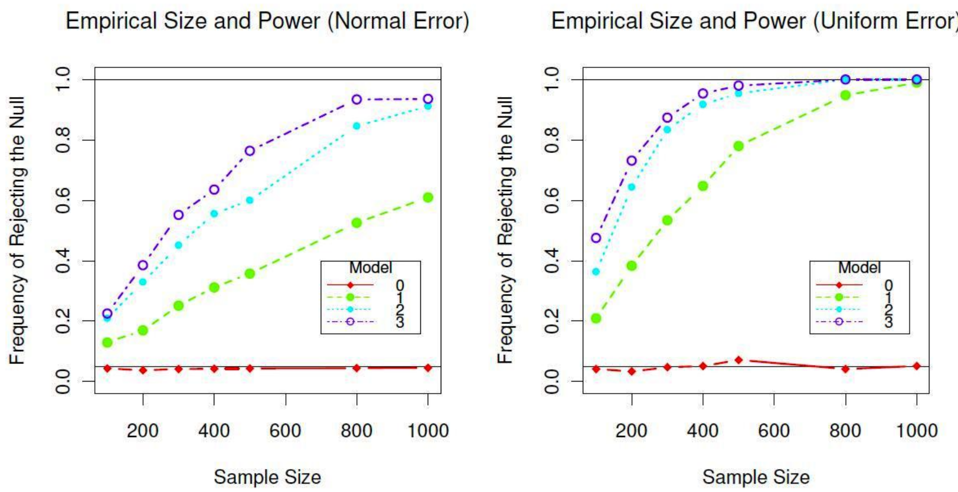

Figure 2.1: *Empirical Size and Power Curves ($h = n^{-1/3}$)*

error are slightly closer to the nominal level than that with the normally distributed error. This indicates that the different distributions of the error terms have some effect on both the accuracy and the power of the test.

To see the effect of the bandwidth on the performance of the minimum distance test, we also conduct a simulation study for $a = 0.5$ and the simulation results are shown in Table 2.4.

Table 2.4: *Empirical size and power for* $h = 0.5n^{-1/3}$

		100	200	300	400	500	800	1000
$\epsilon \sim N(0,1)$	Model 0	0.044	0.044	0.046	0.036	0.048	0.042	0.044
	Model 1	0.124	0.194	0.256	0.316	0.414	0.568	0.650
	Model 2	0.236	0.334	0.466	0.608	0.658	0.872	0.930
	Model 3	0.298	0.416	0.616	0.712	0.806	0.950	0.958
$\epsilon \sim U(-\sqrt{3}, \sqrt{3})$	Model 0	0.040	0.036	0.052	0.054	0.076	0.038	0.050
	Model 1	0.188	0.354	0.516	0.602	0.734	0.924	0.976
	Model 2	0.332	0.622	0.798	0.890	0.940	0.992	1.000
	Model 3	0.450	0.688	0.862	0.944	0.966	0.998	1.000

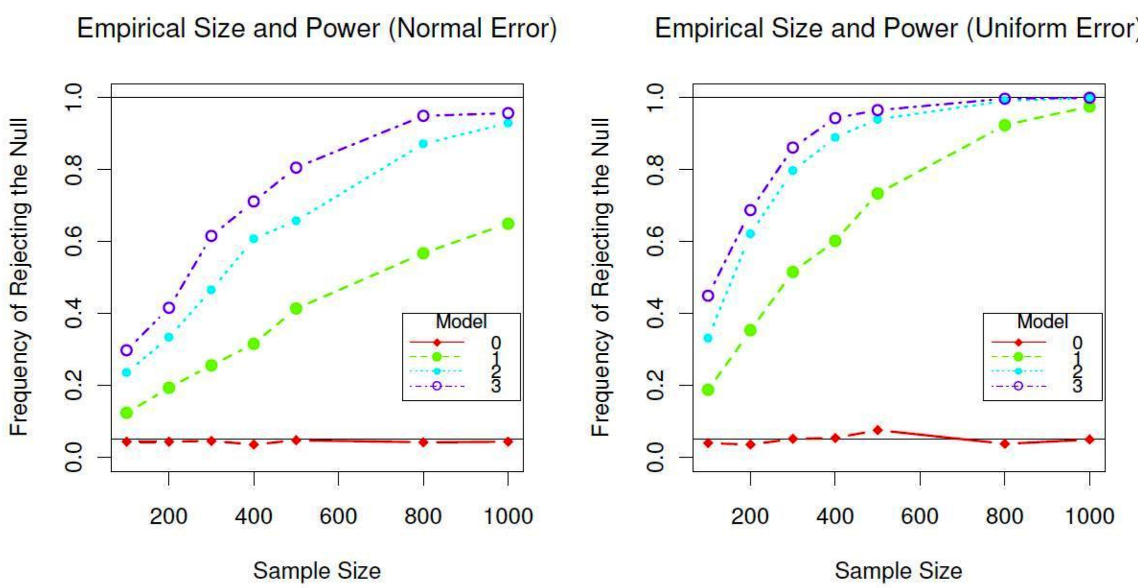

Figure 2.2: *Empirical Size and Power Curves ($h = 0.5n^{-1/3}$)*

Compared to the case of $a = 1$, the simulation results for $a = 0.5$ only vary slightly (see Figure 2.1 and Figure 2.2). The slight difference between these mentioned simulations does indicate that the bandwidth may have certain influence on the test when sample sizes are small to moderate. Therefore, in the real world problem, it is better to perform the test with several values of bandwidth to make a decision to reject or not to reject the null hypothesis.

2.6.3 Simulation Study: Two Dimensional Case

To investigate the performance of the test more deeply, we conduct a simulation study when the design variable has two dimensions. The data are generated from the models $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \sqrt{\theta_0 + \theta_1 X_1 + \theta_2 X_2 + b(X_1^2 + X_2^2)}\epsilon$. The sample from the model with $b = 0$ are used to study the empirical level, while data from models with $b = 0.5, 0.8, 1$ are used to study the empirical power of the test. In the simulation, $X_1 \sim N(0,1), X_2 \sim N(0,1), \beta_0 = 1, \beta_1 = 2, \beta_2 = 1, \theta_0 = 2$ and $\theta_1 = \theta_2 = 0.1$. We study the effect of two error distributions such as $\epsilon \sim N(0,1), \epsilon \sim U(-\sqrt{3}, \sqrt{3})$. The Kernel function K is chosen to be the product of Epanechnikov kernel, i.e. $K(u,v) = 9(1-u^2)(1-v^2)I(|u| \leq 1)I(|v| \leq 1)/16$.

The bandwidth h is chosen to be $n^{-1/5}$ in the two dimensional case (d=2) as the upper bound on the exponent a in n^{-a} is $\min\{1/2d, 4/d(d+4)\}$ in (h_2). The sample sizes used are 100, 200, 300, 400, and 500, and the nominal level used is $\alpha = 0.05$. The weighting measure is chosen to be $dG(x) = \hat{f}_w^2 dF_n(x)$ to make the computation easier., where $F_n(x)$ is the empirical CDF of (X_1, X_2). Similar to the one dimensional case, the test is calculated with 500 simulation runs while the critical value c_α^* is calculated using 400 bootstrap samples per run. The empirical size and power are computed by using $\#\{T_n(\hat{\beta}_n, \hat{\theta}_n) \geq c_\alpha^\star\}/500$.

Table 2.5 gives the empirical sizes and powers regarding the test and it reveals that the proposed test is quite conservative for small to moderate sample sizes. In general, the power would become smaller with the higher dimensional data. In our simulation, we can also see this difference in Table 2.3 and Table 2.5. In the consideration of the influence of the error distribution on the performance of the test, we see that under model 0, the rejection frequencies with normally distributed errors are less than those with the uniformly distributed errors. But under other models (Model 1, Model 2, and Model 3), there is a considerable improvement in the power with uniformly distributed error compared to that with normally distributed errors. This reveals that the different error distributions have some effect on both the accuracy and the power of the test.

Table 2.5: *Empirical size and power for $h = n^{-1/5}$: Two Dimensional Case.*

		100	200	300	400	500	800	1000
$\epsilon \sim N(0,1)$	Model 0	0.020	0.022	0.018	0.038	0.028	0.037	0.029
	Model 1	0.102	0.124	0.102	0.114	0.154	0.212	0.196
	Model 2	0.176	0.198	0.200	0.236	0.258	0.364	0.422
	Model 3	0.260	0.196	0.290	0.350	0.384	0.474	0.586
$\epsilon \sim U(-\sqrt{3}, \sqrt{3})$	Model 0	0.022	0.042	0.036	0.042	0.020	0.046	0.068
	Model 1	0.120	0.160	0.194	0.294	0.356	0.544	0.626
	Model 2	0.176	0.264	0.432	0.538	0.636	0.730	0.862
	Model 3	0.226	0.386	0.532	0.654	0.714	0.850	0.888

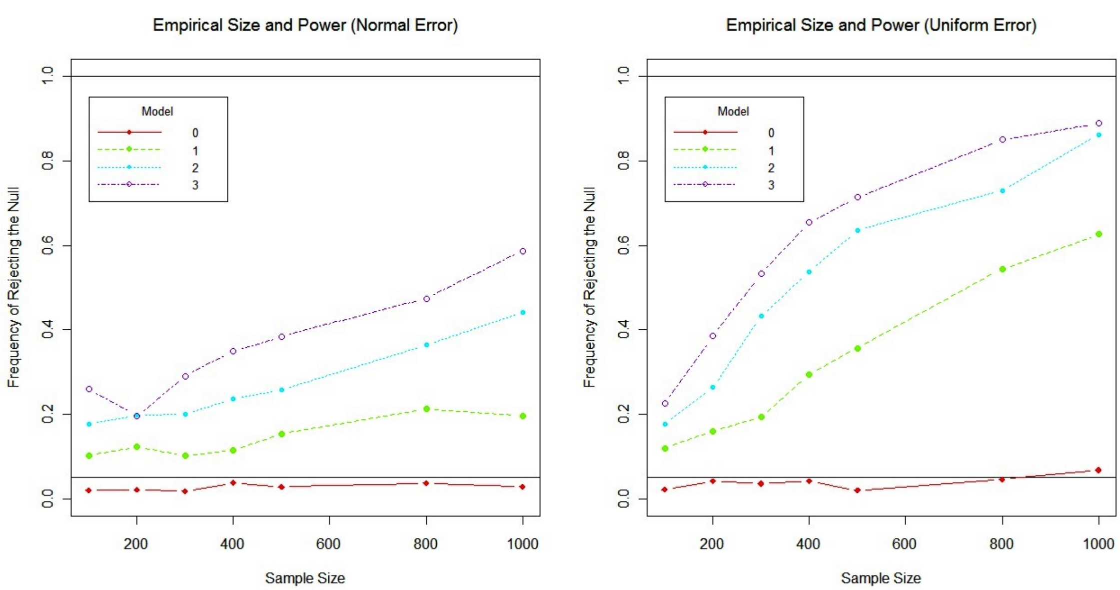

Figure 2.3: *Empirical Size and Power Curves ($h = n^{-1/5}$): Two Dimensional Case*

2.7 Proofs of the Main Results (Minimum Distance Test)

This section is devoted to providing necessary tools for proving the results in chapter 2. We use $\tilde{C}_n(\beta, \theta)$ in (2.19)to denote $C_n(\beta, \theta)$ when $d\hat{\psi}_w(x)$ is replaced by $d\psi(x) = dG(x)/f^2(x)$ and same understanding for $\tilde{\Gamma}_n(\beta, \theta)$ in (2.20). For the sake of convenience, we also define the following:

$$\mu_n(x;\beta) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)(Y_i - m(X_i;\beta))^2, \tag{2.27}$$

$$\eta_n(x;\beta,\theta) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)v(X_i;\beta,\theta), \tag{2.28}$$

$$\dot{\eta}_n(x;\beta,\theta) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i)\dot{v}_\theta(X_i;\beta,\theta). \tag{2.29}$$

The following are the required lemmas to prove the Theorem 2.4.1.

Lemma 2.7.1. *Assume that the conditions* $(e_1), (e_2), (f_1), (h_1), (h_2), (m_1), (m_2)$ *and* $(v_1) - (v_3)$ *hold, then under* H_0*,*

$$(a) \;:\; \tilde{\theta}_n = \underset{\theta\in\Theta}{argmin} T_n^\star(\beta_0, \theta) \textit{ is a consistent estimator of} \theta_0;$$

$$(b) \;:\; sup_{\theta\in\Theta}|T_n^\star(\hat{\beta}_n, \theta) - T_n^\star(\beta_0, \theta)| = o_p(1),$$

where $T_n^\star$ *is as defined in (2.7).*

The proof of (a) is similar to that of corollary 3.1 in Koul and Ni (2004), hence the proof is omitted here.

Proof of part (b):

Let

$$A_{n1} = \int_C \left[\frac{\mu_n(x;\hat\beta_n) - \mu_n(x;\beta_0)}{\hat f_w(x)}\right]^2 dG(x),\ \ A_{n2}(\theta) = \int_C \left[v(x;\hat\beta_n,\theta) - v(x;\beta_0,\theta)\right]^2 dG(x). \tag{2.30}$$

Then the test statistic,

$$T_n^\star(\hat\beta,\hat\theta) = \int_C \left[\frac{\sum_{i=1}^n K_h(x-X_i)(Y_i - m(X_i;\hat\beta))^2}{\sum_{i=1}^n K_w(x-X_i)} - v(x;\hat\beta,\hat\theta)\right]^2 dG(x)$$

can be written as the sum of $T_n^\star(\beta_0,\theta), A_{n1}, A_{n2}(\theta)$ and three other terms which are bounded above by $2\sqrt{A_{n1}A_{n2}(\theta)}, 2\sqrt{A_{n1}T_n^\star(\beta_0,\theta)}$ and $2\sqrt{A_{n2}(\theta)T_n^\star(\beta_0,\theta)}$,using the Cauchy-Schwartz inequality. Therefore it is enough to show that $A_{n1} = o_p(1)$, $\sup_{\theta\in\Theta}|A_{n2}(\theta)| = o_p(1)$ and $\sup_{\theta\in\Theta}|T_n^\star(\beta_0,\theta)| = O_p(1)$. Adding and subtracting $m(x;\beta_0)$ from $Y_i - m(X_i;\hat\beta_n)$, the term A_{n1} is bounded above by A_{n11} and A_{n12}, where

$$\begin{aligned} A_{n11} &= 2\int_C \left[\frac{\sum_{i=1}^n K_h(x-X_i)[m(X_i;\hat\beta_n) - m(X_i;\beta_0)]^2}{\sum_{i=1}^n K_w(x-X_i)}\right]^2 dG(x), \\ A_{n12} &= 8\int_C \left[\frac{\sum_{i=1}^n K_h(x-X_i)(Y_i - m(X_i;\beta_0))(m(X_i;\hat\beta_n) - m(X_i;\beta_0))}{\sum_{i=1}^n K_w(x-X_i)}\right]^2 dG(x). \end{aligned}$$

Let

$$e_{ni} = m(X_i;\hat\beta_n) - m(X_i;\beta_0) - (\hat\beta_n - \beta_0)'\dot m(X_i;\beta_0). \tag{2.31}$$

Then

$$
\begin{aligned}
A_{n11} &= \int_C \left[\frac{\sum_{i=1}^n K_h(x-X_i)(e_{ni} + (\hat{\beta}_n - \beta_0)'\dot{m}(X_i;\beta_0))^2}{\sum_{i=1}^n K_w(x-X_i)} \right]^2 dG(x) \\
&\leq 8 \sup_{1\leq i\leq n} |e_{ni}|^4 \int_C [\hat{f}_h(x)/\hat{f}_w(x)]^2 dG(x) \\
&+8\|\hat{\beta}_n - \beta_0\|^4 \int_C n^{-1}\sum_{i=1}^n K_h(x-X_i)\|\dot{m}(X_i;\beta_0)\|^2/\hat{f}_w(x)]^2 dG(x) \\
&= o_p(n^{-2})O_p(1) + O_p(n^{-2})O_p(1) \\
&= o_p(1)
\end{aligned}
$$

from the conditions $(f_1), (m_1), (m_2), (k), (h_1), (h_2)$, the $\sqrt{n}$-consistency of $\hat{\beta}_n$, and the fact (2.11).

Similarly, we can show that

$$
\begin{aligned}
A_{n12} &= 8\int_C \left[\frac{\sum_{i=1}^n K_h(x-X_i)(Y_i - m(X_i;\beta_0))((e_{ni} + (\hat{\beta}_n - \beta_0)'\dot{m}(X_i;\beta_0)))}{\sum_{i=1}^n K_w(x-X_i)} \right]^2 dG(x) \\
&\leq 16 \sup_{1\leq i\leq n} |e_{ni}|^2 \int_C [n^{-1}\sum_{i=1}^n K_h(x-X_i)|\epsilon_i|\sqrt{v(X_i;\beta_0,\theta_0)}/\hat{f}_w(x)]^2 dG(x) \\
&+ 16\|\hat{\beta}_n - \beta_0\|^2 \int_C [n^{-1}\sum_{i=1}^n K_h(x-X_i)\|\dot{m}(X_i;\beta_0)\|/\hat{f}_w(x)]^2 dG(x) \\
&= o_p(n^{-1})O_p(1) + O_p(n^{-1})O_p(1) \\
&= o_p(1)
\end{aligned}
$$

from the conditions $(m_1), (m_2), (k), (h_1), (h_2)$, the $\sqrt{n}$-consistency of $\hat{\beta}_n$, and the fact 2.11. $\mathrm{Sup}_{\theta\in\Theta} A_{n2}(\theta) = o_p(1)$ can be obtained by using the Lipschitz condition in $(v3)$ and the $\sqrt{n}$ consistency of $\hat{\beta}_n$. The last requirement, $\mathrm{Sup}_{\theta\in\Theta} T_n^\star(\beta_0,\theta) = O_p(1)$ can be shown using (v_1) and

$$
\int_C \left[\frac{\sum_{i=1}^n K_h(x-X_i)(Y_i - m(X_i;\beta_0))^2}{\sum_{i=1}^n K_w(x-X_i)} \right]^2 dG(x) = O_p(1). \tag{2.32}
$$

Hence the proof of part (b) is completed.

$\blacksquare$

To state the second lemma, let $L_2(G)$ denote a class of square integrable real valued functions of $\mathbb{R}^d$ with respect to G. Define

$$\rho(v_1, v_2) = \int_C [v_1(x) - v_2(x)]^2 dG(x), \text{ where } v_1, v_2 \in L_2(G)$$

and the map

$$\mathcal{M}(u) = \underset{\theta\in\Theta}{\text{argmin}}\ \rho(u, v(x; \beta_0, \theta)), u \in L_2(G).$$

Lemma 2.7.2. *Let v satisfy conditions $(v_1) - (v_3)$. Then the following results hold.*

(a). $\mathcal{M}(u)$ *always exists* , $\forall u \in L_2(G)$,

(b). *If* $\mathcal{M}(u)$ *is unique, then* $\mathcal{M}$ *is continuous at* u *in the sense that for any sequence of* $u_n \in L_2(G)$ *converging to* $u \in L_2(G), \mathcal{M}(u_n) \to \mathcal{M}, i.e., \rho(u_n, u) \to 0$ *implies* $\mathcal{M}(u_n) \to \mathcal{M}(u)$, *as* $n \to \infty$.

(c). $\mathcal{M}(v(x; \beta_0, \theta)) = \theta$ *uniquely for* $\forall\theta \in \Theta$.

This lemma is related to Minimum Hellinger Distance Functionals and the proof is omitted as it is similar to Theorem 1 of Beran (1977).

Proof of Theorem 2.4.1 We shall use the part (b) in Lemma 2.7.2 with $u_n(x) = v(x; \beta_0, \theta_n^\star)$ and $u(x) = v(x; \beta_0, \theta_0)$. Note that $\theta_n^\star = \mathcal{M}(u_n), \theta_0 = \mathcal{M}(u)$, uniquely by (v_2). So it suffices to show that

$$\rho(u_n, u) = \int_C [v(x; \beta_0, \theta_n^\star) - v(x; \beta_0, \theta_0)]^2 dG(x) = o_p(1). \tag{2.33}$$

By adding and subtracting $\mu_n(x; \beta_0)/\hat{f}_w(x)$ in the parenthesis of the above integral, $\rho(u_n, u)$, expanding the quadratic, and using the Cauchy- Schwartz inequality on the cross product and can show that it is bounded above by the sum

$$2\int_C [\mu_n(x;\beta_0)/\hat{f}_w(x) - v(x;\beta_0,\theta_n^\star)]^2 dG(x) + 2\int_C [\mu_n(x;\beta_0)/\hat{f}_w(x) - v(x;\beta_0,\theta_0)]^2 dG(x).$$

Using a similar argument, like proving of $C_{n2}(\theta_0) = o_p(1)$ in Koul and Ni (2004), the second term is the order of $o_p(1)$, while the first term is bounded above by the sum of

$$\begin{aligned}
B_{n1} &= 6\int_C [v(x;\beta_0,\theta_n^\star) - v(x;\hat{\beta}_n,\theta_n^\star)]^2 dG(x), \\
B_{n2} &= 6\int_C [\mu_n(x;\hat{\beta}_n)/\hat{f}_w(x) - v(x;\hat{\beta}_n,\theta_n^\star)]^2 dG(x), \\
B_{n3} &= 6\int_C [\mu_n(x;\hat{\beta}_n)/\hat{f}_w(x) - \mu_n(x;\beta_0)/\hat{f}_w(x)]^2 dG(x).
\end{aligned}$$

Lipschitz condition (v_3) and the $\sqrt{n}$-consistency of $\hat{\beta}_n$ imply that $B_{n1} = o_p(1)$. To show that $B_{n2} = o_p(1)$, note that from part (b) of Lemma 2.7.1, $\sup_{\theta\in\Theta}|T_n^\star(\hat{\beta}_n,\theta) - T_n^\star(\beta_0,\theta)| = o_p(1)$, therefore,

$$T_n^\star(\hat{\beta}_n,\theta_n^\star) - T_n^\star(\beta_0,\theta_n^\star) = o_p(1), \quad T_n^\star(\hat{\beta}_n,\tilde{\theta}_n) - T_n^\star(\beta_0,\tilde{\theta}_n) = o_p(1), \tag{2.34}$$

where $\tilde{\theta}_n$ is defined in part (a) of Lemma 2.7.1. Hence

$$T_n^\star(\hat{\beta}_n,\theta_n^\star) - T_n^\star(\hat{\beta}_n,\tilde{\theta}_n) = T_n^\star(\beta_0,\theta_n^\star) - T_n^\star(\beta_0,\tilde{\theta}_n) + o_p(1). \tag{2.35}$$

By the definition of $\theta_n^\star$ and $\tilde{\theta}_n$, the left hand side of (2.35) is non-positive, and the difference $T_n^\star(\beta_0,\theta_n^\star) - T_n^\star(\beta_0,\tilde{\theta}_n)$ on the right hand side is non-negative. Hence, $T_n^\star(\beta_0,\theta_n^\star) - T_n^\star(\beta_0,\tilde{\theta}_n) = o_p(1)$. Notice that since $T_n^\star(\beta_0,\tilde{\theta}_n) \le T_n^\star(\beta_0,\theta_0) = o_p(1)$,then we have $T_n^\star(\beta_0,\tilde{\theta}_n) = o_p(1)$, but this implies $T_n^\star(\hat{\beta}_n,\theta_n^\star) = o_p(1)$ or $B_{n2} = o_p(1)$. Finally, notice that $B_{n3} = A_{n1}$, where A_{n1} is defined in (2.30), and from the proof of 2.7.1, we have $A_{n1} = o_p(1)$, and so is B_{n3}.

Therefore, (2.33) is proved and hence $\theta_n^\star$ is a consistent estimator of θ_0.

Now let's show the consistency of $\hat{\theta}_n$. Again we will use part (b) of 2.7.2 but with $u_n(x) = v(x;\beta_0,\hat{\theta}_n)$ and $u(x) = v(x;\beta_0,\theta_0)$. Note that $\hat{\theta}_n = \mathcal{M}(u_n), \theta_0 = \mathcal{M}(u)$, uniquely by (v_2). It thus suffices to show that

$$\rho(u_n,u) = \int_C [v(x;\beta_0,\hat{\theta}_n) - v(x;\beta_0,\theta_0)]^2 dG(x) = o_p(1). \tag{2.36}$$

Adding and subtracting $v(x;\hat{\beta}_n,\hat{\theta}_n), \mu_n(x;\hat{\beta}_n)/\hat{f}_w(x), \mu_n(x;\beta_0)/\hat{f}_w(x)$ in the brackets of the above integral, $\rho(u_n,u)$ is bounded above by the sum of the following four terms of

$$\begin{aligned}
C_{n1} &= 4\int_C [v(x;\beta_0,\hat{\theta}_n) - v(x;\hat{\beta}_n,\hat{\theta}_n)]^2 dG(x),\\
C_{n2} &= 4\int_C [\mu_n(x;\hat{\beta}_n)/\hat{f}_w(x) - v(x;\hat{\beta}_n,\hat{\theta}_n)]^2 dG(x),\\
C_{n3} &= 4\int_C [\mu_n(x;\hat{\beta}_n)/\hat{f}_w(x) - \mu_n(x;\beta_0)/\hat{f}_w(x)]^2 dG(x),\\
C_{n4} &= 4\int_C [\mu_n(x;\beta_0)/\hat{f}_w(x) - v(x;\beta_0,\theta_0)]^2 dG(x).
\end{aligned}$$

Lipschitz condition (v_3) and the $\sqrt{n}$-consistency of $\hat{\beta}_n$ imply that $C_{n1} = o_p(1)$. Note that the integral in C_{n3} is simply A_{n1} defined in (2.30),so we have $C_{n3} = o_p(1)$. Also it is obvious that $C_{n4} = o_p(1)$. In the following, we shall show that C_{n2} is the order of $o_p(1)$. It is implied by the following claim

$$\sup_{\theta\in\Theta}|T_n(\hat{\beta}_n,\theta) - T_n^\star(\hat{\beta}_n,\theta)| = o_p(1). \tag{2.37}$$

To show this, by adding and subtracting $\eta_n(x;\hat{\beta}_n,\theta)/\hat{f}_w(x)$ in the parenthesis of the integrand in $T_n^\star(\hat{\beta}_n,\theta)$, we can show that $|T_n(\hat{\beta}_n,\theta) - T_n^\star(\hat{\beta}_n,\theta)| \le D_n(\theta) + 2D_n^{1/2}(\theta)T_n^{1/2}(\hat{\beta}_n,\theta)$, where $D_n(\theta) = \int[\eta_n(x;\hat{\beta}_n,\theta)/\hat{f}_w(x) - v(x;\hat{\beta}_n,\theta)]^2 dG(x)$. Therefore it suffices to show that

$$\sup_{\theta\in\Theta} D_n(\theta) = o_p(1), \quad \sup_{\theta\in\Theta} T_n(\hat{\beta}_n,\theta) = o_p(1).$$

For this purpose, adding and subtracting $\eta_n(x;\beta_0,\theta)/\hat{f}_w(x), v(x;\beta_0,\theta)$in the brackets of the integrand of $D_n(\theta)$, we can show that $D_n(\theta)$ is bounded above by $3D_{n1}(\theta) + 3D_{n2}(\theta) + 3D_{n3}(\theta)$, where

$$
\begin{aligned}
D_{n1}(\theta) &= \int_C \{[\eta_n(x;\hat{\beta}_n,\theta) - \eta_n(x;\beta_0,\theta)]/\hat{f}_w(x)\}^2 dG(x), \\
D_{n2}(\theta) &= \int_C \{\eta_n(x;\beta_0,\theta)/\hat{f}_w(x) - v(x;\beta_0,\theta)\}^2 dG(x), \\
D_{n3}(\theta) &= \int_C \{v(x;\hat{\beta}_n,\theta) - v(x;\beta_0,\theta)\}^2 dG(x).
\end{aligned}
\tag{2.38}
$$

From the condition (v_3), we can show that $D_{n1}(\theta) = \|\hat{\beta}_n - \beta_0\|^{2\alpha}.O_p(1)$, and $D_{n3}(\theta) \leq \|\hat{\beta}_n - \beta_0\|^{2\alpha} \int_C \|l(x)\|^2 dG(x)$. Therefore, by the $\sqrt{n}$ consistency of $\hat{\beta}_n$, both $D_{n1}(\theta)$ and $D_{n3}(\theta)$ are of the order of $o_p(1)$ uniformly for $\theta \in \Theta$. The proof of $\sup_{\theta\in\Theta} D_{n2}(\theta)$ is similar to the proof of $\sup_{\theta\in\Theta} C_{n2}(\theta) = o_p(1)$ in Koul and Ni (2004). This concludes that the proof of $\sup_{\theta\in\Theta} D_n(\theta) = o_p(1)$. To show $\sup_{\theta\in\Theta} T_n(\hat{\beta}_n,\theta) = O_p(1)$, note that $T_n(\hat{\beta}_n,\theta)$ is bounded above by $3A_{n1} + 3T_n(\beta_0,\theta) + 3D_{n1}(\theta)$, where A_{n1} is as defined in (2.30). We have already shown that $A_{n1} = o_p(1)$,and $\sup_{\theta\in\Theta} D_{n1}(\theta) = o_p(1)$, so we only have to show that $\sup_{\theta\in\Theta} T_n(\beta_n,\theta) = O_p(1)$, but this can be done by using similar argument in Koul and Ni (2004). Hence, complete the proof of the theorem.

■

The following lemma is necessary to prove the Theorem 2.4.2 which appears as in Theorem 2.2 part (2) in Bosq (1998) .

Lemma 2.7.3. *Let* $\hat{f}_w(x)$ *be the kernel estimate associate with a kernel density function* K *which satisfies a Lipschitz condition. If* (f_2) *holds and* $w = a_n(\log n/n)^{1/(d+4)}$*, where* $a_n \to a_0 > 0$*, then for any positive integer* k*,*

$$\frac{n^{2/(d+4)}}{(\log n)^{2/(d+4)}\log_k n} \underset{c}{sup} |\hat{f}_w(x) - f(x)| \to 0$$

almost surely.

Proof of Theorem 2.4.2:

The first step is to show that

$$nh^d \|\hat{\theta}_n - \theta_0\|^2 = O_p(1). \tag{2.39}$$

For this purpose, let

$$H_n(\theta) = \int_C \left(\frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) [v(X_i, \hat{\beta}_n, \theta) - v(X_i, \hat{\beta}_n, \theta_0)] \right)^2 d\hat{\psi}_w(x).$$

We claim that $nh^d H_n(\hat{\theta}_n) = O_p(1)$. To see this, note that

$$\begin{aligned} H_n(\hat{\theta}_n) &\leq 2\int_C \left(\frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) [(Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i, \hat{\beta}_n, \hat{\theta}_n)] \right)^2 d\hat{\psi}_w(x) \\ &+ 2\int_C \left(\frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) [(Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i, \hat{\beta}_n, \theta_0)] \right)^2 d\hat{\psi}_w(x) \\ &= 2T_n(\hat{\beta}_n, \hat{\theta}_n) + 2T_n(\hat{\beta}_n, \theta_0) \leq 4T_n(\hat{\beta}_n, \theta_0). \end{aligned}$$

Therefore, it is sufficient to show that

$$nh^d T_n(\hat{\beta}_n, \theta_0) = O_p(1). \tag{2.40}$$

Adding and subtracting $(Y_i - m(X_i;\beta_0))^2, v(x:\beta_0;\theta_0)$ from $(Y_i - m(X_i;\hat{\beta}_n))^2 - v(x,\hat{\beta}_n,\theta_0)$ in $T_n(\hat{\beta}_n,\theta_0)$, we can show that $T_n(\hat{\beta}_n,\theta_0)$ is bounded above by $3A_{n1}+3T_n(\beta_0,\theta_0)+3D_{n1}(\theta_0)$, where A_{n1} is defined in (2.30) and $D_{n1}(\theta_0)$ is given in (2.38). Since $A_{n1} = O_P(1/n)$ from the proof of Lemma 2.7.1, $nh^d A_{n1} = O_p(h^d) = o_p(1)$. Note that $D_{n1}(\theta_0)$ is bounded above by $2D_{n11}(\theta_0) + 2D_{n12}(\theta_0)$, where

$$D_{n11}(\theta_0) = \int_C \left[\frac{n^{-1}\sum_{i=1}^n K_h(x-X_i)[v(X_i,\hat{\beta}_n,\theta_0) - v(X_i,\beta_0,\theta_0) - (\hat{\beta}_n - \beta_0)'\dot{v}_\beta(X_i;\beta_0,\theta_0)]}{\hat{f}_w(x)} \right]^2 dG(x)$$

and

$$D_{n12}(\theta_0) = \int_C \left[\frac{n^{-1}\sum_{i=1}^n K_h(x-X_i)[(\hat{\beta}_n - \beta_0)'\dot{v}_\beta(X_i;\beta_0,\theta_0)]}{\hat{f}_w(x)} \right]^2 dG(x).$$

It is easy to see that $D_{n11}(\theta_0)$ is bounded above by

$$\sup_{1\le i\le n} |v(X_i,\hat{\beta}_n,\theta_0) - v(X_i,\beta_0,\theta_0) - (\hat{\beta}_n - \beta_0)'\dot{v}_\beta(X_i;\beta_0,\theta_0)| \int_C \left[\frac{\hat{f}_h(x)}{\hat{f}_w(x)}\right]^2 dG(x)$$

which has the order of $o_p(n^{-1})$ by (v_4). By Cauchy-Schwartz inequality, $D_{n12}(\theta_0)$ is bounded above by

$$\|\hat{\beta}_n - \beta_0\|^2 \int_C \left[\frac{n^{-1}\sum_{i=1}^n K_h(x-X_i)\|\dot{v}_\beta(X_i;\beta_0,\theta_0)\|}{\hat{f}_w(x)} \right]^2 dG(x)$$

which is $O_p(1/n)$ by (v_1) and the $\sqrt{n}$ consistency of $\hat{\beta}_n$. Therefore $nh^d D_{n1}(\theta_0) = nh^d O_p(1/n) = o_p(1)$. So, we only have to show that $nh^d T_n(\beta_0,\theta_0) = O_p(1)$.

Let $\Delta_n(x) = f^2(x)/\hat{f}_w^2(x) - 1$. Then $nh^d T_n(\beta_0,\theta_0)$ is bounded above by the following two terms.

$$\begin{aligned} Q_{n1} &= nh^d \int_C \left[\frac{\sum_{i=1}^n K_h(x - X_i) v(X_i; \beta_0, \theta_0)(\epsilon_i^2 - 1)}{f(x)} \right]^2 dG(x) \quad \text{and} \\ Q_{n2} &= nh^d \int_C \left[\frac{\sum_{i=1}^n K_h(x - X_i) v(X_i; \beta_0, \theta_0)(\epsilon_i^2 - 1)}{f(x)} \right]^2 \Delta_n(x) dG(x). \end{aligned}$$

Note that $\epsilon_i^2 - 1$ are i.i.d. with mean 0, so

$$\begin{aligned} & E \int_C \left[\frac{1}{nh^d} \sum_{i=1}^n K_h(x - X_i) v(X_i; \beta_0, \theta_0)(\epsilon_i^2 - 1) \right]^2 dG(x) \\ = & \frac{1}{nh^{2d}} \int_C EK_h^2(x - X) v(X; \beta_0, \theta_0) \tau(x) d\psi(x) \end{aligned} \tag{2.41}$$

where $d\psi(x) = dG(x)/f^2(x)$. Then from conditions $(e_2), (f_1)$ and (v_1), we can show that the right hand side of (2.41) is the order of $O_p(1/nh^d)$. Hence $Q_{n1} = O_p(1)$. Realizing that $|Q_{n2}| \le \sup_{x \in c} |\Delta_n(x)|.Q_{n1}$, then from 2.11, we have $Q_{n2} = o_p(1)$. These imply that $nh^d T_n(\beta_0, \theta_0) = O_p(1)$, hence $nh^d H_n(\hat{\theta}_n) = O_p(1)$. Similar to proof of (4.6) in Koul and Ni (2004), we can show that

$$\liminf_{n \to \infty} P \left(H_n(\hat{\theta}_n) / \|\hat{\theta}_n - \theta_0\|^2 \ge \frac{1}{2} \inf_{\|b\|=1} b' \Sigma_0 b \right) = 1, \tag{2.42}$$

where Σ_0 is defined in 2.15. To prove, (2.42), let

$$d_{ni} = v(X_i; \hat{\beta}_n, \hat{\theta}_n) - v(X_i; \hat{\beta}_n, \theta_0) - \dot{v}'_\theta(X_i; \hat{\beta}_n, \theta_0)(\hat{\theta}_n - \theta_0). \tag{2.43}$$

Then $H_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2$ can be written as the sum of $H_{n1} + H_{n2} + H_{n3}$, where

$$\begin{aligned} H_{n1} &= \int_C \left[n^{-1} \sum_{i=1}^n K_h(x - X_i) d_{ni} / \|\hat{\theta}_n - \theta_0\| \right]^2 d\hat{\psi}_w(x), \\ H_{n2} &= \int_C \left[n^{-1} \sum_{i=1}^n K_h(x - X_i) \dot{v}_\theta'(X_i; \hat{\beta}_n, \theta_0)(\hat{\theta}_n - \theta_0)/\|\hat{\theta}_n - \theta_0\| \right]^2 d\hat{\psi}_w(x) \end{aligned}$$

and H_{n3} is a term whose absolute value being bounded above by $D_{n1}^{1/2} D_{n2}^{1/2}$, where D_{n1} and D_{n2} are defined as in (2.38). From the condition (v_4), it is easy to see that $H_n = o_p(1)$. Adding and subtracting, $\dot{v}_\theta(X_i; \beta_0, \theta_0)$ from $\dot{v}_\theta(X_i; \hat{\beta}_n, \theta_0)$ in the integrand of H_{n2}, we can show that

$$H_{n2} = H_{n21} + H_{n22} + H_{n23}, \text{ where}$$

$$\begin{aligned} H_{n21} &= \int_C \left((nh^d)^{-1} \sum_{i=1}^n K_h(x - X_i) \frac{\dot{v}_\theta(X_i; \beta_0, \theta_0)(\hat{\theta}_n - \theta_0)}{\|\hat{\theta}_n - \theta_0\|} \right)^2 d\hat{\psi}_w(x), \\ H_{n22} &= \int_C \left((nh^d)^{-1} \sum_{i=1}^n K_h(x - X_i) \frac{[\dot{v}_\theta(X_i; \hat{\beta}_n, \theta_0) - \dot{v}_\theta(X_i; \beta_0, \theta_0)](\hat{\theta}_n - \theta_0)}{\|\hat{\theta}_n - \theta_0\|} \right)^2 d\hat{\psi}_w(x), \end{aligned}$$

and H_{n23} is bounded above by $H_{n21}^{1/2} . H_{n22}^{1/2}$. From the condition (v_6) and the fact of 2.11, we can show that $H_{n22} = O_p(\|\beta - \beta_0\|) = o_p(1)$, while

$$H_{n21} \geq \inf_{\|b\|=1} \int_C b' \dot{\eta}_n(x; \beta_0, \theta_0) \dot{\eta}_n'(x; \beta_0, \theta_0) b d\hat{\psi}_w(x) \equiv \inf_{\|b\|=1} \Sigma_n(b).$$

By the usual calculation, we can show that for each $b \in \mathbb{R}^q, \Sigma_n(b) \to b'\Sigma b$ in probability. Also note that for any $\delta > 0$ and any $b_1, b_2 \in \mathbb{R}^q$ such that $\|b_1 - b_2\| \leq \delta$, we have

$$|\Sigma_n(b_2) - \Sigma_n(b_1)| < \delta(\delta + 2) \int_C \left[n^{-1} \sum_{i=1}^n K_h(x - X_i) \|\dot{v}_\theta(X_i; \beta_0, \theta_0)\| \right]^2 d\hat{\psi}_w(x).$$

Condition (v_7) and the fact that 2.11, imply the above integration is $O_p(1)$. From these observations and the compactness of the unit circle $\{b \in \mathbb{R}^q : \|b\| = 1\}$, we obtain that

$$\sup_{\|b\|=1} |\Sigma_n(b) - b^{'}\Sigma b| = o_p(1).$$

Notice that we also have $H_{n21} = O_p(1)$, therefore, H_{n23} will be the order of $O_p(1)$. Hence we have proved (2.42). The claim (2.39) will then follow from (2.42), $nh^d H_n(\hat{\theta}_n) = O_P(1), \Sigma_0 > 0$ and the fact

$nh^d H_n(\hat{\theta}_n) = nh^d \|\hat{\theta}_n - \theta_0\|^2 . [H_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2]$.

That is the proof of

$$nh^d \|\hat{\theta}_n - \theta_0\|^2 = O_p(1).$$

Asymptotic normality of $\hat{\theta}_n$

Since θ_0 is an interior point of Θ, by the consistency of $\hat{\theta}_n$ for sufficiently large n, $\hat{\theta}_n$ will be in the interior point of Θ, so $\dot{T}_{n,\theta}(\hat{\beta}_n, \hat{\theta}_n) = 0$, where $\dot{T}_{n,\theta}(\hat{\beta}_n, \hat{\theta}_n)$ is the derivative of $T_n(\hat{\beta}_n, \theta)$ with respect to θ, evaluated at $\theta = \hat{\beta}_n$. This is equivalent to

$$\int_C [\mu_n(x; \hat{\beta}_n) - \eta_n(x; \hat{\beta}_n, \hat{\theta}_n)] \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x) = 0.$$

By adding and subtracting $\eta_n(x; \hat{\beta}_n, \theta_0)$ from $\mu_n(x; \hat{\beta}_n) - \eta_n(x; \hat{\beta}_n, \hat{\theta}_n)$, the above can be written as

$$\begin{aligned} & \int_C [\mu_n(x; \hat{\beta}_n) - \eta_n(x; \hat{\beta}_n, \theta_0)] \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x) \\ = & \int_C [\eta_n(x; \hat{\beta}_n, \hat{\theta}_n) - \eta_n(x; \hat{\beta}_n, \theta_0)] \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x). \end{aligned} \tag{2.44}$$

Denote the left hand side as L_n and the right hand side as R_n, then note that L_n can be written as the sum of $L_{n1} + L_{n2} + L_{n3}$, where

$$
\begin{aligned}
L_{n1} &= \int_C [\mu_n(x; \hat{\beta}_n) - \mu_n(x; \beta_0)] \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x), \\
L_{n2} &= \int_C [\mu_n(x; \beta_0) - \eta_n(x; \beta_0, \theta_0)] \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x), \\
L_{n3} &= \int_C [\eta_n(x; \beta_0, \theta_0) - \eta_n(x; \hat{\beta}_n, \theta_0)] \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x).
\end{aligned}
$$

For L_{n1}, we have

$$
\begin{aligned}
L_{n1} &= 2 \int_C n^{-1} \sum_{i=1}^{n} K_h(x - X_i)(Y_i - m(X_i; \beta_0))(m(X_i; \beta_0) - m(X_i; \hat{\beta}_n)) \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x) \\
&\quad + \int_C n^{-1} \sum_{i=1}^{n} K_h(x - X_i)(m(X_i; \beta_0) - m(X_i; \hat{\beta}_n))^2 \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x) \\
&= L_{n11} + L_{n12}.
\end{aligned}
$$

Recall the notation e_{ni} in (2.31) then we have,

$$
\begin{aligned}
L_{n11} = &-2 \int_C \left(n^{-1} \sum_{i=1}^{n} K_h(x - X_i)(Y_i - m(X_i; \beta_0)) e_{ni} \right) \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x) \\
&-2 \int_C \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) \left(n^{-1} \sum_{i=1}^{n} K_h(x - X_i)(Y_i - m(X_i; \beta_0)) \dot{m}'_\beta m(X_i; \beta_0) \right) d\hat{\psi}_w(x)(\hat{\beta}_n - \beta_0).
\end{aligned}
$$

Notice that

$$
\dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) = \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) - \dot{\eta}_n(x; \hat{\beta}_n, \theta_0) + \dot{\eta}_n(x; \hat{\beta}_n, \theta_0) - \dot{\eta}_n(x; \beta_0, \theta_0) - \dot{\eta}_n(x; \beta_0, \theta_0),
$$

then by the condition (v_5) and the fact of 2.11, we can show that

$$\int_C \|\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n)\|^2 d\hat{\psi}_w(x) = \int_C \|\dot{v}_\theta(x;\beta_0,\theta_0)\|^2 dG(x) + o_p(1) \tag{2.45}$$

$$= O_p(1). \tag{2.46}$$

Therefore, by Cauchy-Schwartz inequality and from the condition (m_2),

$$\begin{aligned} & n\| \int_C \left(n^{-1}\sum_{i=1}^n K_h(x-X_i)(Y_i - m(X_i;\beta_0))e_{ni} \right) \dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n) d\hat{\psi}_w(x)\|^2 \\ \le \ & n\int_C \left(n^{-1}\sum_{i=1}^n K_h(x-X_i)(Y_i - m(X_i;\beta_0))e_{ni} \right)^2 d\hat{\psi}_w(x) \int_C \|\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n)\|^2 d\hat{\psi}_w(x) \\ = \ & \sup_{1\le i\le n} |e_{ni}|^2 O_p(1) \\ = \ & o_p(1). \end{aligned}$$

Similarly, we can show that

$$\sqrt{n}\int_C \dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n) \left(n^{-1}\sum_{i=1}^n K_h(x-X_i)(Y_i - m(X_i;\beta_0))\dot{m}'_\beta m(X_i;\beta_0) \right) d\hat{\psi}_w(x)(\hat{\beta}_n - \beta_0) = o_p(1)$$

by the fact that $\sqrt{n}(\hat{\beta}_n - \beta_0) = O_p(1)$ and the fact of

$$\int_C \|n^{-1}\sum_{i=1}^n K_h(x-X_i)(Y_i - m(X_i;\beta_0))\dot{m}'_\beta m(X_i;\beta_0)\|^2 d\hat{\psi}_w(x) = O_p(1/nh^d)$$

which can be shown by the fact of 2.11 and an expectation and variance argument. Therefore, $\sqrt{n}L_{n11} = o_p(1)$. Using Caushy-Schwartz inequality and the conditions of (m_1) and (m_2) on L_{n12}, we can show that $\sqrt{n}L_{n12} = o_p(1)$. Thus we have proved that

$$\sqrt{n}L_{n1} = o_p(1). \tag{2.47}$$

Now, let's consider L_{n2}. For convenience, denote $U_n(x) = \mu_n(x;\beta_0) - \eta_n(x;\beta_0,\theta_0)$. Adding and subtracting $\dot{v}_\theta(X_i,\beta_0,\theta_0)$ from $\dot{v}_\theta(X_i,\hat{\beta}_n,\hat{\theta}_n)$ in $\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n)$, L_{n2} can be written as

$$
\begin{aligned}
L_{n2} &= \int_C U_n(x).n^{-1}\sum_{i=1}^n K_h(x-X_i)(\dot{v}_\theta(X_i,\hat{\beta}_n,\hat{\theta}_n) - \dot{v}_\theta(X_i,\beta_0,\theta_0))d\hat{\psi}_w(x) \\
&\quad + \int_C U_n(x).n^{-1}\sum_{i=1}^n K_h(x-X_i)\dot{v}_\theta(X_i,\beta_0,\theta_0)d\hat{\psi}_w(x) \\
&= L_{n21} + L_{n22}.
\end{aligned}
$$

In the following, we shall prove that $\sqrt{n}L_{n21} = o_p(1)$. In fact

$$
\begin{aligned}
L_{n21} &= \int_C U_n(x).n^{-1}\sum_{i=1}^n K_h(x-X_i)(\dot{v}_\theta(X_i,\hat{\beta}_n,\hat{\theta}_n) - \dot{v}_\theta(X_i,\beta_0,\theta_0))\left(\frac{f^2(x)}{\hat{f}_w(x)} - 1\right)d\psi(x) \\
&\quad + \int_C U_n(x).n^{-1}\sum_{i=1}^n K_h(x-X_i)(\dot{v}_\theta(X_i,\hat{\beta}_n,\hat{\theta}_n) - \dot{v}_\theta(X_i,\beta_0,\theta_0))d\psi(x) \\
&= L'_{n21} + L''_{n22}.
\end{aligned}
$$

Using the Cauchy-Schwartz inequality, the second term is bounded above by the square root of

$$
\int_C U_n^2(x)d\psi(x).\int_C\left[n^{-1}\sum_{i=1}^n K_h(x-X_i)(\dot{v}_\theta(X_i,\hat{\beta}_n,\hat{\theta}_n) - \dot{v}_\theta(X_i,\beta_0,\theta_0))\right]^2 d\psi(x)
$$

which is again bounded above by

$$
\int_C U_n^2(x)d\psi(x).\sup\|\dot{v}_\theta(X_i,\hat{\beta}_n,\hat{\theta}_n) - \dot{v}_\theta(X_i,\beta_0,\theta_0)\|^2\int_C\left[n^{-1}\sum_{i=1}^n K_h(x-X_i)\right]^2 d\psi(x).
$$

Notice that $\|\dot{v}_\theta(X_i,\hat{\beta}_n,\hat{\theta}_n) - \dot{v}_\theta(X_i,\beta_0,\theta_0)\|$ is bounded above by the sum of $\|\dot{v}_\theta(X_i,\hat{\beta}_n,\hat{\theta}_n) -$

$\dot{v}_\theta(X_i, \beta_0, \hat{\theta}_n)\|$ and $\|\dot{v}_\theta(X_i, \beta_0, \hat{\theta}_n) - \dot{v}_\theta(X_i, \beta_0, \theta_0)\|$. By (v_5), both terms are $o_p(h^{d/2})$.
Since $\int_C U_n^2(x)d\psi(x) = O_p(1/nh^d)$,
$\sqrt{n}L''_{n21} = \sqrt{n}.O_p(1/\sqrt{nh^d}).o_p(h^{d/2}) = o_p(1).\sqrt{n}L'_{n21} = o_p(1)$.
Hence we proved

$$\sqrt{n}L_{n21} = o_p(1). \tag{2.48}$$

With considering L_{n22}, we have

$$\begin{aligned} L_{n2} &= \int_C U_n(x).n^{-1}\sum_{i=1}^n K_h(x - X_i)\dot{v}_\theta(X_i, \beta_0, \theta_0)\left(\frac{f^2(x)}{\hat{f}_w(x)} - 1\right)d\psi(x) \\ &\quad + \int_C U_n(x).n^{-1}\sum_{i=1}^n K_h(x - X_i)\dot{v}_\theta(X_i, \beta_0, \theta_0)d\psi(x) \\ &= L'_{n22} + L''_{n22}. \end{aligned}$$

By the the Cauchy-Schwartz inequality,

$$\|L'_{n22}\|^2 \le \int_C U_n^2(x)d\psi(x).\int_C\left[n^{-1}\sum_{i=1}^n K_h(x - X_i)\dot{v}_\theta(X_i, \beta_0, \theta_0)\right]^2 d\psi(x).\sup_{x\in c}\left|\frac{f^2(x)}{\hat{f}_w(x)} - 1\right|^2,$$

so, using Lemma 2.7.3

$$n\|L'_{n22}\|^2 = n.O_p(1/nh^d).o((\log_k n)^2(\log n/n)^{4/(d+4)}) = o_p((\log_k n)^2(\log n)^{4/(d+4)}n^{ad-4/(d+4)})$$

which is $o_p(1)$. Therefore, $\sqrt{n}L'_{n22} = o_p(1)$. This together with the result in (2.48), implies

$$\begin{aligned} \sqrt{n}L_{n2} &= \sqrt{n}\int_C U_n(x).n^{-1}\sum_{i=1}^n K_h(x - X_i)\dot{v}_\theta(X_i, \beta_0, \theta_0)d\psi(x) + o_p(1) \\ &= \sqrt{n}\int_C U_n(x).\dot{\eta}_h(x)d\psi(x) + o_p(1), \end{aligned} \tag{2.49}$$

where

$$\dot{\eta}_h(x) = E[K_h(x - X)\dot{v}_\theta(X_i, \beta_0, \theta_0)]. \tag{2.50}$$

Finally, let us consider L_{n3}. Adding and subtracting $v(x, \beta_0, \theta_0)$ from $v(x, \hat{\beta}_n, \theta_0)$ and taking

$$r_{ni} = v(X_i; \hat{\beta}_n, \theta_0) - v(X_i; \beta_0, \theta_0) - (\hat{\beta}_n - \beta_0)'\dot{v}_\beta(X_i; \beta_0, \theta_0), \tag{2.51}$$

we have

$$\begin{aligned}\sqrt{n}L_{n3} &= -\sqrt{n}\int_C n^{-1}\sum_{i=1}^n K_h(x - X_i)r_{ni}\dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n)d\hat{\psi}_w(x) \\ &\quad -\sqrt{n}\int_C \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n)\dot{\eta}_{n\beta}^{'}(x; \beta_0, \theta_0)d\hat{\psi}_w(x)(\hat{\beta}_n - \beta_0). \end{aligned} \tag{2.52}$$

Condition (v_4) and some routine argument can show that the first term on the right hand side of (2.52) is the order of $o_p(1)$, and the second term is equal to

$$\int_C \dot{\eta}_n(x; \beta_0, \theta_0)\dot{\eta}_{n\beta}^{'}(x; \beta_0, \theta_0)d\psi(x)\sqrt{n}(\hat{\beta}_n - \beta_0) + o_p(1).$$

Note that

$$\int_C \dot{\eta}_n(x; \beta_0, \theta_0)\dot{\eta}_{n\beta}^{'}(x; \beta_0, \theta_0)d\psi(x) = \Pi + o_p(1),$$

where Π is defined in (2.14). then we have

$$\sqrt{n}L_3 = \Pi\sqrt{n}(\hat{\beta}_n - \beta_0) + o_p(1). \tag{2.53}$$

Combining (2.47), (2.49),and (2.53), we have

$$\sqrt{n}L_n = \sqrt{n}\int_C U_n(x)\dot{\eta}_h(x)d\psi(x) - \Pi\sqrt{n}(\hat{\beta}_n - \beta_0) + o_p(1). \tag{2.54}$$

Take $s_{ni} = (\epsilon_i^2 - 1)v(X_i, \beta_0, \theta_0)\int_C K_h(x - X_i)\dot{\eta}_h(x)d\psi(x), t_{ni} = \Pi L(Y_i, X_i; \beta_0, \theta_0)$, where L is defined in 2.12. Then

$$\sqrt{n}L_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(s_{ni} - t_{ni}). \tag{2.55}$$

For convenience, we shall give the proof only for $p = q = 1$. For the multidimensional case, the results can be proved using the Wald Scheme and applying the same argument. Note that $\{s_{ni} - t_{ni}; i \leq 1 \leq n\}$ are i.i.d. centered random variables for each n.

By the Lindeberge-Feller Central Limit Theorem, it suffices to show that as $n \to \infty$,

$$E[(s_{n1} - r_{n1})^2] \to \Sigma, \tag{2.56}$$

$$E[(s_{n1} - r_{n1})^2 I[|s_{n1} - r_{n1}|] > \lambda\sqrt{n}] \to 0 \ \text{ for all } \lambda > 0, \tag{2.57}$$

where Σ is defined in Theorem 2.4.2. Since

$$\begin{aligned} E[(s_{n1} - r_{n1})^2] &= E\left[\tau(x)v^2(X;\beta_0,\theta_0)\left(\int K_h(x - X)\dot{\eta}_h(x)d\psi(x)\right)^2\right] + \Pi^2 E[L^2(Y, X;\beta_0,\theta_0)] \\ &\quad +2\Pi E\left[\rho(X)v(X;\beta_0,\theta_0)\int K_h(x - X)\dot{\eta}_h(x)d\psi(x)\right] \\ &= \sigma_{11} + \Pi^2 E[L^2(Y, X;\beta_0,\theta_0)] + 2\Pi\sigma_{12}. \end{aligned}$$

By Fubini theorem,

$$\sigma_{11} = \int\int EK_h(x - X)K_h(y - X)\tau(x)v^2(X;\beta_0,\theta_0)\dot{\eta}_h(x)\dot{\eta}_h(y)d\psi(x)d\psi(y).$$

By the transformations of $x - z = uh, y - z = vh$, and using the assumed continuity of

$\tau(x), v(X;\beta_0,\theta_0), f$ and g, we, obtain

$$\begin{aligned}\sigma_{11} &= \int\int\int K(u)K(v)\tau(z)v^2(z;\beta_0,\theta_0)\dot{\eta}_h(z+uh)\dot{\eta}_h(z+vh)f(z)\frac{g(z+uh)g(z+vh)}{f^2(z+uh)f^2(z+vh)}dudvdz \\ &\to \int \frac{\tau(x)v^2(x;\beta_0,\theta_0)\dot{v}_\theta^2(x;\beta_0,\theta_0)g^2(x)}{f(x)}dx\end{aligned}$$

as $h \to 0$. By the transformation $x - z = uh$, we can obtain

$$\sigma_{12} = \int \rho(x)v(x;\beta_0,\theta_0)\dot{v}_\theta(x;\beta_0,\theta_0)g(x)dx.$$

Therefore, Σ has the form in Theorem 2.4.2.

To show (2.57), we use the inequality,

$$(a+b)^r \le 2^{r-1}(a^r + b^r) \text{ for } a, b > 0, r > 1.$$

Then the left side of (2.57) is bounded above by

$$\lambda^{-\delta}n^{-\delta/2}E[(s_{n1} - t_{n1})^{2+\delta}] \le 2^{1+\delta}\lambda^{-\delta}n^{-\delta/2}E(s_{n1})^{2+\delta} + 2^{1+\delta}\lambda^{-\delta}n^{-\delta/2}E(t_{n1})^{2+\delta}.$$

Using the Hölder's inequality, and the continuity of $\tau(x), v(x;\beta_0,\theta_0)$, and $\dot{v}_\theta(x;\beta_0,\theta_0)$ with respect to x,

$$E(s_{n1})^{2+\delta} \le E\left[\left(\int_C (K_h(x-X)\dot{\eta}_h(x))^{(2+\delta)/2}d\psi(x)\right)^2 (\tau(x)v(x;\beta_0,\theta_0))^{2+\delta}\right] = O(h^{-\delta d/2}).$$

Therefore, $2^{1+\delta}\lambda^{-\delta}n^{-\delta/2}E(s_{n1})^{2+\delta} = O(nh^{-\delta d/2}) = o_p(1)$. From 2.13, we can see that $2^{1+\delta}\lambda^{-\delta}n^{-\delta/2}E(t_{n1})^{2+\delta} = O(n^{-\delta/2}) = o(1)$. It is the proof of (2.57). Hence

$$\sqrt{n}L_n \Rightarrow N(0,\Sigma) \text{ in distribution} \tag{2.58}$$

Now let us consider the term R_n. In the following, we shall show that $R_n = H_n(\hat{\theta}_n - \theta_0)$ with $H_n = \Sigma_0 + o_p(1)$ where Σ_0 is defined in 2.15. To see this, define

$$d_{ni} = v(X_i; \hat{\beta}_n, \hat{\theta}_n) - v(X_i; \hat{\beta}_n, \theta_0) - \dot{v}_\theta'(X_i; \hat{\beta}_n, \theta_0)(\hat{\theta}_n - \theta_0). \tag{2.59}$$

Then R_n can be written as the sum of $R_{n1} + R_{n2}$, where,

$$\begin{aligned} R_{n1} &= \int_C n^{-1} \sum_{i=1}^n K_h(x - X_i) d_{ni}.\dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x), \\ R_{n2} &= \int_C \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n)\dot{\eta}_n'(x; \hat{\beta}_n, \theta_0) d\hat{\psi}_w(x)(\hat{\theta}_n - \theta_0). \end{aligned}$$

Let

$$R_{n11} = \int_C n^{-1} \sum_{i=1}^n K_h(x - X_i) \frac{d_{ni}}{\|\hat{\theta}_n - \theta_0\|} \dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n) d\hat{\psi}_w(x) \frac{(\hat{\theta}_n - \theta_0)'}{\|\hat{\theta}_n - \theta_0\|}.$$

Then R_{n1} can be written as $R_{n1} = R_{n11}(\hat{\theta}_n - \theta_0)$. But

$$\|R_{n11}\| \leq \sup_{1 \leq i \leq n} \frac{|d_{ni}|}{\|\hat{\theta}_n - \theta_0\|} \int_C n^{-1} \sum_{i=1}^n K_h(x - X_i).\|\dot{\eta}_n(x; \hat{\beta}_n, \hat{\theta}_n)\| d\hat{\psi}_w(x).$$

From (v_4), we know the first factor of the above inequality is of $o_p(1)$. Applying Cauchy-Schwartz inequality to the second factor and using (2.43), the integral is $O_p(1)$. Therefore, $\sqrt{n} R_{n1} = o_p(1)\sqrt{n}(\hat{\theta}_n - \theta_0)$.

Note that the usual calculations show that

$$\int_C \dot{\eta}(x; \hat{\beta}_n, \hat{\theta}_n)\dot{\eta}(x; \hat{\beta}_n, \theta_0) d\hat{\psi}_w(x) = \Sigma_0 + o_p(1).$$

Hence, $\sqrt{n}R_{n2} = (\Sigma_0 + o_p(1))\sqrt{n}(\hat{\theta}_n - \theta_0)$. Therefore,

$$\sqrt{n}R_n = (\Sigma_0 + o_p(1))\sqrt{n}(\hat{\theta}_n - \theta_0).$$

This, together with (2.58), proved the theorem of

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\Rightarrow} N(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1}).$$

■

Proof of Theorem 2.4.3: In order to prove this theorem, it is necessary to state and prove the following lemmas.

Lemma 2.7.4. *Suppose all the conditions in Theorem 2.4.2 hold, then*

$$\begin{aligned}
(i).\quad nh^{d/2}[T_n(\hat{\beta}_n, \hat{\theta}_n) - T_n(\hat{\beta}_n, \theta_0)] &= o_p(1),\\
(ii).\quad nh^{d/2}[T_n(\hat{\beta}_n, \theta_0) - T_n(\beta_0, \theta_0)] &= o_p(1),\\
(iii).\quad nh^{d/2}[T_n(\beta_0, \theta_0) - \tilde{T}_n(\beta_0, \theta_0)] &= o_p(1).
\end{aligned}$$

Proof: Recall

$$T_n(\hat{\beta}_n, \hat{\theta}_n) = \int_C \left[\frac{\sum_{i=1}^n K_h(x - X_i)[(Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \hat{\theta}_n)]}{\sum_{i=1}^n K_w(x - X_i)}\right]^2 dG(x)$$

and $d\hat{\psi}_w(x) = dG(x)/\hat{f}_w^2(x)$. By adding and subtracting $v(X_i; \hat{\beta}_n, \theta_0)$ from $\xi_i(\hat{\beta}_n; \hat{\theta}_n) = (Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \hat{\beta}_n)$ and expanding the square terms, we can show that $T_n(\hat{\beta}_n, \hat{\theta}_n) - T_n(\hat{\beta}_n, \theta_0) = Q_{n1} - 2Q_{n2}$, where

$$
\begin{aligned}
Q_{n1} &= \int_C \left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)[v(X_i;\hat\beta_n,\hat\theta_n)-v(X_i;\hat\beta_n,\theta_0)]\right]^2 d\hat\psi_w(x),\\
Q_{n2} &= \int_C \left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\hat\xi_i\right]\cdot\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)[v(X_i;\hat\beta_n,\hat\theta_n)-v(X_i;\hat\beta_n,\theta_0)]\right] d\hat\psi_w(x).
\end{aligned}
$$

To show part (i) in Lemma 2.7.3, it necessary to show that

$$
nh^{d/2}Q_{n1}=o_p(1),\quad nh^{d/2}Q_{n2}=o_p(1). \tag{2.60}
$$

Recall the definition of d_{ni} in (2.43), we can show that $Q_{n1}\le 2Q_{n11}+2Q_{n12}$, where

$$
\begin{aligned}
Q_{n11} &= \int_C \left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)d_{ni}\right]^2 d\hat\psi_w(x),\\
Q_{n12} &= \int_C \left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\dot v_\theta'(X_i;\hat\beta_n,\theta_0)(\hat\theta_n-\theta_0)\right]^2 d\hat\psi_w(x).
\end{aligned}
$$

From the assumption of (v_4), we can show that

$$
Q_{n11}\le \|\hat\theta_n-\theta_0\|^2\sup_{1\le i\le n}\frac{|d_{ni}|^2}{\|\hat\theta_n-\theta_0\|^2}\int_C \hat f_h^2(x)d\hat\psi_w(x)=o_p(1/n),
$$

and from the assumption of (v_5),

$$
\begin{aligned}
Q_{n12} &\le 2\|\hat\theta_n-\theta_0\|^2\int_C\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\|\dot v_\theta(X_i;\hat\beta_n,\theta_0)-\dot v_\theta(X_i;\beta_0,\theta_0)\|\right]^2 d\hat\psi_w(x)\\
&+ 2\|\hat\theta_n-\theta_0\|^2\int_C\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\|\dot v_\theta(X_i;\beta_0,\theta_0)\|\right]^2 d\hat\psi_w(x)\\
&= O_p(1/n).
\end{aligned}
$$

This imply, $nh^{d/2}Q_{n1}=o_p(1)$ in (2.60). Now we'll consider Q_{n2}. By adding and subtracting $\dot{v}_\theta'(X_i;\hat{\beta}_n,\theta_0)(\hat{\theta}_n-\theta_0)$ to and from $v(X_i;\hat{\beta}_n,\hat{\theta}_n)-v(X_i;\hat{\beta}_n,\theta_0)$, we can write Q_{n2} as a sum of Q_{n21} and Q_{n22}, where

$$\begin{aligned} Q_{n21} &= \int_C [\mu_n(x;\hat{\beta}_n)-\eta_n(x;\hat{\beta}_n,\hat{\theta}_n)].\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)d_{ni}d\hat{\psi}_w(x), \\ Q_{n22} &= (\hat{\theta}_n-\theta_0)'\int_C [\mu_n(x;\hat{\beta}_n)-\eta_n(x;\hat{\beta}_n,\hat{\theta}_n)].\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\dot{v}_\theta(X_i;\hat{\beta}_n,\theta_0)d\hat{\psi}_w(x). \end{aligned}$$

By the Cauchy-Schwartz inequality, assumption (v_4) and (2.40),

$$\|Q_{n21}\|^2 \le T_n(\hat{\beta}_n,\theta_0)\|\hat{\theta}_n-\theta_0\|^2 \sup_{1\le i\le n}\left(\frac{|d_{ni}|}{\|\hat{\theta}_n-\theta_0\|}\right)^2\int_C \hat{f}_w^2(x)d\hat{\psi}_w(x) = o_p(1/n^2h^d).$$

Therefore, $nh^{d/2}Q_{21} = nh^{d/2}o_p(1/\sqrt{n^2h^d}) = o_p(1)$. Note that Q_{22} can be written as $Q_{22}'-Q_{22}''$, where

$$\begin{aligned} Q_{n22}' &= (\hat{\theta}_n-\theta_0)'\int_C [\mu_n(x;\hat{\beta}_n)-\eta_n(x;\hat{\beta}_n,\hat{\theta}_n)].\dot{\eta}_n(X_i;\hat{\beta}_n,\hat{\theta}_n)d\hat{\psi}_w(x) \\ Q_{n22}'' &= (\hat{\theta}_n-\theta_0)'\int_C [\mu_n(x;\hat{\beta}_n)-\eta_n(x;\hat{\beta}_n,\hat{\theta}_n)].[\dot{\eta}_n(X_i;\hat{\beta}_n,\hat{\theta}_n)-\dot{\eta}_n(X_i;\hat{\beta}_n,\theta_0)]d\hat{\psi}_w(x). \end{aligned}$$

By Cauchy-Schwartz inequality, we can show that

$$|Q_{n22}''|^2 \le \|\hat{\theta}_n-\theta_0\|^2\sup_{1\le i\le n}\|\dot{\eta}_n(X_i;\hat{\beta}_n,\hat{\theta}_n)-\dot{\eta}_n(X_i;\hat{\beta}_n,\theta_0)\|^2.T_n(\hat{\beta}_n,\theta_0).\int_C \hat{f}_w^2(x)d\hat{\psi}_w(x).$$

From the assumption of (v_5), the $\sqrt{n}-$ consistency of $\hat{\theta}_n$, it is clear that $|Q_{n22}''|^2 = O_p(1/n^2)$. Therefore, $nh^{d/2}Q_{n22}'' = nh^{d/2}o_p(1/n) = o_p(h^{d/2}) = o_p(1)$.

With considering Q_{n22}', note that the integration is same as the left side of (2.44), hence

$$Q_{n22}^{'} = (\hat{\theta}_n - \theta_0)^{'} \int_C [\eta_n(x;\hat{\beta}_n,\hat{\theta}_n) - \eta_n(x;\hat{\beta}_n,\theta_0)]\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n) d\hat{\psi}_w(x).$$

By adding and subtracting $\dot{\eta}_n(x;\hat{\beta}_n,\theta_0), \dot{\eta}_n(x;\beta_0,\theta_0)$ from $\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n)$, $Q_{n22}^{'}$ can be written as the sum of $Q_{n221}^{'} + Q_{n222}^{'} + Q_{n223}^{'}$, where

$$\begin{aligned}
Q_{n221}^{'} &= (\hat{\theta}_n - \theta_0)^{'} \int_C [\eta_n(x;\hat{\beta}_n,\hat{\theta}_n) - \eta_n(x;\hat{\beta}_n,\theta_0)][\dot{\eta}_n(x;\hat{\beta}_n,\hat{\theta}_n) - \dot{\eta}_n(x;\hat{\beta}_n,\theta_0)] d\hat{\psi}_w(x), \\
Q_{n222}^{'} &= (\hat{\theta}_n - \theta_0)^{'} \int_C [\eta_n(x;\hat{\beta}_n,\hat{\theta}_n) - \eta_n(x;\hat{\beta}_n,\theta_0)][\dot{\eta}_n(x;\hat{\beta}_n,\theta_0) - \dot{\eta}_n(x;\beta_0,\theta_0)] d\hat{\psi}_w(x), \\
Q_{n223}^{'} &= (\hat{\theta}_n - \theta_0)^{'} \int_C [\eta_n(x;\hat{\beta}_n,\hat{\theta}_n) - \eta_n(x;\hat{\beta}_n,\theta_0)]\dot{\eta}_n(x;\beta_0,\theta_0) d\hat{\psi}_w(x).
\end{aligned}$$

Then from the conditions (v_4), (v_5) and 2.11, we can show that $nh^{d/2}Q_{n221}^{'} = o_p(1), nh^{d/2}Q_{n222}^{'} = o_p(1)$ and $nh^{d/2}Q_{n223}^{'} = o_p(1)$. That is, $nh^{d/2}Q_{n22}^{'} = o_p(1)$. Therefore, $nh^{d/2}Q_{n22} = o_p(1)$, and $nh^{d/2}Q_{n2} = o_p(1)$ which is the second part of the (2.60) and hence the proof of (i).

Following is the proof of part (ii). By the definition of μ_n and η_n, $T_n(\hat{\beta}_n,\theta_0) - T_n(\beta_0,\theta_0)$ can be written as the sum of $A_{n1} + A_{n2} + 2A_{n3} + 2A_{n4} + 2A_{n5}$, where

$$\begin{aligned}
A_{n1} &= \int_C [\mu_n(x;\hat{\beta}_n) - \mu_n(x;\beta_0)]^2 d\hat{\psi}_w(x), \\
A_{n2} &= \int_C [\eta_n(x;\hat{\beta}_n,\theta_0) - \eta_n(x;\beta_0,\theta_0)]^2 d\hat{\psi}_w(x), \\
A_{n3} &= \int_C [\mu_n(x;\hat{\beta}_n) - \mu_n(x;\beta_0)][\mu_n(x;\beta_0) - \eta_n(x;\beta_0,\theta_0)] d\hat{\psi}_w(x), \\
A_{n4} &= \int_C [\mu_n(x;\hat{\beta}_n) - \mu_n(x;\beta_0)][\eta_n(x;\beta_0,\theta_0) - \eta_n(x;\hat{\beta}_n,\theta_0)] d\hat{\psi}_w(x), \\
A_{n5} &= \int_C [\mu_n(x;\beta_0) - \eta_n(x;\beta_0,\theta_0)][\eta_n(x;\beta_0,\theta_0) - \eta_n(x;\hat{\beta}_n,\theta_0)] d\hat{\psi}_w(x).
\end{aligned}$$

From (2.30), $nh^{d/2}A_{n1} = nh^{d/2}O_p(1/n) = O_p(h^{d/2}) = o_p(1)$ and from (2.38), $A_{n2} = D_{n1}(\theta_0) = O_p(1/n)$, hence $nh^{d/2}A_{n2} = O_p(h^{d/2}) = o_p(1)$. Now let us consider A_{n3}.. For convenience, let $U_n(x) = \mu_n(x;\beta_0) - \eta_n(x;\beta_0,\theta_0)$ and then A_{n3} can be written as $A_{n31} - 2A_{n32}$, where

$$\begin{aligned}
A_{n31} &= \int_C \frac{1}{n}\sum_{i=1}^n K_h(x - X_i)[m(X_i;\hat{\beta}_n) - m(X_i;\beta_0)]^2 U_n(x)d\hat{\psi}_w(x), \\
A_{n32} &= \int_C \frac{1}{n}\sum_{i=1}^n K_h(x - X_i)[Y_i - m(X_i;\beta_0)][m(X_i;\hat{\beta}_n) - m(X_i;\beta_0)]U_n(x)d\hat{\psi}_w(x).
\end{aligned}$$

Using the assumption of (m_2), the definition if e_{ni} in (2.31), and the Cauchy-Schwartz inequality, one can show that

$$\begin{aligned}
|A_{n31}| &\leq 2\sup_{1\leq i\leq n}|e_{ni}|^2\left(\int_C \hat{f}_h^2(x)d\hat{\psi}_w(x).T_n(x;\beta_0,\theta_0)\right)^{1/2} \\
&\quad +2\|\hat{\beta}_n - \beta_0\|^2\left(\int_C \left[\frac{1}{n}\sum_{i=1}^n K_h(x - X_i)\|\dot{m}(X_i;\beta_0)\|^2\right]^2 d\hat{\psi}_w(x).T_n(x;\beta_0,\theta_0)\right)^{1/2} \\
&= o_p(1/n)O_p(1/\sqrt{nh^d}) + O_p(1/n)O_p(1/\sqrt{nh^d}).
\end{aligned}$$

Hence, $nh^{d/2}A_{n31} = o_p(1/\sqrt{n}) + O_p(1/\sqrt{n}) = o_p(1)$. Now A_{n32} can be written as a sum of $A_{n321} + A_{n322}$, where

$$\begin{aligned}
A_{n321} &= \int_C \left[\frac{1}{n}\sum_{i=1}^n K_h(x - X_i)[Y_i - m(X_i;\beta_0)]e_{ni}\right]U_n(x)d\hat{\psi}_w(x), \\
A_{n322} &= (\hat{\beta}_n - \beta_0)\int_C \left[\frac{1}{n}\sum_{i=1}^n K_h(x - X_i)[Y_i - m(X_i;\beta_0)]\dot{m}(X_i;\beta_0)\right]U_n(x)d\hat{\psi}_w(x).
\end{aligned}$$

By Cauchy-Schwartz inequality,

$$\begin{aligned} |A_{n321}| &\le \sup_{1\le i\le n}|e_{ni}|\left(\int_C\left[\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)|Y_i-m(X_i;\beta_0)|\right]^2 d\hat{\psi}_w(x).T_n(x;\beta_0,\theta_0)\right)^{1/2} \\ &= o_p(1/\sqrt{n}).O_p(1).O_p(1/\sqrt{nh^d}). \end{aligned}$$

Hence, $nh^{d/2}A_{n321} = o_p(1)$. Again using the Cauchy-Schwartz inequality,

$$\begin{aligned} |A_{n322}| &\le \|\hat{\beta}_n-\beta_0\|\left(\int_C \|\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)[Y_i-m(X_i;\beta_0)]\dot{m}(X_i;\beta_0)\|^2 d\hat{\psi}_w(x).T_n(x;\beta_0,\theta_0)\right)^{1/2} \\ &= O_p(1/\sqrt{n}).O_p(1/\sqrt{nh^d}).O_p(1/\sqrt{nh^d}). \end{aligned}$$

Therefore, $nh^{d/2}A_{n322} = O_p(1/\sqrt{nh^d}) = o_p(1)$. This implies that $nh^{d/2}A_{n32} = o_p(1)$ and hence $nh^{d/2}A_{n3} = o_p(1)$.

Using Cauchy-Schwartz inequality on A_{n4}^2, we get

$$A_{n4}^2 \le \int_C[\mu_n(x;\hat{\beta}_n)-\mu_n(x;\beta_0)]^2 d\hat{\psi}_w(x).\int_C[\eta_n(x;\hat{\beta}_n,\theta_0)-\eta_n(x;\beta_0,\theta_0)]^2 d\hat{\psi}_w(x).$$

From (2.30) and (2.38), both above integrations are $O_p(1/n)$. Therefore, $nh^{d/2}A_{n4} = O_p(h^{d/2}) = o_p(1)$.

Finally, let's consider A_{n5}. By the definition of r_{ni} in (2.51), we have

$$\eta_n(x;\hat{\beta}_n,\theta_0)-\eta_n(x;\beta_0,\theta_0) = \frac{1}{n}\sum_{i=1}^n K_h(x-X_i)r_{ni} + (\hat{\beta}_n-\beta_0)'\frac{1}{n}\sum_{i=1}^n K_h(x-X_i)\dot{v}_\beta(X_i;\beta_0,\theta_0).$$

So, A_{n5} can be written as the sum of A_{n51} and A_{n52}, where

$$\begin{aligned} A_{n51} &= \int_C U_n(x) \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) r_{ni} d\hat{\psi}_w(x), \\ A_{n52} &= (\hat{\beta}_n - \beta_0)' \int_C \frac{1}{n} \sum_{i=1}^n K_h(x - X_i) \dot{v}_\beta(X_i; \beta_0, \theta_0) U_n(x) d\hat{\psi}_w(x). \end{aligned}$$

From (v_4) and Cauchy-Schwartz inequality,

$$|A_{n51}|^2 \leq \sup_{1 \leq i \leq n} |r_{ni}|^2 . \int_C U_n^2(x) d\hat{\psi}_w(x) . \int_C \hat{f}_h^2 d\hat{\psi}_w(x) = o_p(1/n) O_p(1/nh^d),$$

So, $nh^{d/2} A_{n51} = nh^{d/2} o_p(1/nh^{d/2}) = o_p(1)$. By adding and subtracting $EK_h(x - X)\dot{v}_\beta(X; \beta_0, \theta_0)$ from $\frac{1}{n}\sum_{i=1}^n K_h(x - X)\dot{v}_\beta(X; \beta_0, \theta_0)$, A_{n52} can be written as the sum of $A_{n511} + A_{n512}$, where

$$\begin{aligned} A_{n511} &= (\hat{\beta}_n - \beta_0)' \int_C \left[n^{-1} \sum_{i=1}^n K_h(x - X_i) \dot{v}_\beta(X_i; \beta_0, \theta_0) - EK_h(x - X) \dot{v}_\beta(X; \beta_0, \theta_0) \right] U_n(x) d\hat{\psi}_w(x), \\ A_{n512} &= (\hat{\beta}_n - \beta_0)' \int_C \left[EK_h(x - X) \dot{v}_\beta(X; \beta_0, \theta_0) \right] U_n(x) d\hat{\psi}_w(x). \end{aligned}$$

By the routing calculations, we can show that

$$\int_C \left[n^{-1} \sum_{i=1}^n K_h(x - X_i) \dot{v}_\beta(X_i; \beta_0, \theta_0) - EK_h(x - X) \dot{v}_\beta(X; \beta_0, \theta_0) \right]^2 d\hat{\psi}_w(x) = O_P(1/nh^d).$$

Therefore, $nh^{d/2} A_{n511} = nh^{d/2} O_p(1/\sqrt{n}) O_p(1/nh^d) = O_p(1/\sqrt{nh^d}) = o_p(1)$. As for A_{n512}, we first claim that $nh^{d/2} A_{n512} = nh^{d/2} \tilde{A}_{n512} + o_p(1)$, where $\tilde{A}_{n512}$ is same as for A_{n512} but with $\hat{f}_w^2(x)$ is replaced by $f^2(x)$. In fact $nh^{d/2} |A_{n512} - \tilde{A}_{n512}|$ is bounded above by

$$
\begin{aligned}
& nh^{d/2}\|\hat{\beta}_n - \beta_0\|' \int_C EK_h(x-X)\|\dot{v}_\beta(X;\beta_0,\theta_0)\||U_n(x)|\left|\frac{f^2(x)}{\hat{f}_w^2(x)} - 1\right| dG(x) \\
\leq \; & nh^{d/2} O_p(1/\sqrt{n}) O_p(1/\sqrt{nh^d}) \sup_{x\in C}\left|\frac{f^2(x)}{\hat{f}_w^2(x)} - 1\right| = o_p(1)
\end{aligned}
$$

by (2.11). So, we only have to show that $nh^{d/2}\tilde{A}_{n512} = o_p(1)$. Since $EK_h(x-X)\dot{v}_\beta(X;\beta_0,\theta_0) = \dot{v}_\beta(x;\beta_0,\theta_0)f(x) + o(1)$ uniformly for $x \in C$, hence we only need to show that

$$
nh^{d/2}(\hat{\beta}_n - \beta_0)' \int_C U_n(x)\frac{\dot{v}_\beta(x;\beta_0,\theta_0)g(x)}{f(x)}dx = o_p(1). \tag{2.61}
$$

To see this, note that

$$
\int_C U_n(x)\frac{\dot{v}_\beta(x;\beta_0,\theta_0)g(x)}{f(x)}dx = \frac{1}{n}\sum_{i=1}^n \left[\frac{1}{h^d}\int_C K\left(\frac{x-X_i}{h}\right)\frac{\dot{v}_\beta(x;\beta_0,\theta_0)g(x)}{f(x)}dx\right]\xi_i.
$$

Since, $f(x)$ has a compact support and $\dot{v}_\beta, g$, and f are continuous, we can show that

$$
\frac{1}{h^d}\int_C K\left(\frac{x-X_i}{h}\right)\frac{\dot{v}_\beta(x;\beta_0,\theta_0)g(x)}{f(x)}dx = \frac{\dot{v}_\beta(X_i;\beta_0,\theta_0)g(X_i)}{f(X_i)} + o_p(1).
$$

Hence,

$$
\int_C U_n(x)\frac{\dot{v}_\beta(x;\beta_0,\theta_0)g(x)}{f(x)}dx = \frac{1}{n}\sum_{i=1}^n \frac{\dot{v}_\beta(X_i;\beta_0,\theta_0)g(X_i)\xi_i}{f(X_i)} + o_p(1)\frac{1}{n}\sum_{i=1}^n \xi_i = O_p(1/\sqrt{n}).
$$

which implies (2.61) has order of $nh^{d/2}O_p(1/n) = o_p(1)$. So is the desired result.

$\blacksquare$

Lemma 2.7.5. *Suppose all the conditions in Theorem 2.4.2 hold, then*

$$\begin{aligned}
(i). \quad nh^{d/2}[C_n(\hat{\beta}_n, \hat{\theta}_n) - C_n(\hat{\beta}_n, \theta_0)] &= o_p(1); \\
(ii). \quad nh^{d/2}[C_n(\hat{\beta}_n, \theta_0) - C_n(\beta_0, \theta_0)] &= o_p(1); \\
(iii). \quad nh^{d/2}[C_n(\beta_0, \theta_0) - \tilde{C}_n(\beta_0, \theta_0)] &= o_p(1),
\end{aligned}$$

where C_n is as defined in (2.19).

Proof: By adding and subtracting $v(X_i; \hat{\beta}_n, \theta_0)$ from $\hat{\xi}_i = (Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \hat{\beta}_n)$, $C_n(\hat{\beta}_n, \hat{\theta}_n)$ can be written as the sum of $C_n(\hat{\beta}_n, \theta_0) + 2B_{n1} + B_{n2}$, where

$$\begin{aligned}
B_{n1} &= \frac{1}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x - X_i)[(Y_i - m(X_i; \beta))^2 - v(X_i; \hat{\beta}_n, \theta_0)][v(X_i; \hat{\beta}_n, \theta_0) - v(\hat{\beta}_n, \hat{\theta}_n)]d\hat{\psi}_w(x), \\
B_{n2} &= \frac{1}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x - X_i)[v(X_i; \hat{\beta}_n, \theta_0) - v(\hat{\beta}_n, \hat{\theta}_n)]^2 d\hat{\psi}_w(x).
\end{aligned}$$

We can see that B_{n2} is bounded above by the sum of $B_{n21} + B_{n22}$, where

$$\begin{aligned}
B_{n21} &= \frac{2}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x - X_i)d_{ni}^2 d\hat{\psi}_w(x), \\
B_{n22} &= \frac{2}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x - X_i)[\dot{v}_\theta'(X_i; \hat{\beta}_n, \theta_0)(\hat{\theta}_n - \theta_0)]^2 d\hat{\psi}_w(x),
\end{aligned}$$

and d_{ni} is as defined in (2.59).

By (v_4), and the $\sqrt{n}-$consistency of $\hat{\theta}_n$,

$$B_{n21} \le \frac{2}{n^2}\sup_{1\le i\le n}\frac{|d_{ni}|^2}{\|\hat{\theta}_n - \theta_0\|^2}.\|\hat{\theta}_n - \theta_0\|^2\sum_{i=1}^{n}\int_C K_h^2(x - X_i)d\hat{\psi}_w(x).$$

Since

$$\frac{1}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)d\hat{\psi}_w(x) = O_p(1/nh^d),$$

we can show that $nh^{d/2}|B_{n21}| = nh^{d/2}o_p(1)O_p(1/n)O_p(1/nh^d) = o_p(1)$. For B_{n22}, we have

$$\begin{aligned} B_{n22} &\leq \frac{4}{n^2}\|\hat{\theta}_n-\theta_0\|^2\sum_{i=1}^{n}\int_C K_h^2(x-X_i)\|\dot{v}_\beta(X_i;\hat{\beta}_n,\theta_0)-\dot{v}_\beta(X_i;\beta_0,\theta_0)\|^2 d\hat{\psi}_w(x), \\ &+ \frac{4}{n^2}\|\hat{\theta}_n-\theta_0\|^2\sum_{i=1}^{n}\int_C K_h^2(x-X_i)\|\dot{v}_\beta(X_i;\beta_0,\theta_0)\|^2 d\hat{\psi}_w(x) \end{aligned}$$

From (v_5), and the $\sqrt{n}-$consistency of $\hat{\theta}_n$ and $\hat{\beta}_n$, one can show that the first term is $o_p(1/n^2)$, and the second term is $O_p(1/n)O_p(1/nh^d)$. Therefore, $nh^{d/2}B_{n22} = o_p(1)$. This implies $nh^{d/2}B_{n2} = o_p(1)$. As for B_{n1}, by adding and subtracting $(Y_i - m(X_i;\beta_0))^2 - v(X_i;\beta_0,\theta_0)$ from $(Y_i - m(X_i;\hat{\beta}_n))^2 - v(X_i;\hat{\beta}_n,\theta_0)$, it can be written as the sum of $B_{n11}+B_{n12}+B_{n13}$, where

$$\begin{aligned} B_{n11} &= \frac{1}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)[(Y_i-m(X_i;\hat{\beta}_n))^2-(Y_i-m(X_i;\beta_0))^2]V_n(x)d\hat{\psi}_w(x), \\ B_{n12} &= \frac{1}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)[(Y_i-m(X_i;\beta_0))^2-v(X_i;\beta_0,\theta_0)]V_n(x)d\hat{\psi}_w(x), \\ B_{n13} &= \frac{1}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)[v(X_i;\beta_0,\theta_0)-v(X_i;\hat{\beta}_n,\theta_0)]V_n(x)d\hat{\psi}_w(x) \end{aligned}$$

and $V_n(x) = v(X_i;\hat{\beta}_n,\theta_0) - v(X_i;\hat{\beta}_n,\hat{\theta}_n)$. B_{n11} can be written as the sum of $B'_{n11}+B''_{n11}$, where

$$\begin{aligned} B'_{n11} &= \frac{1}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)[m(X_i;\hat{\beta}_n)-m(X_i;\beta_0)]V_n(x)d\hat{\psi}_w(x), \\ B''_{n11} &= \frac{2}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)[Y_i-m(X_i;\beta_0)][m(X_i;\hat{\beta}_n)-m(X_i;\beta_0)]V_n(x)d\hat{\psi}_w(x). \end{aligned}$$

Notice that

$$m(X_i;\hat{\beta}_n) - m(X_i;\beta_0) = e_{ni} + \dot{m}^{'}(X_i;\beta_0)(\hat{\beta}_n - \beta_0), \tag{2.62}$$

and

$$V_n(x) = -d_{ni} - \dot{v}_{\theta}^{'}(X_i;\hat{\beta}_n,\theta_0)(\hat{\theta}_n - \theta_0). \tag{2.63}$$

Then from $(m_2), (v_4), (v_5)$, and the $\sqrt{n}-$consistency of $\hat{\theta}_n$ and $\hat{\beta}_n$, one can show that $B_{n11}^{'} = O_p(1/n\sqrt{n})O_p(1/nh^d)$, hence $nh^{d/2}B_{n11}^{'} = o_p(1)$. Notice that

$$\frac{1}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)|Y_i - m(X_i;\beta_0)|dG(x) = O_p(1/nh^d),$$

and then by a similar argument leads to $B_{n11}^{''} = O_p(1/n)O_p(1/nh^d)$. So $nh^{d/2}B_{n11}^{''} = o_p(1)$. This implies $nh^{d/2}B_{n11} = o_p(1)$. Using (2.63), we have

$$nh^{d/2}|B_{n12}| = nh^{d/2}O_p(1/\sqrt{n})O_p(1/nh^d) = O_p(1/\sqrt{nh^d}) = o_p(1).$$

By the condition (v_4) and (2.63),

$$nh^{d/2}|B_{n13}| = nh^{d/2}O_p(1/n)O_p(1/nh^d) = O_p(1/nh^{d/2}) = o_p(1).$$

Therefore $nh^{d/2}B_{n1} = o_p(1)$ and hence the part (i) is proved.

To see (ii), adding and subtracting $(Y_i - m(X_i;\beta_0))^2 - v(X_i;\beta_0,\theta_0)$ from $(Y_i - m(X_i;\hat{\beta}_n))^2 - v(X_i;\hat{\beta}_n,\theta_0)$, then $C_n(\hat{\beta}_n,\theta_0) - C_n(\beta_0,\theta_0)$ can be written as the sum of the following five terms

$$\frac{1}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)[(Y_i-m(X_i;\hat{\beta}_n))^2-(Y_i-m(X_i;\beta_0))^2]^2 d\hat{\psi}_w(x),$$
$$\frac{1}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)[v(X_i;\beta_0,\theta_0)-v(X_i;\hat{\beta}_n,\theta_0)]^2 d\hat{\psi}_w(x),$$
$$\frac{2}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)\xi_i[(Y_i-m(X_i;\hat{\beta}_n))^2-(Y_i-m(X_i;\beta_0))^2] d\hat{\psi}_w(x),$$
$$\frac{2}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)[(Y_i-m(X_i;\hat{\beta}_n))^2-(Y_i-m(X_i;\beta_0))^2][v(X_i;\beta_0,\theta_0)-v(X_i;\hat{\beta}_n,\theta_0)] d\hat{\psi}_w(x),$$
$$\frac{2}{n^2}\sum_{i=1}^{n}\int_C K_h^2(x-X_i)\xi_i[v(X_i;\beta_0,\theta_0)-v(X_i;\hat{\beta}_n,\theta_0)] d\hat{\psi}_w(x).$$

Usual calculations will show that all five terms are $o_p(1/nh^{d/2})$. This implies the result of (ii). Finally, the claim (iii) can be shown in a similar way as in Koul and Ni (2004).

■

Lemma 2.7.6. *Suppose all the conditions in Theorem 2.4.2 hold, then*

$$(i).\quad \Gamma_n(\hat{\beta}_n,\hat{\theta}_n)-\Gamma_n(\beta_0,\theta_0)=o_p(1);$$
$$(ii).\quad \Gamma_n(\beta_0,\theta_0)-\tilde{\Gamma}_n(\beta_0,\theta_0)=o_p(1).$$

Proof: By the definition of $\hat{\xi}_i$ and ξ_i , and denoting $t_i=(Y_i-m(X_i;\hat{\beta}_n))^2-(Y_i-m(X_i;\beta_0))^2, s_i=v(X_i;\hat{\beta}_n,\hat{\theta}_n)-v(X_i;\beta_0,\theta_0)$, we have $\hat{\xi}_i=\xi_i+t_i-s_i$. Hence

$$\hat{\xi}_i\hat{\xi}_j=\xi_i\xi_j+\xi_i t_j-\xi_i s_j+t_i\xi_j+t_i t_j-t_i s_j-s_i\xi_j-s_i t_j+s_i s_j.$$

For convenience, define $\delta_{ij}=\hat{\xi}_i\hat{\xi}_j-\xi_i\xi_j$ and $K_{hi}=K_h(x-X_i)$. Then

$$\begin{aligned}\Gamma_n(\hat{\beta}_n, \hat{\theta}_n) &= \frac{2h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj}\xi_i\xi_j d\hat{\psi}_\omega(x)\right)^2 + \frac{2h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj}\delta_{ij} d\hat{\psi}_\omega(x)\right)^2 \\ &+ \frac{4h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj}\xi_i\xi_j d\hat{\psi}_\omega(x)\right)\left(\int_C K_{hi}K_{hj}\delta_{ij} d\hat{\psi}_\omega(x)\right).\end{aligned}$$

Note that the first term is just $\Gamma_n(\beta_0, \theta_0)$, we have

$$\left|\Gamma_n(\hat{\beta}_n, \hat{\theta}_n) - \Gamma_n(\beta_0, \theta_0)\right| \leq \frac{2h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj}\left|\delta_{ij}\right| d\hat{\psi}_\omega(x)\right)^2$$

$$+ \frac{4h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj}\left|\xi_i\xi_j\right| d\hat{\psi}_\omega(x)\right)\left(\int_C K_{hi}K_{hj}\left|\delta_{ij}\right| d\hat{\psi}_\omega(x)\right). \quad (2.64)$$

To proceed , we need the following facts which can be proved using similar argument in Koul and Ni (2004). For the sake of simplicity , details are omitted.

$$\frac{h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj}\left|\xi_i\xi_j\right| d\hat{\psi}_\omega(x)\right)^2 = O_p(1), \quad (2.65)$$

$$\frac{h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj}\left|\xi_i\right| d\hat{\psi}_\omega(x)\right)^2 = O_p(1), \quad (2.66)$$

$$\frac{h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj}\left|\xi_i\right| K(X_i) d\hat{\psi}_\omega(x)\right)^2 = O_p(1), \quad (2.67)$$

$$\frac{h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj} d\hat{\psi}_\omega(x)\right)^2 = O_p(1), \quad (2.68)$$

where $K(x)$ is such that $\int_C K^2(x)dG(x) < \infty$. Note that the first term on the right hand side of (2.64) is bounded above by eight terms, such as $\frac{8h^d}{n^2}\sum_{i\neq j}(\int_C K_{hi}K_{hj}\left|\xi_i t_j\right| d\hat{\psi}_\omega(x))^2$, $\frac{8h^d}{n^2}\sum_{i\neq j}(\int_C K_{hi}K_{hj}\left|\xi_i s_j\right| d\hat{\psi}_\omega(x))^2$, etc. All these eight terms can be shown as $o_p(1)$. Since the proofs are similar, we only show that the first term above is $o_p(1)$. Since $t_i = (m(X_i; \hat{\beta}_n) - m(X_i; \beta_0))^2 - 2(Y_i - m(X_i; \beta_0))(m(X_i; \hat{\beta}_n) - m(X_i; \beta_0))$ we have that $\frac{2h^d}{n^2}\sum_{i\neq j}(\int_C K_{hi}K_{hj}\left|\xi_i t_j\right| d\hat{\psi}_\omega(x))^2$

will be bounded above by the following two terms:

$$\frac{8h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj}|\xi_i(Y_i - m(X_i;\beta_0))||(m(X_i;\hat{\beta}_n) - m(X_i;\beta_0))|d\hat{\psi}_\omega(x)\right)^2 \tag{2.69}$$

and

$$\frac{4h^d}{n^2}\sum_{i\neq j}\left(\int_C K_{hi}K_{hj}|\xi_i||(m(X_i;\hat{\beta}_n) - m(X_i;\beta_0))|d\hat{\psi}_\omega(x)\right)^2. \tag{2.70}$$

By $(m2)$ and (2.67) , we can show that (2.69) has the order $O_p(1/n)$, and (2.66) has the order $O_p(1/n^2)$. Hence$\frac{2h^d}{n^2}\sum_{i\neq j}(\int_C K_{hi}K_{hj}|\xi_i t_j|d\hat{\psi}_\omega(x))^2 = o_p(1)$. By applying the Cauchy-Schwartz inequality to the double sum, we can also show that the second term on the right hand side of (2.64) is $o_p(1)$. Hence we have proven the first claim of this lemma .

Similar to the proof of Lemma 2.7.5 in Koul and Ni (2004), one can show that (ii) holds.

■

Lemma 2.7.7. *Suppose* $(e1), (e2), (e4), (f1), (g), (k), (h1)$, *and* $(v1)$ *hold; then*

$$nh^{d/2}(\tilde{T}_n(\beta_0,\theta_0) - \tilde{C}_n(\beta_0,\theta_0)) \overset{d}{\Rightarrow} N(0,\Gamma).$$

where Γ *is as defined in (2.21).*

Proof: Details of the proof of this theorem are similar to that of Lemma 5.1 in Koul and Ni (2004) with obvious modifications.

Proof of the Theorem 2.5.1: Let $Y_i^a = m(X_i;\beta_a) + \sqrt{v_a(X_i)\epsilon_i}$. Denote $K_{hi}(x) = K_h(x - X_i)$, and $K_{wi}(x) = K_w(x - X_i)$. Adding and subtracting Y_i^a from Y_i in $T_n(\hat{\beta}_n,\hat{\theta}_n)$, it can be written as the sum of $T_{n1} + 4T_{n2} + T_{n3} + 4T_{n4} + 2T_{n5} + 4T_{n6}$, where

$$
\begin{aligned}
T_{n1} &= \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(Y_i - Y_i^a)^2}{\sum_{i=1}^n K_{\omega i}(x)}\right]^2 dG(x),\\
T_{n2} &= \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(Y_i - Y_i^a)(Y_i^a - m(X_i;\hat{\beta}_n))}{\sum_{i=1}^n K_{\omega i}(x)}\right]^2 dG(x),\\
T_{n3} &= \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)[Y_i^a - m(X_i;\hat{\beta}_n) - v(X_i;\hat{\beta}_n,\hat{\theta}_n)]}{\sum_{i=1}^n K_{\omega i}(x)}\right]^2 dG(x),\\
T_{n4} &= \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(Y_i - Y_i^a)^2}{\sum_{i=1}^n K_{\omega i}(x)}\right]\left[\frac{\sum_{i=1}^n K_{hi}(x)(Y_i - Y_i^a)(Y_i^a - m(X_i;\hat{\beta}_n))}{\sum_{i=1}^n K_{\omega i}(x)}\right] dG(x),\\
T_{n5} &= \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(Y_i - Y_i^a)^2}{\sum_{i=1}^n K_{\omega i}(x)}\right]\left[\frac{\sum_{i=1}^n K_{hi}(x)[(Y_i^a - m(X_i;\hat{\beta}_n))^2 - v(X_i;\hat{\beta}_n,\hat{\theta}_n)]}{\sum_{i=1}^n K_{\omega i}(x)}\right] dG(x),\\
T_{n6} &= \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(Y_i - Y_i^a)(Y_i^a - m(X_i;\hat{\beta}_n))}{\sum_{i=1}^n K_{\omega i}(x)}\right]\cdot\\
&\qquad\qquad \left[\frac{\sum_{i=1}^n K_{hi}(x)[(Y_i^a - m(X_i;\hat{\beta}_n))^2 - v(X_i;\hat{\beta}_n,\hat{\theta}_n)]}{\sum_{i=1}^n K_{\omega i}(x)}\right] dG(x).
\end{aligned}
$$

Using Cauchy-Schwartz inequality , one can show that T_{n5}, T_{n6} are the order of $o_p(1)$. Note that under $H_a, Y_i - Y_i^a = m(X_i;\beta_0) - m(X_i;\beta_a) + [\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)}]\epsilon_i$. Then T_{n1} can be written as the sum $T_{n11} + T_{n12} + T_{n13} + T_{n14} + T_{n15} + T_{n16}$, where

$$\begin{aligned}
T_{n11} &= \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(m_0(X_i) - m_a(X_i))^2}{\sum_{i=1}^n K_{\omega i}(x)} \right]^2 dG(x), \\
T_{n12} &= \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})^2 \epsilon_i^2}{\sum_{i=1}^n K_{\omega i}(x)} \right]^2 dG(x), \\
T_{n13} &= 2\int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(m_0(X_i) - m_a(X_i))^2}{\sum_{i=1}^n K_{\omega i}(x)} \right] \cdot \\
&\qquad \left[\frac{\sum_{i=1}^n K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})^2 \epsilon_i^2}{\sum_{i=1}^n K_{\omega i}(x)} \right] dG(x), \\
T_{n14} &= 4\int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(m_0(X_i) - m_a(X_i))(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})\epsilon_i}{\sum_{i=1}^n K_{\omega i}(x)} \right]^2 dG(x), \\
T_{n15} &= 4\int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(m_0(X_i) - m_a(X_i))^2}{\sum_{i=1}^n K_{\omega i}(x)} \right] \cdot \\
&\qquad \left[\frac{\sum_{i=1}^n K_{hi}(x)(m_0(X_i) - m_a(X_i))(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})\epsilon_i}{\sum_{i=1}^n K_{\omega i}(x)} \right] dG(x), \\
T_{n16} &= 4\int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})^2 \epsilon_i^2}{\sum_{i=1}^n K_{\omega i}(x)} \right] \cdot \\
&\qquad \left[\frac{\sum_{i=1}^n K_{hi}(x)(m_0(X_i) - m_a(X_i))(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})\epsilon_i}{\sum_{i=1}^n K_{\omega i}(x)} \right]^2 dG(x).
\end{aligned}$$

While $T_{n11} \to \int_C [m_0(x) - m_a(x)]^4 dG(x)$, $T_{n12} \to \int_C [\sqrt{v_1(x)} - \sqrt{v_a(x)}]^4 dG(x)$ and $T_{n13} \to 2\int_C [m_0(x) - m_a(x)]^2 [\sqrt{v_1(x)} - \sqrt{v_a(x)}]^2 dG(x)$. The remainder terms converges to 0 in probability.

So

$$T_{n1} \to \int_C \left([m_0(x) - m_a(x)]^2 + [\sqrt{v_1(x)} - \sqrt{v_a(x)}]^2 \right)^2 dG(x) \qquad (2.71)$$

in probability.

Now, let us consider T_{n2}. Denote $m_n(x) = m(x; \hat{\beta}_n)$. By the definition of Y_i^a, T_{n2} can be written as the sum of T_{n21} and a remainder, where

$$T_{n21} = \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)[\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)}]\sqrt{v_a(X_i)}\epsilon_i^2}{\sum_{i=1}^n K_{\omega i}(x)} \right]^2 dG(x) \tag{2.72}$$

Condition (m2), the $\sqrt{n}$-consistency of $\hat{\beta}_n$, and Cauchy-Schwartz inequality imply the remainder term is $o_p(1)$, and a routing argument leads to $T_{n21} = \int_C [\sqrt{v_1(x)} - \sqrt{v_a(x)}]^2 v_a(x) dG(x) + o_p(1)$. Hence $T_{n21} \to \int_C [\sqrt{v_1(x)} - \sqrt{v_a(x)}]^2 v_a(x) dG(x)$ in probability. As for T_{n3}, similar to the arguments in proving Theorem 2.4.3 , one can show that

$$nh^{d/2}(T_{n3} - C_n^a) \Rightarrow N(0, \Gamma_a) \tag{2.73}$$

Where

$$C_n^a = \frac{1}{n^2} \sum_{i=1}^n \int_C K_h^2(x - X_i)[(Y_i^a) - m(X_i; \hat{\beta}_n)^2 - v(X_i; \hat{\beta}_n; \hat{\theta}_n)]^2 d\hat{\psi}_w(x)$$

and Γ_a is the same as Γ in the null case except for β_0 and θ_0 being replaced by $\beta_a$and θ_a, respectively.

Using the definition of Y_i^a, T_{n4} can be written as a sum of twelve terms. One can show that all other terms are negligible in probability, except for the following two terms,

$$\begin{aligned}
B_{n1} &= \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(m_0(X_i) - m_a(X_i))^2}{\sum_{i=1}^n K_{\omega i}(x)} \right] \cdot \\
&\qquad \left[\frac{\sum_{i=1}^n K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})(\sqrt{v_a(X_i)})\epsilon_i^2}{\sum_{i=1}^n K_{\omega i}(x)} \right] dG(x), \\
B_{n2} &= \int_C \left[\frac{\sum_{i=1}^n K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})^2 \epsilon_i^2}{\sum_{i=1}^n K_{\omega i}(x)} \right] \cdot \\
&\qquad \left[\frac{\sum_{i=1}^n K_{hi}(x)(\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})(\sqrt{v_a(X_i)})\epsilon_i^2}{\sum_{i=1}^n K_{\omega i}(x)} \right] dG(x).
\end{aligned}$$

In fact, one can show that

$$\begin{aligned} B_{n1} &= \int_C [m_0(x) - m_a(x)]^2(\sqrt{v_1(x)} - \sqrt{v_a(x)})\sqrt{v_a(x)}dG(x) + o_p(1) \\ B_{n2} &= \int_C (\sqrt{v_1(x)} - \sqrt{v_a(x)})^3\sqrt{v_a(x)}dG(x) + o_p(1) \end{aligned}$$

That is

$$\begin{aligned} T_{n4} &= \int_C [m_0(x) - m_a(x)]^2(\sqrt{v_1(x)} - \sqrt{v_a(x)})\sqrt{v_a(x)}dG(x) \\ &\qquad + \int_C (\sqrt{v_1(x)} - \sqrt{v_a(x)})^3\sqrt{v_a(x)}dG(x) + o_p(1). \end{aligned}$$

By some simple algebra, one can show that

$$T_{n1} + 4T_{n2} + 4T_{n4} = \Delta + o_p(1). \tag{2.74}$$

Under the alternative hypothesis H_1, $C_n(\hat{\beta}_n, \hat{\theta}_n)$ can be written as C_n^a plus a remainder which can be shown as a negligible term. While Γ_n, after adding and subtracting Y_i^a from Y_i, Y_j^a from Y_j , can be written as a sum of bounded in probability terms. Details are similar to that of Koul and Song (2009) and hence we omit the proof for the sake of simplicity. Combining the results from (2.74), and the asymptotic distributions of $\Gamma_n(\hat{\beta}_n, \hat{\theta}_n)$, and $C_n(\hat{\beta}_n, \hat{\theta}_n)$, one can see that $nh^{d/2}\Gamma_n^{-1/2}(\hat{\beta}_n, \hat{\theta}_n)[T_n(\hat{\beta}_n, \hat{\theta}_n) - C_n(\hat{\beta}_n, \hat{\theta}_n)] = nh^{d/2}\Gamma_n^{-1/2}[T_{n1} + 4T_{n2} + 4T_{n4}] + o_p(nh^{d/2})$ which tends to ∞ as long as $\Delta > 0$. This implies the consistency of the minimum distance test. Hence the proof of the theorem.

■

Proof of Theorem 2.5.2: Details of the proof of this theorem are similar to that of to the Theorem 2.5.1 with obvious modification.

Chapter 3

Empirical Smoothing Lack-of-Fit Tests For Variance Function

This section discusses a nonparametric Empirical Smoothing Lack-of-Fit test for the functional form of the variance in regression models. The proposed test can be treated as a nontrivial modification of Zheng (1996)'s nonparametric smoothing test and Koul and Ni (2004)'s minimum distance test for the mean function in the classic regression models. The section establishes the asymptotic normality of the proposed test under the null hypothesis. Consistency at some fixed alternatives and asymptotic power under some local alternatives are also discussed. A simulation study is conducted to assess the finite sample performance of the proposed test. The simulation study also shows that the proposed test is more powerful and computationally more efficient than some existing tests.

3.1 Introduction

The proposed test in the previous section using minimum distance method requires the calculation of the integrations in the test statistics. These integrations usually do not have a tractable form, so some numerical methods are needed to approximate the integrations. The empirical L_2 test proposed in this section is much simpler and computationally easier than that using the minimum distance method.

This section is organized as follows. The Empirical Smoothing Lack-of-Fit test statistic and the technical assumptions are stated in subsection 3.2. The asymptotic null distribution, the consistency and local power study of the test are presented in subsection 3.3. Subsection 3.4 contains simulation studies to show the finite sample performance of the test. Subsection 4.5 gives a comparison remarks of Minimum Distance test, Empirical Smoothing Lack-of-Fit test, and the test proposed by Wang and Zhou (2006). All the proofs of main results regarding this section are presented in subsection 4.6.

3.2 Test Statistic and Assumptions

In this subsection, a new lack-of-fit test is proposed to check the adequacy of a parametric form of the variance function in the heteroscedastic regression models. To be specific, consider the following regression model,

$$Y = m(X; \beta) + \sqrt{v(X)}\epsilon \tag{3.1}$$

where Y is a one dimensional response variable, X is a d-dimensional explanatory variable, $m(x; \beta)$ is the mean function of known form characterized by the unknown p-dimensional parameter β, and $v(x)$ is the conditional variance function of Y given $X = x$.

The hypothesis to be tested is

$$H_0 : v(X) = v(X; \beta_0, \theta_0) \text{ for some } (\beta_0, \theta_0) \in \Gamma \times \Theta \quad \text{v.s.} \quad H_1 : H_0 \text{ is not true.} \tag{3.2}$$

Assuming that the error term ϵ satisfies $E(\epsilon|X) = 0$ and $E(\epsilon^2|X) = 1$, we have

$$E[(Y - m(X; \beta))^2 | X = x] = v(x), \tag{3.3}$$

which implies that testing the variance function in model (3.1) is equivalent to testing the mean function in the following regression model

$$(Y - m(X; \beta))^2 = v(X) + \xi \tag{3.4}$$

if β is known, where $(Y - m(X; \beta))^2$ is viewed as the response variable and, $\xi = (Y - m(X; \beta))^2 - E[(Y - m(X; \beta))^2 | X]$ is the error term, uncorrelated with X. Similar to Koul and Ni (2004), a lack-of-fit test is developed in the previous section for H_0 in (3.2) based on the quantity of

$$T_n(\beta, \theta) = \int_C \left[\frac{h^{-d} \sum_{i=1}^n K_h(x - X_i)[(Y_i - m(X_i; \beta))^2 - v(X_i; \beta, \theta)]}{w^{-d} \sum_{i=1}^n K_w(x - X_i)} \right]^2 dG(x), \tag{3.5}$$

where C is a compact set in R^d, G is a weighting measure with C being a compact subset of its support, K is a kernel function, $K_h(\cdot) = K(\cdot/h)$, and h, w are the bandwidths. In real applications, β and θ are usually unknown. In the previous method, β is estimated in advance, θ is estimated by $\hat{\theta}_n = \arg\min_{\theta \in \Theta} T_n(\hat{\beta}_n, \theta)$. The test statistic is then constructed from $T_n(\hat{\beta}_n, \hat{\theta}_n)$. The integral in $T_n(\hat{\beta}_n, \hat{\theta}_n)$ usually does not have a tractable form. Therefore, one has to approximate the integration using some numerical methods to implement the test. These numerical methods either take a long execution time because of the complex iterations or provide unstable results because of the subjectivity of choosing some tuning parameters in the algorithms. Zheng (1996) provided a nonparametric smoothing test for checking the adequacy of mean function forms. This test has a close connection with Koul and Ni (2004) minimum distance method, but it does not need to calculate any integrations. Wang and Zhou (2006) applied Zheng (1996) method to test the hypothesis (3.2). Denote $\xi_i = (Y_i - m(X_i;\beta_0))^2 - v(X_i;\beta_0,\theta_0)$. Note that under H_0,

$$E(\xi_i|X_i) = 0 \quad \text{and} \quad E[\xi_i E(\xi_i|X_i) f(X_i)] = 0, \quad \text{for} \quad i = 1, 2, \cdots, n \tag{3.6}$$

while under H_1, since $E(\xi_i|X_i) = v(X_i) - v(X_i;\beta_0,\theta_0)$, it is clear that

$$E[\xi_i E(\xi_i|X_i) f(X_i)] = E[((E(\xi_i|X_i))^2) f(X_i)] = E[(v(X_i) - v(X_i;\beta_0,\theta_0))^2 f(X_i)] > 0. \tag{3.7}$$

Applying Zheng (1996)'s idea, Wang and Zhou (2006)'s test is based on the quantity

$$n^{-1} \sum_{i=1}^{n} \xi_i E(\xi_i|X_i) f(X_i) \tag{3.8}$$

which is a sample analogue of $E[\xi_i E(\xi_i|X_i) f(X_i)]$. The estimators of $E(\xi_i|X_i)$ and $f(X_i)$ using the leave-one-out Nadaraya-Watson kernel estimates,

$$\hat{E}(\xi_i|X_i) = \frac{1}{(n-1)\hat{f}(X_i)} \sum_{j \neq i} \frac{1}{h^d} K_h(i,j) e_j, \quad \text{and} \quad \hat{f}(X_i) = \frac{1}{(n-1)} \sum_{j \neq i} \frac{1}{h^d} K_h(i,j), \tag{3.9}$$

respectively, where $e_i = (Y_i - m(X_i;\hat{\beta}_n))^2 - v(X_i;\hat{\beta}_n,\hat{\theta}_n)$, $i = 1, 2, \ldots, n$, $\hat{\beta}_n$ and $\hat{\theta}_n$ are any $\sqrt{n}$-consistent estimator of β_0 and θ_0, the true parameter of β and θ under the null hypothesis, respectively, and $K_h(i,j) = K((X_i - X_j)/h)$. Wang and Zhou (2006) test is then constructed from the following quantity

$$Z_n = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{h^d} K_h(i,j) e_i e_j.$$

Similar to the question raised in Song and Du (2011) when checking the adequacy of mean function, we wonder why not use the empirical version of the second term in (3.7) to build the test statistic? An attractive feature of the empirical version of $E(E^2(\xi|X)f(X))$ is that the variance of this empirical version will be less than that of (3.8) used in Wang and Zhou (2006), which is derived from the following fact

$$E(E^4(\xi|X)f^2(X)) \leq EE(\xi^2|X)E^2(\xi|X)f^2(X) = E\xi^2 E^2(\xi|X)f^2(X) \tag{3.10}$$

by applying Cauchy-Schwartz inequality. So if a new test is constructed based on the standardized sample analogue of the second term in (3.7), comparing to Zheng (1996)'s test which uses the standardized sample analogue of the first term in (3.7) as the test statistic, we will find that these two test statistics might have similar numerators based on the first equality in (3.7), while the new test statistic has a smaller denominator than Wang and Zhou (2006)'s test statistic. This implies that the new test might be more powerful than Wang and Zhou (2006)'s test. Although that the variance of the population version (3.10) is smaller than that of (3.8) does not necessarily imply their empirical counterparts posses the same relationship, in particular, after replacing all unknown quantities with the estimators, but it is intuitively appealing to investigate the actual performance of the new test. Comparing with Wang and Zhou (2006)'s test, the new test statistic is relatively complicated, in particular, the appearance of the kernel estimator of $f(x)$ in the denominator needs some extra conditions to avoid the possible asymptotic negligibility at the boundary points and the possible numeric instability when $f(x)$ is small. In real applications, if we are not sure whether or not these conditions hold for $f(x)$, then special attention should be paid when employing the proposed method. But except that, the new test shares the same advantages as Wang and Zhou (2006)'s test. Another important fact revealed in the current work is the inherent connection between the selection of smoothing parameter and the choice of kernel functions, which is also found in Song and Du (2011).

Using the leave-one-out estimators in (3.9), the sample analogue of $E[(E(\xi|X))^2 f(X)]$ is given by

$$\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{(n-1)h^d}\sum_{j\neq i}K_h(i,j)e_j\right]^2 \hat{f}^{-1}(X_i). \tag{3.11}$$

By expanding the square term, it can be written as

$$\frac{1}{n(n-1)^2h^{2d}}\sum_{i=1}^{n}\left[\sum_{j\neq i}\sum_{k\neq i}K_h(i,j)K_h(i,k)e_je_k\right]\hat{f}^{-1}(X_i). \tag{3.12}$$

Similar to the leave-one-out technique in (3.9), we drop all the terms with $k=j$ from the third sum in (3.12), accordingly, change one $1/(n-1)$ into $1/(n-2)$. Then the test statistic we are proposing has the form

$$Z_n = \frac{1}{n(n-1)(n-2)h^{2d}}\sum_{i=1}^{n}\left[\sum_{j\neq i}\sum_{k\neq i,j}K_h(i,j)K_h(i,k)e_je_k\right]\hat{f}^{-1}(x_i), \tag{3.13}$$

Denote $\dot{m}(x;\beta)$ as the derivative of $m(x;\beta)$ with respect to β, and $\dot{v}_\beta(x;\beta,\theta), \dot{v}_\theta(x;\beta,\theta)$ be derivatives of $v(x;\beta,\theta)$ with respect to β and θ respectively. The following is a list of technical assumptions needed for proving the main results in the paper.

C1: The design variable X has a compact support I and $min_{x\in I}f(x) \geq c$, where c is a positive constant.This typical restriction avoids a nonparametric estimator of $f(x)$ from vanishing near the boundary of the design space.

C2: $m(x;\beta)$, $v(x;\beta,\theta)$ and their derivatives $\dot{m}(x;\beta)$, $\dot{v}_\beta(x;\beta,\theta), \dot{v}_\theta(x;\beta,\theta)$ are continuous in x for all θ and β.

C3: $E[\epsilon^4|X = x]$ is continuous in x.

C4: For any $\sqrt{n}$-consistent estimator of β_0,

$$\sup_{1\le i\le n} |m(X_i; \hat{\beta}_n) - m(X_i; \beta_0) - (\hat{\beta}_n - \beta_0)' \dot{m}(X_i; \beta_0)| = O_p(1/n).$$

C5: For any $\sqrt{n}$-consistent estimators $\hat{\beta}_n, \hat{\theta}_n$ of β_0 and θ_0 respectively,

$$\sup_{1\le i\le n} |v(X_i; \hat{\beta}_n, \hat{\theta}_n) - v(X_i; \beta_0, \theta_0) - (\hat{\beta}_n - \beta_0)' \dot{v}_\beta(X_i; \beta_0, \theta_0) - (\hat{\theta}_n - \theta_0)' \dot{v}_\theta(X_i; \beta_0, \theta_0)| = O_p(1/n).$$

C6: The Kernel function K is nonnegative, bounded, continuous, and symmetric function such that $\int K(u)du = 1$. This is the most commonly used one in the nonparametric literature. Note that the boundedness of K implies $\int K^2(u)du < \infty$.

C7: The bandwidth h is chosen so that $h \to 0$ and $nh^{2d} \to \infty$ as $n \to \infty$.

Condition (C1) is a typical restriction that avoids a nonparametric estimator of $f(x)$ from vanishing near the boundary of the design space; Conditions (C4) and (C5) might appear stronger, but if the second derivatives of $m(x; \beta)$ and $v(x; \beta, \theta)$ with respect to β and θ are bounded in a neighborhood of β_0, θ_0, then (C4) and (C5) hold. The conditions (C6) and (C7) are the typical assumptions adopted in nonparametric smoothing literature. The condition (C3) is imposed when proving the theorem regarding the local power of the test under a fixed alternative.

3.3 Main Results

This subsection is devoted to present the main results of the proposed nonparametric Empirical Smoothing Lack-of-Fit test. For the sake of simplicity, denote,

$$\tau^2(x;\beta,\theta) = E(\xi^2|X=x) = v^2(x;\beta,\theta)[E(\epsilon^4|X=x)-1]. \tag{3.14}$$

The asymptotic distribution of Z_n under the null hypothesis is given in the following theorem.

Theorem 3.3.1. *Assume that the conditions (C1)-(C7) hold, then under H_0 in (3.2), $nh^{d/2}Z_n \Longrightarrow N(0,\sigma^2)$, where*

$$\sigma^2 = 2\int\left[\int K(u+v)K(v)dv\right]^2 du \cdot \int [\tau^2(x;\beta_0,\theta_0)]^2 f^2(x)dx. \tag{3.15}$$

Let $H(u) = \int K(u+v)K(v)dv$, which is the convolution of K. Then σ^2 can be consistently estimated by $\hat{\sigma}^2$, where

$$\hat{\sigma}^2 = \frac{2}{n(n-1)}\sum_{i\neq j}\frac{1}{h^d}H^2\left(\frac{x_i-x_j}{h}\right)e_i^2 e_j^2.$$

Thus, the test that rejects H_0 whenever,

$$T_n = \frac{nh^{d/2}|Z_n|}{\hat{\sigma}} > Z_{\alpha/2} \tag{3.16}$$

is of the asymptotic size α, where Z_α is the $(1-\alpha)100$th percentile of the standard normal distribution.

The result above is similar to that in Wang and Zhou (2006) except for the first integration in σ^2. The integration in Wang and Zhou (2006)'s result is $\int K^2(v)dv$. Note that H is the convolution of K, by Cauchy-Schwartz inequality, one can easily show that

$\int H^2(v)dv \le \int K^2(v)dv$. That is, our test has a smaller asymptotic variance than that of Wang and Zhou (2006)'s test.

Although the motivation of the current research is to construct a more precise test by modifying Wang and Zhou (2006)'s test, it turns out that there are some interesting connections with the method proposed in the previous section using minimum distance test based on (3.5). In fact, if we choose $w = h$, $dG(x) = \hat{f}_h(x)dF_n(x)$ in (3.5), where $F_n(x)$ is the empirical cumulative distribution function of X_i's, then after a slight and obvious modification, $T_n(\hat{\beta}_n, \hat{\theta}_n)$ is simply Z_n defined in (3.13).

The proof of Theorem 3.3.1, which is postponed to subsection 5.5, shows that

$$nh^{d/2}Z_n = \frac{1}{(n-1)h^{d/2}} \sum_{j \neq k} H\left(\frac{X_j - X_k}{h}\right) \epsilon_j \epsilon_k + o_p(1) := V_n + o_p(1). \tag{3.17}$$

This also gives an interesting connection between our test and Wang and Zhou (2006)'s test: Our test is asymptotically equivalent to Wang and Zhou (2006)'s test with the kernel function K replaced with the convolution H of K, nevertheless, our test is more powerful.

If one wants to construct a test based on V_n in (3.17) with the random errors ϵ_i's replaced by the residual e_i's, denoted it as $\hat{V}_n$, that is, we will rejects H_0 whenever

$$R_n = nh^{d/2}|\hat{V}_n|/\hat{\sigma} > z_{\alpha/2}, \tag{3.18}$$

then the conditions needed for the asymptotic theory can be greatly simplified. For example, (C1) can be removed, and (C7) can be changed to $nh^d \to \infty$.

Typically, nonparametric tests are design to be omnibus, in the sense that they are consistent against a very wide class of fixed alternatives. A test is said to be consistent against a given alternative if the power of the test under that alternative tends to 1 as sample size tends to ∞. Let $v_1(x)$ be a known positive real valued function such that

$v_1 \notin \{v(x;\beta,\theta) : (\beta,\theta) \in \Gamma \times \Theta\}$. Consider the alternative hypothesis

$$H_a : E((Y - m(X,\beta))^2|X) = v_1(X). \tag{3.19}$$

Under the null hypothesis, we have assumed that estimator $\hat{\theta}_n$ is $\sqrt{n}$-consistent for the true parameter θ_0. Would this estimator still have the similar property under the alternative hypothesis H_a? The question is of interest in its own right. In the classic regression setup, Jennrich (1969) and (White, 1981, 1982) showed that, under some mild regularity conditions, the nonlinear least squares estimator converges in probability and is asymptotically normal even in the presence of model misspecification. Suppose the true value of β under H_a is still β_0, the estimator $\hat{\beta}_n$ is usually not a consistent estimator for β_0. But under some regularity conditions, it is a consistent estimator of some other value, say β_a; moreover $\hat{\beta}_n$ is still asymptotically normal. In the following, we simply assume that $\sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1)$ and $\sqrt{n}(\hat{\beta}_n - \beta_a) = O_p(1)$ under the alternative H_a for some $\beta_a \in \Gamma, \theta_a \in \Theta$. We will not justify this assumption rigorously here.

The following theorem states the asymptotic property of the test statistic under H_a.

Theorem 3.3.2. *Suppose the conditions $(C1) - (C7)$ hold with β_0 and θ_0 replaced by β_a and θ_a. Then under the alternative hypothesis H_a in (3.19),*
$Z_n \to E\left[(m(X;\beta_0) - m(X;\beta_a))^2 + (v_1(X) - v(X;\beta_a,\theta_a))\right]^2 f(X)$ *in probability and,*

$$\begin{aligned}\hat{\sigma}^2 \to \int \left[\int K(u+v)K(v)dv\right]^2 du \cdot \\ \int \left[\tau^2(x;\beta_a,\theta_a) + (m(X;\beta_0) - m(X;\beta_a))^2 + (v_1(x) - v(x;\beta_a,\theta_a))\right]^2 f^2(x)dx,\end{aligned} \tag{3.20}$$

in probability, where $\tau^2(x;\beta_a,\theta_a)$ is defined as in (3.14)

The consistency of the test is thus implied by the positiveness of $E[(m(X;\beta_0) - m(X;\beta_a))^2 +$

$(v_1(X) - v(X;\beta_a,\theta_a))]^2 f(X)$ and the finiteness of the right hand side of (3.20). Comparing with the corresponding result in the previous method, this result only differs in the denominator, like the null case. This implies the current test will be more powerful than Zheng (1996)'s test for fixed alternatives.

Sometimes, it would also be desirable to investigate how sensitive the test is to local alternatives. For this purpose, let $\delta(x)$ be a positive real valued function such that $\int_c \delta^2(x)dG(x) < \infty$. Note that $\delta(x)$ is a function that is not in the parametric class of $\{v(X;\beta_0,\theta_0) : (\beta,\theta) \in \Gamma \times \Theta\}$. Consider the following local alternative

$$H_{Loc} : v(x) = v(X;\beta_0,\theta_0) + c_n\delta(x),\ \forall x \in I, \tag{3.21}$$

where c_n is a sequence of numbers converging to zero.

Under H_{Loc}, the regression model has the form

$$Y = m(X;\beta_0) + \sqrt{v(X;\beta_0,\theta_0) + c_n\delta(X)}\epsilon.$$

The following theorem states that the proposed test has nontrivial power against a sequence of local alternatives which approaches to the null hypothesis at the rate of $1/\sqrt{nh^{d/2}}$.

Theorem 3.3.3. *Given the assumptions $(C1) - (C7)$ hold, then under the local alternative hypothesis H_{Loc} in (3.21), $nh^{d/2}Z_n/\hat{\sigma} \Rightarrow N(\mu, 1)$, where $\mu = E[\delta^2(x)f(x)]/\sigma$, and σ is defined as in (3.15).*

3.4 Simulation

This section investigates the finite sample performance of the proposed test through a Monte Carlo simulation study. We generate samples from the following models:

$$\begin{aligned}
\text{Model0}: &\quad Y_i = \beta_1 + \beta_2 X_i + \sqrt{\theta_1 + \theta_2 X_i}\epsilon_i,\\
\text{Model1}: &\quad Y_i = \beta_1 + \beta_2 X_i + \sqrt{\theta_1 + \theta_2 X_i + 0.5X_i^2}\epsilon_i,\\
\text{Model2}: &\quad Y_i = \beta_1 + \beta_2 X_i + \sqrt{\theta_1 + \theta_2 X_i + 0.8X_i^2}\epsilon_i,\\
\text{Model3}: &\quad Y_i = \beta_1 + \beta_2 X_i + \sqrt{\theta_1 + \theta_2 X_i + X_i^2}\epsilon_i,
\end{aligned}$$

for $i = 1, 2, \cdots, n$.

The data from model 0 are used to study the empirical level, while the data from model 1-3 are used to study the empirical power of the test. In this simulation, $X_i \sim U(-3, 3)$, for $i = 1, 2, \cdots, n$, with $\beta_1 = 1, \beta_2 = 2, \theta_1 = 2$ and $\theta_2 = 0.1$. Two types of error distributions are considered, $\epsilon \sim N(0, 1)$ and $\epsilon \sim \mathrm{U}(-\sqrt{3}, \sqrt{3})$. The kernel function K is chosen to be the standard normal and the bandwidth is set to be $h = an^{-1/3}$ where a is a positive constant and the sample sizes are chosen to be $n = 100, 200, 300, 400, 500$, and 800. In the simulation, we chose $a = 0.5, 0.8$, and 1 to see the influence of the bandwidth on the power of the test. For all scenarios, the nominal significance level is chosen to be 0.05, and the test is repeated 500 times. The empirical size and power are computed by using the relative frequency of the event $\#\{T_n(\hat{\beta}_n, \hat{\theta}_n) \geq 1.96\}/500$ with T_n being defined in (3.16). Table 3.1 shows the empirical level and power of the test for $a = 1$.

This simulation study shows that the empirical levels are all less than the nominal level 0.05 and hence the proposed test is conservative for all chosen sample sizes and for both error types. This is common for nonparametric smoothing tests. The empirical powers against all alternative models get larger when the sample sizes get larger. The power performance

Table 3.1: *Empirical sizes and powers of the Empirical Smoothing test ($a = 1$)*

		100	200	300	400	500	800
$\epsilon \sim N(0,1)$	Model 0	0.014	0.022	0.020	0.042	0.024	0.024
	Model 1	0.056	0.342	0.666	0.844	0.942	0.998
	Model 2	0.134	0.650	0.926	0.982	1.000	1.000
	Model 3	0.224	0.748	0.960	1.000	1.000	1.000
$\epsilon \sim U(-\sqrt{3}, \sqrt{3})$	Model 0	0.008	0.002	0.012	0.014	0.022	0.016
	Model 1	0.358	0.872	0.990	0.998	1.000	1.000
	Model 2	0.602	0.998	1.000	1.000	1.000	1.000
	Model 3	0.730	0.966	1.000	1.000	1.000	1.000

is satisfactory and it is even higher for uniform errors than normal errors with the same means and standard deviations.

We also conduct a simulation study using a bootstrap method as it generally provides more accurate approximation to the distribution of the test statistic than asymptotic normal theory does when the sample size is small to moderate. The bootstrap method we use in this study is same as that of in the previous section. We use 400 bootstrap samples per run to obtain the critical value c^*_α. The empirical size and power are computed by using the relative frequency of the event $\#\{T_n(\hat{\beta}_n, \hat{\theta}_n) \geq c^*_\alpha\}/500$. Table 3.2 shows the empirical level and power of the test for $a = 1$ using the bootstrap method.

Table 3.2: *Empirical sizes and powers of the Empirical Smoothing test using bootstrapping method ($a = 1$)*

		100	200	300	400	500	800
$\epsilon \sim N(0,1)$	Model 0	0.062	0.046	0.054	0.055	0.043	0.048
	Model 1	0.120	0.390	0.716	0.880	0.968	0.994
	Model 2	0.174	0.694	0.932	0.999	1.000	1.000
	Model 3	0.260	0.782	0.976	0.998	1.000	1.000
$\epsilon \sim U(-\sqrt{3}, \sqrt{3})$	Model 0	0.061	0.059	0.062	0.052	0.050	0.050
	Model 1	0.474	0.950	0.996	0.998	1.000	1.000
	Model 2	0.744	0.996	1.000	1.000	1.000	1.000
	Model 3	0.790	1.000	1.000	1.000	1.000	1.000

Similar pattern as in Table 3.1 can be seen in the Table 3.2, but the empirical levels are close to the nominal level 0.05 when the sample sizes get larger. We also conduct some simulation studies for different values of a. Since the simulation results are similar, we will not report them here for the sake of brevity.

3.5 Results Comparison

As a comparison, we carry out a simulation study for the test proposed in Chapter 2 using bootstrapping method. The simulation results are shown in Table 3.3 for $a = 1$. The simulation results for other values of a are similar, hence omitted here. Comparing Table 3.2 and 3.3, we can see that the proposed test is more powerful than the minimum distance test proposed in the previous section.

Table 3.3: *Empirical sizes and powers of the Minimum Distance test using bootstrapping method ($a = 1$)*

		100	200	300	400	500	800
$\epsilon \sim N(0,1)$	Model 0	0.044	0.038	0.042	0.040	0.044	0.045
	Model 1	0.130	0.170	0.252	0.312	0.358	0.526
	Model 2	0.214	0.330	0.452	0.556	0.600	0.846
	Model 3	0.226	0.386	0.552	0.636	0.764	0.934
$\epsilon \sim U(-\sqrt{3}, \sqrt{3})$	Model 0	0.042	0.034	0.048	0.052	0.072	0.042
	Model 1	0.210	0.384	0.534	0.648	0.780	0.948
	Model 2	0.364	0.664	0.834	0.918	0.954	1.000
	Model 3	0.476	0.732	0.874	0.954	0.980	1.000

By extending the comparison further, a simulation is conducted using the test proposed in Wang and Zhou (2006) method too. Figure 3.1 - Figure 3.8 show the curves of empirical sizes and powers of Minimum Distance(MD) test, Empirical Smoothing Lack-of-Fit(ES) test, and the test proposed by Wang & Zhou(WZ). Figure 3.1 and Figure 3.2 show that the empirical sizes with uniform errors are close to the nominal significance level of $\alpha = 0.05$ for all three tests. More concisely, the empirical sizes of Empirical Smoothing Lack-of-Fit test with uniform error and $h = n^{-1/3}$ are closer to the nominal levels than that of any other tests. Figure 3.3 - Figure 3.8 show the finite sample powers of the tests and it is clear that the Empirical Smoothing Lack-of-Fit test surpasses that of other two tests in almost all cases.

Another interesting finding from the simulation study is that the finite sample powers are not stable for different choices of bandwidths (different values of a) and for different error distributions (i.e. normal or uniform). Furthermore, we can see the bigger values of a, the larger the power. Note that the convolution of normal densities is still a normal density, so increasing the values of $a = 1$ to $a = \sqrt{2}$ is equivalent to replacing a standard normal kernel with the convolution of two standard normal kernels. According to the theory developed in the study, this will decrease the asymptotic variance of the test statistic and hence leads to a more powerful test.

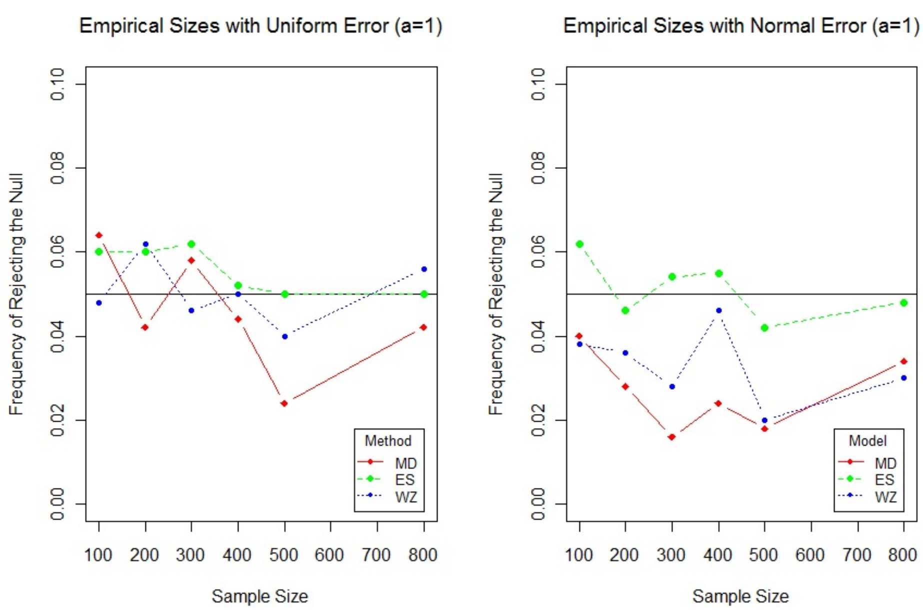

Figure 3.1: *Comparison of Empirical Sizes, Model 0 ($h = n^{-1/3}$)*

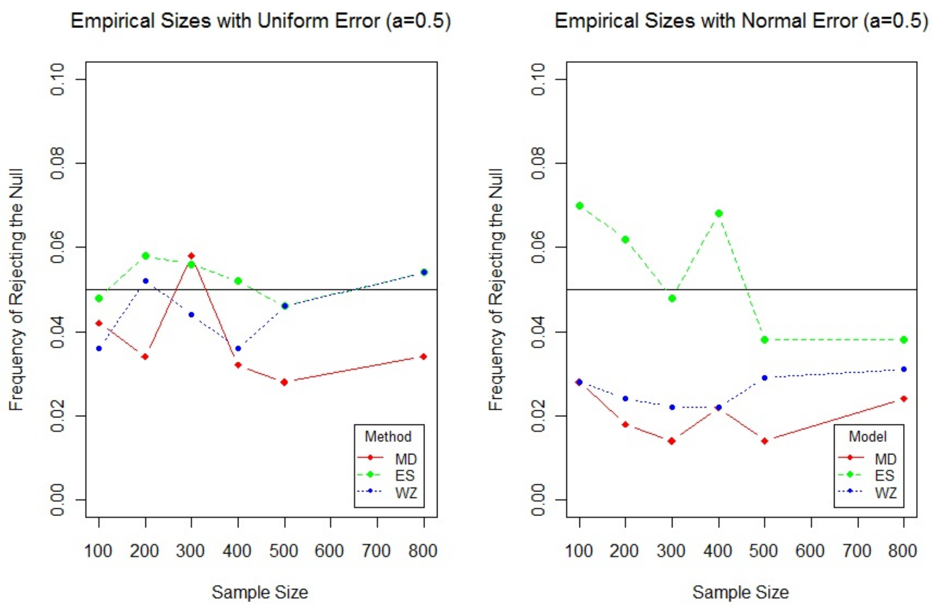

Figure 3.2: *Comparison of Empirical Sizes, Model 0* ($h = 0.5n^{-1/3}$)

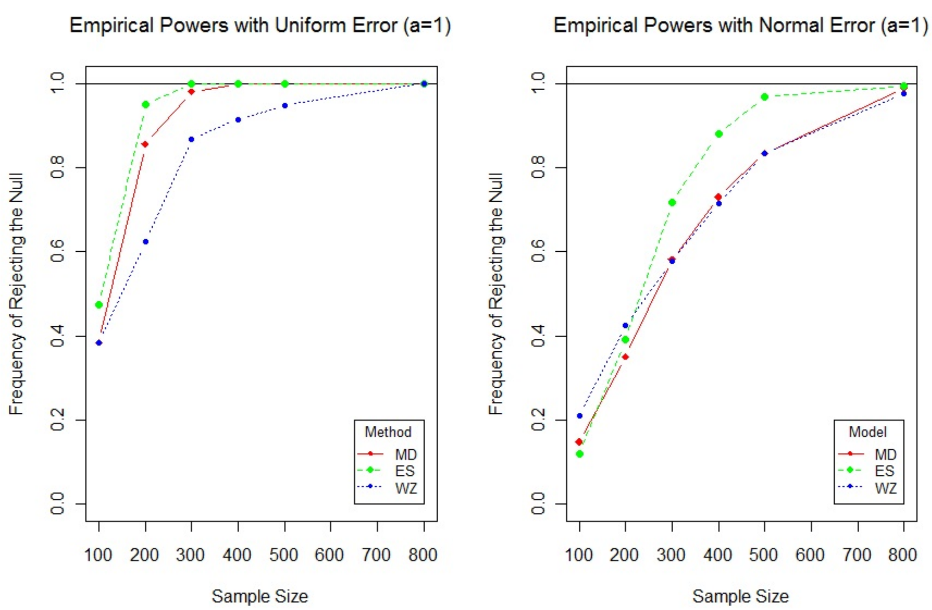

Figure 3.3: *Comparison of Empirical Powers, Model 1* ($h = n^{-1/3}$)

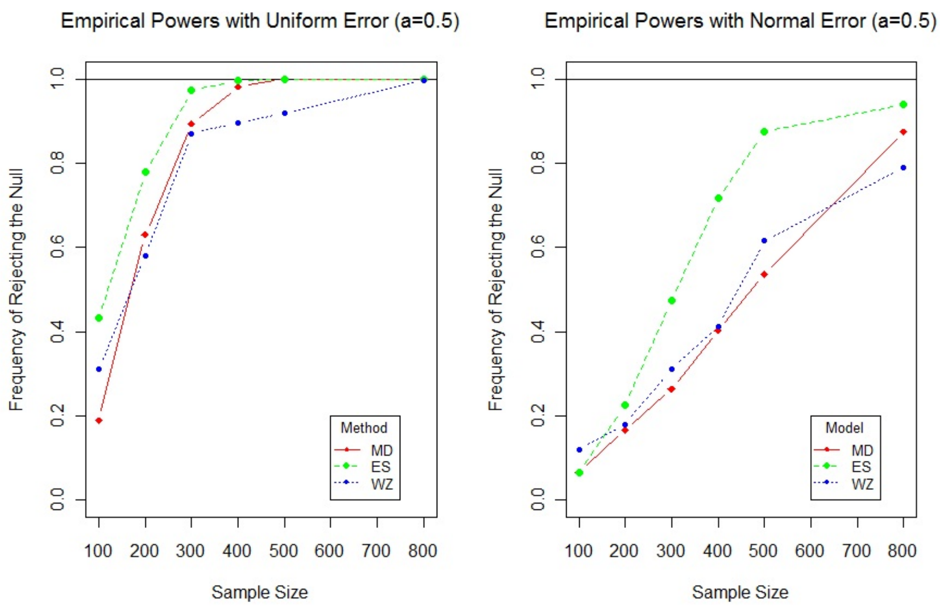

Figure 3.4: *Comparison of Empirical Powers, Model 1* ($h = 0.5n^{-1/3}$)

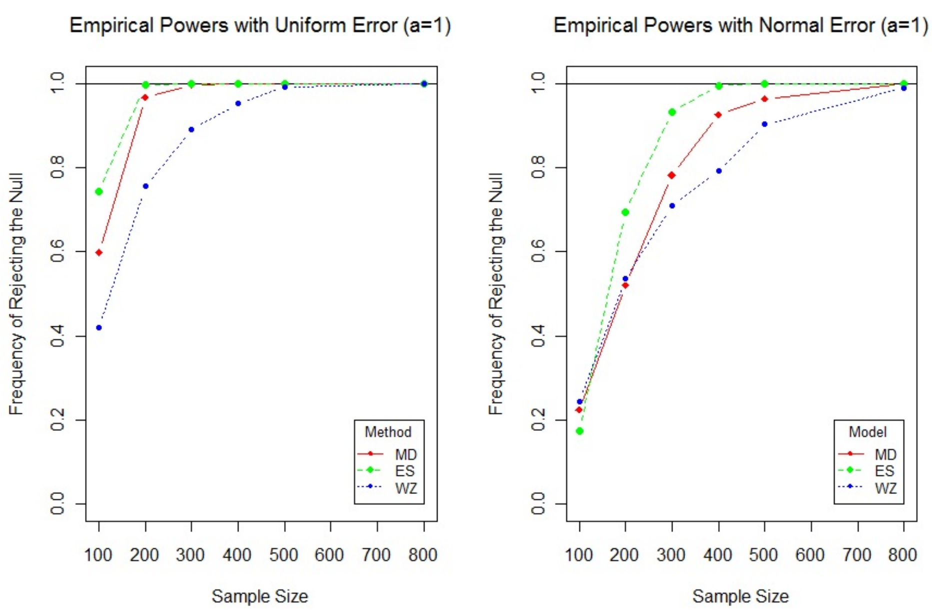

Figure 3.5: *Comparison of Empirical Powers, Model 2* ($h = n^{-1/3}$)

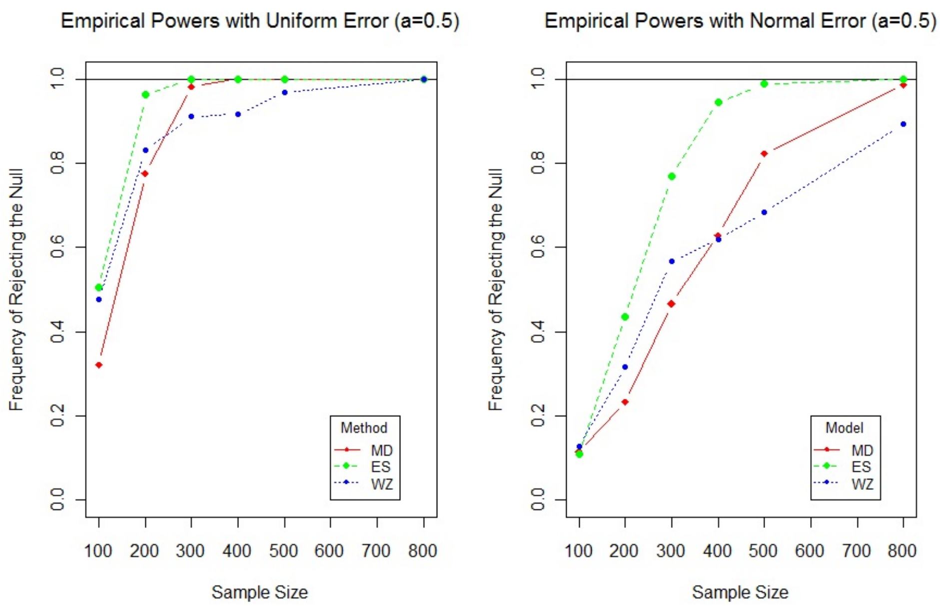

Figure 3.6: *Comparison of Empirical Powers, Model 2 ($h = 0.5n^{-1/3}$)*

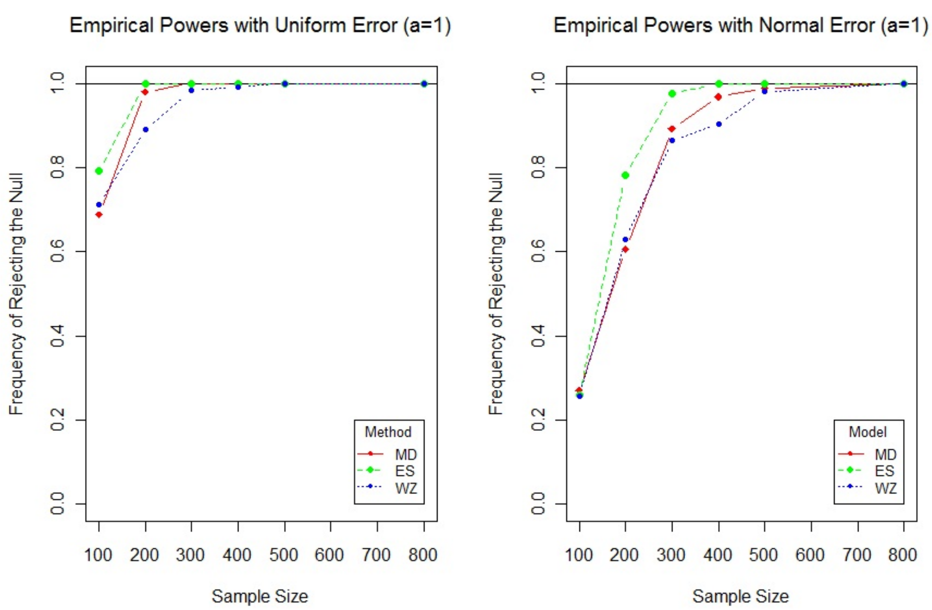

Figure 3.7: *Comparison of Empirical Powers, Model 3 ($h = n^{-1/3}$)*

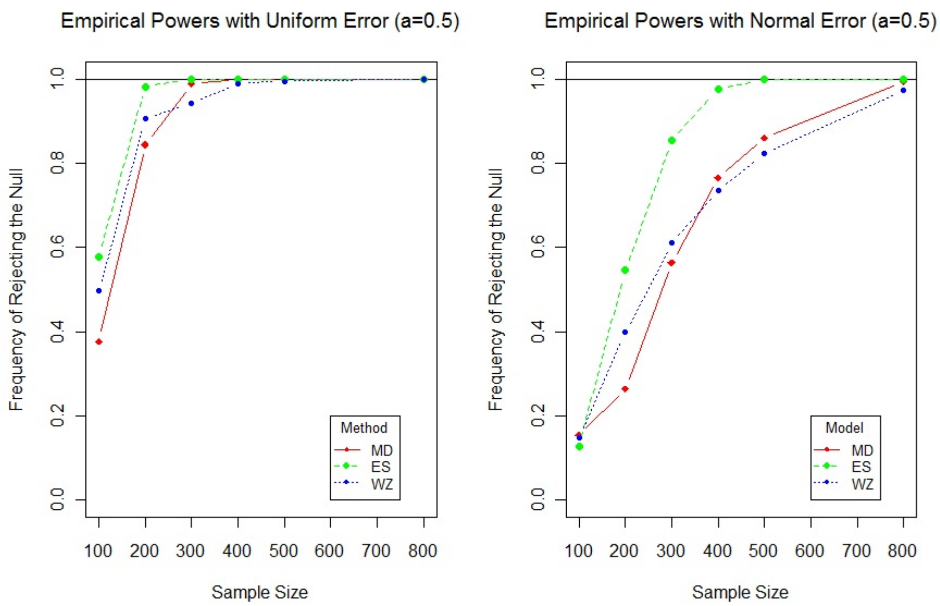

Figure 3.8: *Comparison of Empirical Powers, Model 3 ($h = 0.5n^{-1/3}$)*

3.6 Proof of the Main Results (Empirical Smoothing Test)

Proof of Theorem (3.3.1): Adding and subtracting $m(X_i;\beta_0)$ and $v(X_i;\beta_0,\theta_0)$ from e_i, e_i can be written as

$$\begin{aligned} e_i &= (Y_i - m(X_i;\beta_0) + m(X_i;\beta_0) - m(X_i;\hat{\beta}_n))^2 - v(X_i;\beta_0,\theta_0) + v(X_i;\beta_0,\theta_0) - v(X_i;\hat{\beta}_n,\hat{\theta}_n) \\ &= \xi_i + (\Delta m_i)^2 - 2\Delta m_i(Y_i - m(X_i;\beta_0)) - \Delta v_i, \end{aligned}$$

where $\Delta m_i = m(X_i;\hat{\beta}) - m(X_i;\beta_0)$ and $\Delta v_i = v(X_i;\hat{\beta}_n,\hat{\theta}_n) - v(X_i;\beta_0,\theta_0)$. Then Z_n can be further written as the sum of $Z_{n1}, Z_{n2}, \cdots, Z_{n10}$ where

$$Z_{nl} = \frac{1}{n(n-1)(n-2)h^{2d}} \sum_{i=1}^{n} \left[\sum_{j\neq i} \sum_{k\neq i,j} K_h(i,j)K_h(i,k) \right] \hat{f}^{-1}(X_i)P_{jk,l}$$

for $l = 1, 2, \cdots, 10$ with,

$$P_{jk,1} = \xi_j \xi_k, \quad P_{jk,2} = 2\xi_j (\Delta m_k)^2, \quad P_{jk,3} = -2\xi_j \Delta v_k, \quad P_{jk,4} = \Delta v_j \Delta v_k,$$
$$P_{jk,5} = -2\Delta v_j (\Delta m_k)^2, \quad P_{jk,6} = (\Delta m_j)^2 (\Delta m_k)^2, \quad P_{jk,7} = 4\Delta m_j \Delta v_k (Y_j - m(X_j; \beta_0)),$$
$$P_{jk,8} = -4\Delta m_j \xi_k (Y_j - m(X_j; \beta_0)), \quad P_{jk,9} = -4(\Delta m_k)^2 \Delta m_j (Y_j - m(X_j; \beta_0)),$$
$$P_{jk,10} = 4\Delta m_j \Delta m_k (Y_j - m(X_j, \beta_0))(Y_k - m(X_k; \beta_0)).$$

In the following, We use $\tilde{Z}_{nl}$ to denote Z_{nl} when $\hat{f}(X_i)$ is replaced by $f(X_i)$ for $l = 1, 2, \cdots, 10$.

Now let's consider

$$\begin{aligned} nh^{d/2}\tilde{Z}_{n1} &= \frac{1}{(n-1)(n-2)h^{3d/2}} \sum_{i=1}^{n} \left[\sum_{j \neq i} \sum_{k \neq i,j} K_h(i,j) K_h(i,k) \xi_j \xi_k \right] f^{-1}(X_i) \\ &= \frac{1}{(n-1)h^{3d/2}} \sum_{j=1}^{n} \sum_{k \neq j} \left[\frac{1}{(n-2)} \sum_{i \neq j,k} K_h(i,j) K_h(i,k) f^{-1}(X_i) \right] \xi_j \xi_k. \end{aligned}$$

By changing variable, we have

$$\begin{aligned} E\left[K_h(i,j)K_h(i,k)f^{-1}(X_i)|X_j, X_k\right] &= \int K\left(\frac{x - X_j}{h}\right) K\left(\frac{x - X_k}{h}\right) f^{-1}(x) f(x) dx \\ &= h^d \int K\left(u + \frac{X_j - X_k}{h}\right) K(u) du \\ &= h^d H_h(j,k), \end{aligned} \tag{3.22}$$

where $H_h(j,k) = H((X_j - X_k)/h)$. Notice that this H is the convolution of K. If K is a nonnegative, bounded, continuous, and symmetric density function, so is H.

Now we can write, $nh^{d/2}\tilde{Z}_{n1} = A_{n1} + A_{n2}$ where,

$$\begin{aligned}
A_{n1} =& \frac{1}{(n-1)h^{3d/2}} \sum_{j=1}^{n} \sum_{k \neq j} \left[\frac{1}{(n-2)} \sum_{i \neq j,k} K_h(i,j) K_h(i,k) f^{-1}(X_i) - h^d H_h(j,k) \right] \xi_j \xi_k, \\
A_{n2} =& \frac{1}{(n-1)h^{d/2}} \sum_{j=1}^{n} \sum_{k \neq j} H_h(j,k) \xi_j \xi_k.
\end{aligned}$$

Using the expectation-variance argument, we can show that A_{n1} is the order of $o_p(1)$. In fact, it is clear that $EA_{n1} = 0$ and next we'll consider the second moment of A_{n1}. Let

$$G_h(j,k) = \frac{1}{(n-2)} \sum_{i \neq j,k} K_h(i,j) K_h(i,k) f^{-1}(X_i) - h^d H_h(k,j).$$

Notice that $G_h(j,k) = G_h(k,j)$ and then A_{n1} can be rewritten as

$$A_{n1} = \frac{1}{(n-1)h^{3d/2}} \sum_{j=1}^{n} \sum_{k \neq j} G_h(j,k) \xi_j \xi_k.$$

The independence of ξ_j and ξ_k when $j \neq k$, and $E(\xi|X) = 0$ imply

$$\begin{aligned}
EA_{n1}^2 =& \frac{1}{(n-1)^2 h^{3d}} E \left(\sum_{j=1}^{n} \sum_{k \neq j} G_h(j,k) \xi_j \xi_k \right)^2 \\
=& \frac{1}{(n-1)^2 h^{3d}} E \left(\sum_{j=1}^{n} \sum_{k \neq j} G_h^2(j,k) \xi_j^2 \xi_k^2 + \sum_{j=1}^{n} \sum_{\substack{l \neq j \\ k \neq j}} G_h(j,k) G_n(j,l) \xi_j^2 \xi_k \xi_l \right) \\
=& \frac{2}{(n-1)^2 h^{3d}} \sum_{j=1}^{n} \sum_{k \neq j} EG_h^2(j,k) \xi_j^2 \xi_k^2 \\
=& \frac{2n}{(n-1)h^{3d}} EG_h^2(1,2) \tau^2(X_1) \tau^2(X_2),
\end{aligned}$$

where $\tau^2(x) = \tau^2(x; \beta_0, \theta_0)$ is as defined in (3.14). Conditioning on $(X_1, X_2), G_h(1,2)$ is a sum of i.i.d. centered random variables. Therefore the expectation of $G_h^2(1,2)\tau^2(X_1)\tau^2(X_2)$

equals

$$
\begin{aligned}
&E\left[\frac{1}{(n-2)}\sum_{i\neq 1,2} K_h(i,1)K_h(i,2)f^{-1}(X_i) - h^d H_h(1,2)\right]^2 \tau^2(X_1)\tau^2(X_2)\\
\leq&\frac{1}{(n-2)}E[K_h(3,1)K_h(3,2)f^{-1}(X_3) - h^d H_h(1,2)]^2\tau^2(X_1)\tau^2(X_2)\\
\leq&\frac{1}{(n-2)}EK_h^2(1,3)K_h^2(2,3)f^{-2}(X_3)\tau^2(X_1)\tau^2(X_2)\\
=&\frac{1}{n-2}\iiint K^2\left(\frac{x-y}{h}\right)K^2\left(\frac{x-z}{h}\right)\tau^2(y)\tau^2(z)f^{-1}(x)f(y)f(z)dxdydz\\
=&\frac{h^{2d}}{n-2}\iiint K^2(u)K^2(v)\tau^2(x-uh)\tau^2(x-vh)f^{-1}(x)f(x-uh)f(x-vh)dxdudv
\end{aligned}
$$

From the continuity and boundedness of K, τ^2, f and by (C1), (C3), (C6), we have

$$EG_h^2(1,2)\tau^2(X_1)\tau^2(X_2) = O(h^{2d}/(n-2)) \tag{3.23}$$

So $EA_{n1}^2 = o(1)$ from (C7). This implies

$$nh^{d/2}\tilde{Z}_{n1} = A_{n2} + o_p(1). \tag{3.24}$$

To show that $nh^{d/2}\tilde{Z}_{n2} = o_p(1)$, denote

$$d_{ni} = m(X_i;\hat{\beta}_n) - m(X_i;\beta_0) - (\hat{\beta}_n - \beta_0)'\dot{m}(X_i;\beta_0) = \Delta m_i - (\hat{\beta}_n - \beta_0)'\dot{m}(X_i;\beta_0)$$

By (C4), $\sup_{1\leq i\leq n}|d_{ni}| = O_p(1/n)$.

Using the notation d_{ni} and $a_n = 1/(n(n-1)(n-2)h^{2d})$, $\tilde{Z}_{n2}$ is the sum of $\tilde{Z}_{n21} + \tilde{Z}_{n22} + \tilde{Z}_{n23}$ where,

$$
\begin{aligned}
\tilde{Z}_{n21} =& 2a_n \sum_{i=1}^{n} \left[\sum_{j \neq k} \sum_{k \neq i,j} K_h(i,j) K_h(i,k) \xi_j d_{nk}^2 \right] f^{-1}(X_i) \\
\tilde{Z}_{n22} =& 4a_n \sum_{i=1}^{n} \left[\sum_{j \neq k} \sum_{k \neq i,j} K_h(i,j) K_h(i,k) \xi_j d_{nk} (\hat{\beta}_n - \beta_0)' \dot{m}(X_k; \beta_0) \right] f^{-1}(X_i) \\
\tilde{Z}_{n23} =& 2a_n \sum_{i=1}^{n} \left[\sum_{j \neq k} \sum_{k \neq i,j} K_h(i,j) K_h(i,k) \xi_j (\hat{\beta}_n - \beta_0)' \dot{m}(X_k; \beta_0) \dot{m}'(X_k; \beta_0) (\hat{\beta}_n - \beta_0) \right] f^{-1}(X_i)
\end{aligned}
$$

Notice that $|\tilde{Z}_{n21}|$ is bounded above by

$$
2 \sup_{1 \leq k \leq n} |d_{nk}|^2 \cdot a_n \sum_{i=1}^{n} \left[\sum_{j \neq k} \sum_{k \neq i,j} K_h(i,j) K_h(i,k) |\xi_j| \right] f^{-1}(X_i).
$$

The expectation of the second term is further bounded by

$$
\begin{aligned}
& \frac{1}{h^{2d}} E\left[K_h(1,2) K_h(1,3) E(|\xi_2||X_2) f^{-1}(X_1) \right] \\
& \leq \frac{1}{h^{2d}} \iint K\left(\frac{x-y}{h}\right) K\left(\frac{x-z}{h}\right) \tau(y) f(y) f(z) dx dy dz \\
& = \iint K(u) K(v) \tau(x - uh) f(x - vh) f(x - uh) dx du dv = O(1)
\end{aligned}
$$

Therefore,

$$
nh^{d/2} \tilde{Z}_{n21} = nh^{d/2} \cdot O_p(1/n^2) \cdot O_p(1) = o_p(1). \tag{3.25}
$$

Now $\tilde{Z}_{n22}$ can be written as the sum of $\tilde{Z}'_{n22}$ and $\tilde{Z}''_{n22}$, where

$$
\begin{aligned}
\tilde{Z}'_{n22} =& \frac{2(\hat{\beta}_n - \beta_0)'}{n(n-1)h^{2d}} \sum_{i=1}^{n} \sum_{j \neq k} \left[\frac{1}{n-2} \sum_{i \neq j,k} K_h(i,j) K_h(i,k) f^{-1}(X_i) - h^d H_h(j,k) \right] \xi_j d_{nk} \dot{m}(X_k; \beta_0), \\
\tilde{Z}''_{n22} =& \frac{2(\hat{\beta}_n - \beta_0)'}{(n-1)h^d} \sum_{j=1}^{n} \sum_{j \neq k} H_h(j,k) \xi_j d_{nk} \dot{m}(X_k; \beta_0).
\end{aligned}
$$

To show $\tilde{Z}'_{n22} = o_p(1)$, note that $|\tilde{Z}'_{n22}|$ is bounded above by

$$2 \sup_{1 \le k \le n} |d_{nk}| \frac{|\hat{\beta}_n - \beta_0|}{n(n-1)h^{2d}} \sum_{i=1}^{n} \sum_{j \ne k} \left[\frac{1}{n-2} \sum_{i \ne j,k} K_h(i,j) K_h(i,k) f^{-1}(X_i) - h^d H_h(j,k) \right] \xi_j \dot{m}(X_k; \beta_0).$$

Using the expectation-variance argument to the second term in the above expression, consider

$$
\begin{aligned}
& E\left[\frac{1}{(n-1)h^{2d}} \sum_{j \ne k} \left[\frac{1}{n-2} \sum_{i \ne j,k} K_h(i,j) K_h(i,k) f^{-1}(X_i) - h^d H_h(j,k) \right] \xi_j \dot{m}(X_k; \beta_0) \right]^2 \\
&= \frac{1}{(n-1)^2 h^{4d}} E\left[\sum_{k=2}^{n} \left(\frac{1}{n-2} \sum_{i \ne 1,k} K_h(i,1) K_h(i,k) f^{-1}(X_i) - h^d H_h(1,k) \right) \dot{m}(X_k; \beta_0) \right]^2 \tau^2(X_1) \\
&= \frac{1}{(n-1)^2 h^{4d}} \sum_{k=2}^{n} E\left(\frac{1}{n-2} \sum_{i \ne 1,k} K_h(i,1) K_h(i,k) f^{-1}(X_i) - h^d H_h(1,k) \right)^2 \dot{m}(X_k; \beta_0) \dot{m}'(X_k; \beta_0) \tau^2(X_1) \\
&+ \frac{1}{(n-1)^2 h^{4d}} \sum_{k_1 \ne k_2}^{n} E\left(\frac{1}{n-2} \sum_{i \ne 1,k_1} K_h(i,1) K_h(i,k_1) f^{-1}(X_i) - h^d H_h(1,k_1) \right) \dot{m}(X_{k_1}; \beta_0) \\
&\qquad \left(\frac{1}{n-2} \sum_{i \ne 1,k_2} K_h(i,1) K_h(i,k_2) f^{-1}(X_i) - h^d H_h(1,k_2) \right) \dot{m}(X_{k_2}; \beta_0) \tau^2(X_1) \\
&= \frac{(n-1)}{(n-1)^2 h^{4d}} E\left(\frac{1}{n-2} \sum_{i \ne 1,2} K_h(i,1) K_h(i,2) f^{-1}(X_i) - h^d H_h(1,2) \right)^2 \dot{m}(X_k; \beta_0) \dot{m}'(X_k; \beta_0) \tau^2(X_1) \\
&+ \frac{(n-1)(n-2)}{(n-1)^2 h^{4d}} E\left(\frac{1}{n-2} \sum_{i \ne 1,2} K_h(i,1) K_h(i,2) f^{-1}(X_i) - h^d H_h(1,2) \right) \dot{m}(X_2; \beta_0) \\
&\qquad \left(\frac{1}{n-2} \sum_{i \ne 1,3} K_h(i,1) K_h(i,3) f^{-1}(X_i) - h^d H_h(1,3) \right) \dot{m}(X_3; \beta_0) \tau^2(X_1).
\end{aligned}
$$

With the results of (3.23), the boundedness of $\|\dot{m}(x; \beta_0)\|$ and the assumptions of $(C3)$, the expectation of the first term on the right hand side is bounded above by $O(h^{2d}/(n-2))$. By a trivial argument, the second term on the right is $O_p(1/nh^{d/2})$.

Since $\hat{\beta}_n$ is a $\sqrt{n}$-consistent estimator and by (C4),

$$nh^{d/2}\tilde{Z}'_{n22} = nh^{d/2}.O_p(1/\sqrt{n}).O_p(1/\sqrt{n}).O(1/nh^{d/2}) = o_p(1).$$

Now let's consider $\tilde{Z}''_{n22}$.

According to the Lemma 3.3b in Zheng (1996),

$$\frac{1}{n(n-1)h^d}\sum_{j=1}^{n}\sum_{j\neq k} H_h(k,j)\xi_j \dot{m}(X_k;\beta_0) = O_p(1/\sqrt{n}).$$

So, one can show that $nh^{d/2}\tilde{Z}''_{n22} = nh^{d/2}.O_p(1/\sqrt{n}).O_p(1/\sqrt{n}).O_p(1/\sqrt{n}) = o_p(1)$. This implies that

$$nh^{d/2}\tilde{Z}_{n22} = o_p(1) \tag{3.26}$$

Finally we show that $nh^{d/2}\tilde{Z}_{n23} = o_p(1)$. For simplicity, we only prove the result for $p = 1$. The general case can be argued element wise. It is easily see that $\tilde{Z}_{n23}$ is bounded above by

$$\frac{|\hat{\beta}_n - \beta_0|^2}{n(n-1)(n-2)h^{2d}} \cdot \sum_{i=1}^{n}\sum_{j\neq k}\sum_{k\neq i,j} K_h(i,j)K_h(i,k)|\xi_j||\dot{m}(X_k;\beta_0)\dot{m}^{'}(X_k;\beta_0)|f^{-1}(X_i)$$

The expectation of the second term is further bounded above by

$$n(n-1)(n-2)E\left[K_h(1,2)K_h(1,3)E(|\xi_2||X_2)f^{-1}(X_1)\dot{m}(X_k;\beta_0)\dot{m}^{'}(X_k;\beta_0)\right]$$

By the boundedness of $\|\dot{m}(x;\beta_0)\|$, $\sqrt{n}-$consistency of $\hat{\beta}_n$, and the conditions imposed in the previous derivation, we can show that

$$nh^{d/2}\tilde{Z}_{n23} = nh^{d/2}.O_p(1/n).O(1) = o_p(1) \tag{3.27}$$

Using the results of (3.25), (3.26), and (3.27), it is clear that

$$nh^{d/2}\tilde{Z}_{n2} = o_p(1). \tag{3.28}$$

Next, let's show that $nh^{d/2}\tilde{Z}_{n3} = o_p(1)$. Adding and subtracting $(\hat{\beta}_n - \beta_0)'\dot{v}_\beta(X_k;\beta_0,\theta_0)$ and $(\hat{\theta}_n - \theta_0)'\dot{v}_\theta(X_k;\beta_0,\theta_0)$ from $\Delta v_k = v(X_k;\hat{\beta}_n,\hat{\theta}_n) - v(X_k;\beta_0,\theta_0)$ and denoting

$$u_{nk} = \Delta v_k - (\hat{\beta}_n - \beta_0)'\dot{v}_\beta(X_k;\beta_0,\theta_0) - (\hat{\theta}_n - \theta_0)'\dot{v}_\theta(X_k;\beta_0,\theta_0),$$

$nh^{d/2}\tilde{Z}_{n3}$ can be written as the sum of the following three terms.

$$B_{n1} = -\frac{2(\hat{\beta}_n - \beta_0)'}{(n-1)(n-2)h^{3d/2}}\sum_{i=1}^{n}\left[\sum_{j\neq k}\sum_{k\neq i,j} K_h(i,j)K_h(i,k)\right] f^{-1}(X_i)\xi_j\dot{v}_\beta(X_k;\beta_0,\theta_0),$$

$$B_{n2} = -\frac{2(\hat{\theta}_n - \theta_0)'}{(n-1)(n-2)h^{3d/2}}\sum_{i=1}^{n}\left[\sum_{j\neq k}\sum_{k\neq i,j} K_h(i,j)K_h(i,k)\right] f^{-1}(X_i)\xi_j\dot{v}_\theta(X_k;\beta_0,\theta_0),$$

$$B_{n3} = -\frac{2}{(n-1)(n-2)h^{3d/2}}\sum_{i=1}^{n}\left[\sum_{j\neq k}\sum_{k\neq i,j} K_h(i,j)K_h(i,k)\right] f^{-1}(X_i)\xi_j u_{nk}.$$

By adding and subtracting $h^dH_h(j,k)$, B_{n1} can be written as the sum of $B_{n11} + B_{n12}$, where

$$B_{n11} = -\frac{2(\hat{\beta}_n - \beta_0)'}{(n-1)h^{3d/2}}\sum_{j\neq k}\left[\frac{1}{(n-2)}\sum_{i\neq j,k} K_h(i,j)K_h(i,k)f^{-1}(X_i) - h^dH_h(j,k)\right]\xi_j\dot{v}_\beta(X_k;\beta_0,\theta_0),$$

$$B_{n12} = -\frac{2(\hat{\beta}_n - \beta_0)'}{(n-1)h^{d/2}}\sum_{j\neq k} H_h(j,k)\xi_j\dot{v}_\beta(X_k;\beta_0,\theta_0).$$

Let $\dot{v}_{\beta l}(X_k;\beta_0,\theta_0)$ denote the $l-$th element of the $p\times 1$ vector $\dot{v}_\beta(X_k;\beta_0,\theta_0)$. Note that

$$E\left[\frac{1}{(n-1)h^{3d/2}}\sum_{j\neq k}\left[\frac{1}{(n-2)}\sum_{i\neq j,k}K_h(i,j)K_h(i,k)f^{-1}(X_i)-h^dH_h(j,k)\right]\xi_j\dot{v}_{\beta l}(X_k;\beta_0,\theta_0)\right]^2$$

$$=\frac{n}{(n-1)^2h^{3d}}E\left[\sum_{k=2}^{n}\left(\frac{1}{(n-2)}\sum_{i\neq 1,k}K_h(i,1)K_h(i,k)f^{-1}(X_i)-h^dH_h(1,k)\right)\dot{v}_{\beta l}(X_k;\beta_0,\theta_0)\right]^2\tau^2(X_1)$$

$$=\frac{n(n-1)}{(n-1)^2h^{3d}}E\left(\frac{1}{(n-2)}\sum_{i\neq 1,2}K_h(i,1)K_h(i,2)f^{-1}(X_i)-h^dH_h(1,2)\right)^2\dot{v}^2_{\beta l}(X_2;\beta_0,\theta_0)\tau^2(X_1)$$

$$+\frac{n(n-1)(n-2)}{(n-1)^2h^{3d}}E\left(\frac{1}{(n-2)}\sum_{i\neq 1,2}K_h(i,1)K_h(i,2)f^{-1}(X_i)-h^dH_h(1,2)\right)\dot{v}_{\beta l}(X_2;\beta_0,\theta_0)$$

$$\left(\frac{1}{(n-2)}\sum_{i\neq 1,3}K_h(i,1)K_h(i,3)f^{-1}(X_i)-h^dH_h(1,3)\right)\dot{v}_{\beta l}(X_3;\beta_0,\theta_0)\tau^2(X_1).$$

With the assumption of the continuity of $\dot{v}(X;\beta_0,\theta_0)$, $\tau^2(X)$ and using the result of (3.23), one can show that the expectation in the first term on the right is $O(h^{2d}/(n-2))$. Therefore the first term on the right hand side is $O_p(1/nh^d)$ which is $o_p(1)$. By a lengthy but trivial argument, one can show that the second term on the right is $O_p(1)$. Using these results and the $\sqrt{n}-$consistency of $\hat{\beta}_n$, $B_{n11}=O_p(1/\sqrt{n})O_p(1)=o_p(1)$.

Next, we'll consider B_{n12}. According to the Lemma 3.3b in Zheng (1996),

$$\frac{1}{n(n-1)h^d}\sum_{j=1}^{n}\sum_{j\neq k}H_h(k,j)\xi_j\dot{v}_\beta(X_k;\beta_0,\theta_0)=O_p(1/\sqrt{n}).$$

Therefore $B_{n12}=nh^{d/2}O_p(1/\sqrt{n})O_p(1/\sqrt{n})=o_p(1)$. Hence B_{n1} is $o_p(1)$.

Using the similar arguments used in showing $B_{n1}=o_p(1)$, the $\sqrt{n}-$consistency of $\hat{\theta}_n$, and the continuity of $\dot{v}_\theta(x;\beta_0,\theta_0)$, we can show that $B_{n2}=o_p(1)$.

To show that $B_{n3} = o_p(1)$, note that $|B_{n3}|$ is bounded above by

$$\sup_{1\le k\le n} |u_{nk}| \frac{2}{(n-1)h^{3d/2}} \sum_{j=1}^{n} \sum_{j\ne k} \left[\frac{1}{((n-2))} \sum_{i\ne j,k} K_h(i,j)K_h(i,k)f^{-1}(X_i)|\xi_j| \right].$$

It can be shown that the expectation of the second term is

$$\frac{2n(n-1)}{(n-1)h^{3d/2}} E\left[\frac{K_h(1,2)K_h(1,3)}{f(X_1)} E(|\xi_2||X_2) \right] = O(nh^{d/2}).$$

By (C5) and $\sup_{1\le k\le n} |u_{nk}| = O_p(1/n)$

$$B_{n3} = O_p(1/n)O(nh^{d/2}) = o_p(1).$$

Hence

$$nh^{d/2}\tilde{Z}_{n3} = o_p(1). \tag{3.29}$$

Using similar arguments, one can show that

$$nh^{d/2}\tilde{Z}_{nl} = o_p(1) \quad \text{for all } l = 4, 5, \cdots, 10. \tag{3.30}$$

Note that the above results are obtained by replacing $\hat{f}(x)$ with $f(x)$. Next we will consider this modification. We denote

$$\begin{aligned} C_n &= \frac{nh^{d/2}}{n(n-1)(n-2)h^{2d}} \sum_{i\ne j,k} K_h(i,j)K_h(i,k)f^{-1}(X_i) \left[\frac{f(X_i)}{\hat{f}(X_i)} - 1 \right] \xi_j\xi_k \\ &= \frac{h^{d/2}}{(n-1)} \sum_{j\ne k} M_n(X_j, X_k)\xi_j\xi_k, \end{aligned}$$

where

$$M_n(X_j, X_k) = \frac{1}{(n-2)h^{2d}} \sum_{i \neq j,k} K_h(i,j) K_h(i,k) f^{-1}(X_i) \left[\frac{f(X_i)}{\hat{f}(X_i)} - 1 \right].$$

The symmetry of $M_n(X_j, X_k)$ and the assumptions of error terms imply

$$EC_n^2 = \frac{4h^d}{(n-1)^2} \sum_{j \neq k} EM_n^2(X_j, X_k) \tau^2(X_j) \tau^2(X_k) = O(h^d) EM_n^2(X_1, X_2) \tau^2(X_1) \tau^2(X_2).$$

Note that

$$\begin{aligned}
& EM_n^2(X_1, X_2)\tau^2(X_1)\tau^2(X_2) \\
&= E\left[\frac{1}{(n-2)h^{2d}} \sum_{i=3}^{n} K_h(i,1) K_h(i,2) f^{-1}(X_i) \left[\frac{f(X_i)}{\hat{f}(X_i)} - 1 \right] \right]^2 \tau^2(X_1)\tau^2(X_2) \\
&\leq \frac{1}{(n-2)h^{4d}} \sum_{i=3}^{n} EK_h^2(i,1) K_h^2(i,2) f^{-2}(X_i) \left[\frac{f(X_i)}{\hat{f}(X_i)} - 1 \right]^2 \tau^2(X_1)\tau^2(X_2) \\
&= \frac{1}{h^{4d}} EK_h^2(3,1) K_h^2(3,2) f^{-2}(X_3) \left[\frac{f(X_3)}{\hat{f}(X_3)} - 1 \right]^2 \tau^2(X_1)\tau^2(X_2). \qquad (3.31)
\end{aligned}$$

From the condition (C1), we can see that the last expectation in (3.31) has the same order as

$$\frac{1}{h^{4d}} EK_h^2(3,1) K_h^2(3,2) f^{-2}(X_3) \left[f(X_3) - \hat{f}(X_3) \right]^2 \tau^2(X_1)\tau^2(X_2) \qquad (3.32)$$

Notice that

$$\begin{aligned}
\hat{f}(X_3) - f(X_3) &= \frac{1}{nh^d} \sum_{j=4}^{n} K_h(j,3) - \frac{1}{h^d} E\left[K_h(4,3) | X_3 \right] \\
&+ \frac{1}{h^d} E\left[K_h(4,3) | X_3 \right] - f(X_3) \\
&+ \frac{1}{nh^d} K_h(1,3) + \frac{1}{nh^d} K_h(2,3) + \frac{1}{nh^d} K_h(1,1).
\end{aligned}$$

Using the previous results, we can show that (3.32) is bounded above by the sum of the following three terms:

$$3E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}\left[\frac{1}{nh^d}\sum_{j=4}^{n}K_h(j,3)-\frac{1}{h^d}E\left[K_h(4,3)|X_3\right]\right]^2\tau^2(X_1)\tau^2(X_2)$$
$$3E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}\left[\frac{1}{h^d}E\left[K_h(4,3)|X_3\right]-f(X_3)\right]^2\tau^2(X_1)\tau^2(X_2),$$
$$3E\frac{K_h^2(3,1)K_h^2(3,2)}{h^{4d}f^2(X_3)}\left[\frac{1}{nh^d}K_h(1,3)+\frac{1}{nh^d}K_h(2,3)+\frac{1}{nh^d}K_h(1,1)\right]^2\tau^2(X_1)\tau^2(X_2).$$

By changing variables when calculating the above expectations, one can show that the first term has the order of $O(1/nh^{3d})$, the second term has the order of $O(1)$, and the third one has the order of $O(1/n^2h^{4d})$. Therefore,

$$C_n^2 = O_p(1/nh^{2d}) + O_p(h^d) + O_p(1/n^2h^{3d})$$

By condition (C7) together with the above results, $C_n = o_p(1)$.

Now we will consider,

$$D_n = \frac{nh^{d/2}}{n(n-1)(n-2)h^{2d}}\sum_{i\neq j,k}K_h(i,j)K_h(i,k)f^{-1}(X_i)\left[\frac{f(X_i)}{\hat{f}(X_i)}-1\right]\xi_j(\Delta m_k)^2$$

Using the same definition of $M_n(X_j, X_k)$, we can write

$$D_n = \frac{h^{d/2}}{(n-1)}\sum_{j\neq k}M_n(X_j,X_k)\xi_j(\Delta m_k)^2.$$

Note that $\Delta m_k = d_{nk}+(\hat{\beta}_n-\beta_0)'\dot{m}(X_k;\beta_0)$. D_n can be written as the sum of $D_{n1}+D_{n2}+D_{n3}$

where,

$$D_{n1} = \frac{2h^{d/2}}{(n-1)} \sum_{j \neq k} M_n(X_j, X_k)\xi_j d_{nk}^2,$$

$$D_{n2} = 2\frac{4h^{d/2}}{(n-1)} \sum_{j \neq k} M_n(X_j, X_k)\xi_j d_{nk}(\hat{\beta}_n - \beta_0)^{'} \dot{m}(X_k; \beta_0),$$

$$D_{n3} = \frac{2h^{d/2}}{(n-1)} \sum_{j \neq k} M_n(X_j, X_k)\xi_j (\hat{\beta}_n - \beta_0)^{'} \dot{m}(X_k; \beta_0).\dot{m}^{'}(X_k; \beta_0)(\hat{\beta}_n - \beta_0).$$

By the symmetry of $M_n(X_j, X_k)$ in its arguments, the result of (3.31), and $(C3)$ imply

$$\begin{aligned}
ED_{n1}^2 &= \frac{h^d}{(n-1)^2} \sum_{j \neq k} EM_n^2(X_j, X_k)E(\xi_i^2|X_j)(d_{nk}^2)^2 \\
&\leq O(h^d) \left[\sup_{1 \leq k \leq n} |d_{nk}|^2 \right]^2 EM_n^2(X_j, X_k)E(\xi_i^2|X_j) \\
&\leq O(h^2) \left[\sup_{1 \leq k \leq n} |d_{nk}|^2 \right]^2 \left[O(1/nh^{3d}) + O(1) + O(1/n^2h^{4d}) \right] \\
&= O(h^d) \cdot O(1/n^2) \left[O(1/nh^{3d}) + O(1) + O(1/n^2h^{4d}) \right]
\end{aligned}$$

Hence the condition $(C7)$ and the above results imply $D_{n1} = o_p(1)$.

Using the similar arguments in proving $nh^{d/2}\tilde{Z}_{n22} = o_p(1), nh^{d/2}\tilde{Z}_{n23} = o_p(1)$, $\sqrt{n}-$consistency of $\hat{\beta}_n$, and the boundedness of $\|\dot{m}(x; \beta_0)\|$, we can show that $D_{n2} = o_p(1)$ and $D_{n3} = o_p(1)$. Therefore $D_n = o_p(1)$. By continuing this way, one can deal with all other modified terms.

By (3.24), (3.28), (3.29), and (3.30)

$$nh^{d/2} Z_n = \frac{1}{(n-1)h^{d/2}} \sum_{j=1}^{n} \sum_{k \neq j} H_h(j, k)\xi_j\xi_k + o_p(1). \tag{3.33}$$

From Lemma 3.3a in Zheng (1996) and from Theorem 2.1 in Song and Du (2011),

$$nh^{d/2}Z_n \Rightarrow N(0, \sigma^2),$$

where

$$\sigma^2 = 2\int \left[\int K(u+v)K(v)dv\right]^2 du. \int [\tau^2(x)]^2 f^2(x)dx, \tag{3.34}$$

with $\tau^2(x) = E(\xi^2|X = x)$.

■

Proof of Theorem 3.3.2:

Under H_a, we write $Y_i^a = m(X_i; \beta_a) + \sqrt{v_a(X_i)}\epsilon_i$ and $Y_i = m(X_i; \beta_0) + \sqrt{v_1(X_i)}\epsilon_i$. Define $m_0(x) = m(x; \beta_0), m_a(x) = m(x; \beta_a), v_a(x) = v(x; \beta_a, \theta_a)$ and $K_{ij} = K_h(i, j)$.

The test statistic can be written as

$$Z_n = a_n \sum_{i \neq j \neq k} K_{ij}K_{ik}e_je_k\hat{f}^{-1}(x_i),$$

where $a_n = \frac{1}{n(n-1)(n-2)h^{2d}}$ and $e_i = (Y_i - m(X_i; \hat{\beta}))^2 - v(X_i; \hat{\beta}_n, \hat{\theta}_n)$. By adding and subtracting Y_i^a from Y_i in the test statistic, we can write it as the sum of the following six terms:

$$
\begin{aligned}
U_{n1} =& a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}(y_j - y_j^a)^2(y_k - y_k^a)^2 \hat{f}^{-1}(x_i),\\
U_{n2} =& 4a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}(y_j - y_j^a)^2(y_k - y_k^a)(y_k^a - m(X_k;\hat{\beta}_n))\hat{f}^{-1}(x_i),\\
U_{n3} =& 2a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}(y_j - y_j^a)^2 \left[(y_k^a - m(X_k;\hat{\beta}))^2 - v(X_k;\hat{\beta}_n,\hat{\theta}_n)\right] \hat{f}^{-1}(x_i),\\
U_{n4} =& 4a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}(y_j - y_j^a)(y_j^a - m(X_j;\hat{\beta}_n))(y_k - y_k^a)(y_k^a - m(X_k;\hat{\beta}_n))\hat{f}^{-1}(x_i),\\
U_{n5} =& 4a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}(y_j - y_j^a)(y_j^a - m(X_j;\hat{\beta}_n)) \left[(y_k^a - m(X_k;\hat{\beta}_n))^2 - v(X_k;\hat{\beta}_n,\hat{\theta}_n)\right] \hat{f}^{-1}(x_i),\\
U_{n6} =& a_n \sum_{i\neq j\neq k} K_{ij}K_{ik} \left[(y_j^a - m(X_j;\hat{\beta}_n))^2 - v(X_j;\hat{\beta}_n,\hat{\theta}_n)\right] \left[(y_k^a - m(X_k;\hat{\beta}_n))^2 - v(X_k;\hat{\beta}_n,\hat{\theta}_n)\right] \hat{f}^{-1}(x_i).
\end{aligned}
$$

Since $Y_i - Y_i^a = m(X_i;\beta_0) - m(X_i;\beta_a) + (\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})\epsilon_i$, and taking $\Delta m_i = m(X_i;\beta_0) - m(X_i;\beta_a)$, $\Delta v_i = (\sqrt{v_1(X_i)} - \sqrt{v_a(X_i)})$, U_{n1} can be written as the following six terms:

$$
\begin{aligned}
U_{n11} =& a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\Delta^2 m_j \Delta^2 m_k \hat{f}^{-1}(x_i),\\
U_{n12} =& 4a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\Delta^2 m_j \Delta m_k \Delta v_k \epsilon_k \hat{f}^{-1}(x_i),\\
U_{n13} =& 2a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\Delta^2 m_j \Delta^2 v_k \epsilon_k^2 \hat{f}^{-1}(x_i),\\
U_{n14} =& 4a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\Delta m_j \Delta m_k \Delta v_j \Delta v_k \epsilon_j \epsilon_k \hat{f}^{-1}(x_i),
\end{aligned}
$$

$$U_{n15} = 4a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\Delta m_j \Delta v_j \Delta^2 v_k \epsilon_j \epsilon_k^2 \hat{f}^{-1}(x_i),$$
$$U_{n16} = a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\Delta^2 v_j \Delta^2 v_k \epsilon_j^2 \epsilon_k^2 \hat{f}^{-1}(x_i).$$

With the assumption on ϵ, one can show that $U_{n11} \to E[(m(X;\beta_0)-m(X;\beta_a))^4 f(X)], U_{n13} \to 2E[(m(X;\beta_0)-m(X;\beta_a))^2(\sqrt{v_1(X)}-\sqrt{v_a(X)})^2 f(X)], U_{n16} \to E[(\sqrt{v_1(X)}-\sqrt{v_a(X)})^4 f(X)]$, and all other terms are of order $o_p(1)$.
Hence, $U_{n1} \to E\left[(m(X;\beta_0)-m(X;\beta_a))^2 + (\sqrt{v_1(X)}-\sqrt{v_a(X)})^2\right]^2 f(X)$.
Using the $\sqrt{n}-$ consistency of $\hat{\beta}_n$ under H_a and (C4), we can write U_{n2} as the sum of the following two terms and remainders of order $o_p(1)$:

$$U_{n21} = a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\Delta^2 m_j \Delta v_k \sqrt{v_a(X_i)}\epsilon_k^2 \hat{f}^{-1}(x_i),$$
$$U_{n22} = a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\Delta^2 v_j \Delta v_k \sqrt{v_a(X_i)}\epsilon_j^2 \epsilon_k^2 \hat{f}^{-1}(x_i).$$

Using the same arguments as in the previous, one can show that
$U_{n2} \to 4E\left[(m(X;\beta_0)-m(X;\beta_a))^2(\sqrt{v_1(X)}-\sqrt{v_a(X)}) + (\sqrt{v_1(X)}-\sqrt{v_a(X)})^3\right]\sqrt{v_a(X)}f(X)$
in probability.

Using the condition (C4) and the $\sqrt{n}-$ consistency of $\hat{\beta}_n$ to β_a, U_{n4} can be written as the sum of the following term and a remainder of order $o_p(1)$:

$$U_{n41} = a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\Delta v_j \Delta v_k \sqrt{v_a(X_j)}\sqrt{v_a(X_k)}\epsilon_j^2 \epsilon_k^2 \hat{f}^{-1}(x_i).$$

Again using the same arguments, $U_{n4} \to 4E\left[(\sqrt{v_1(X)} - \sqrt{v_a(X)})^2 v_a(X)\right] f(X)$. Using the same methods used in null case, one can show that $U_{n3} = U_{n5} = U_{n6} = o_p(1)$. After doing some algebraic manipulations,we can show that

$$Z_n \to E\left[(m(X;\beta_0) - m(X;\beta_a))^2 + (v_1(X) - v_a(X))\right]^2 f(X)$$

in probability.

Finally, similar to the Lemma 3.4 in Zheng (1996) and from Theorem 3.1 in Song and Du (2011), we have

$$\hat{\sigma}^2 \to \int \left[\int K(u+v)K(v)dv\right]^2 du.$$
$$\int [\tau^2(x) + (m(X;\beta_0) - m(X;\beta_a))^2 + (v_1(X) - v(X;\beta_a,\theta_a))]^2 f^2(x)dx,$$

in probability.

■

Proof of Theorem (3.3.3): Under the local alternative,

$$H_{Loc} : v(x) = v(x;\beta_0,\theta_0) + c_n\delta(x),\ \forall x \in \mathbb{R}^d,$$

we write $Y_i^L = m(X_i;\beta_0) + \sqrt{v(X_i)}\epsilon_i$ and $Y_i = m(X_i;\beta_0) + \sqrt{v(X_i) + c_n\delta(X_i)}\epsilon_i$. Define $v(x) = v(x;\beta_0,\theta_0)$ and $K_{ij} = K_h(i,j)$.

The test statistic can be written as

$$Z_n = a_n \sum_{i \neq j \neq k} K_{ij}K_{ik}e_je_k\hat{f}^{-1}(X_i),$$

where $a_n = \frac{1}{n(n-1)(n-2)h^{2d}}$ and $e_i = (Y_i - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \hat{\theta}_n)$. By adding and subtracting Y_i^L from Y_i in the test statistic, we can write it as the sum of the following six terms:

$$
\begin{aligned}
W_{n1} =& a_n \sum_{i \neq j \neq k} K_{ij} K_{ik} \xi_j \xi_k \hat{f}^{-1}(X_i), \\
W_{n2} =& 2a_n \sum_{i \neq j \neq k} K_{ij} K_{ik} (Y_j - Y_j^L)^2 \xi_k \hat{f}^{-1}(X_i), \\
W_{n3} =& a_n \sum_{i \neq j \neq k} K_{ij} K_{ik} (Y_j - Y_j^L)^2 (Y_k - Y_k^L)^2 \hat{f}^{-1}(X_i), \\
W_{n4} =& 4a_n \sum_{i \neq j \neq k} K_{ij} K_{ik} (Y_j - Y_j^L)(Y_j^L - m(X_j; \hat{\beta}_n)) \xi_k \hat{f}^{-1}(X_i), \\
W_{n5} =& 4a_n \sum_{i \neq j \neq k} K_{ij} K_{ik} (Y_j - Y_j^L)^2 (Y_k - Y_k^L)(Y_k^L - m(X_k; \hat{\beta}_n)) \hat{f}^{-1}(X_i), \\
W_{n6} =& 4a_n \sum_{i \neq j \neq k} K_{ij} K_{ik} (Y_j - Y_j^L)(Y_j^L - m(X_j; \hat{\beta}_n))(Y_k - Y_k^L)(Y_k^L - m(X_k; \hat{\beta}_n)) \hat{f}^{-1}(X_i),
\end{aligned}
$$

where $\xi_i = (Y_i^L - m(X_i; \hat{\beta}_n))^2 - v(X_i; \hat{\beta}_n, \hat{\theta}_n)$.

Similar to the proof of the null case,

$$
\begin{aligned}
nh^{d/2} W_{n1} =& \frac{1}{(n-1)h^{d/2}} \sum_{j=1}^{n} \sum_{k \neq j} \frac{1}{h^d} \frac{1}{(n-2)} \sum_{i \neq j \neq k} K_h(i,j) K_h(i,k) \hat{f}^{-1}(X_i) \xi_j \xi_k + o_p(1) \\
=& \frac{1}{(n-1)h^{d/2}} \sum_{j=1}^{n} \sum_{k \neq j} H_h(j,k) \xi_j \xi_k + o_p(1) \\
\Rightarrow& N(0, \sigma^2),
\end{aligned}
$$

where σ^2 is given in (3.34).

Since $Y_i - Y_i^L = (\sqrt{v(X_i) + c_n \delta(X_i)} - \sqrt{v(X_i)})\epsilon_i$ with $v(X_i) = v(X_i; \beta_0, \theta_0)$, we can write

$$
\begin{aligned}
W_{n2} =& 2a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}(\sqrt{v(X_j)+c_n\delta(X_j)} - \sqrt{v(X_j)})^2\epsilon_j^2\xi_k\hat{f}^{-1}(X_i), \\
=& 2a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\frac{c_n^2\delta^2(X_j)}{(\sqrt{v(X_j)+c_n\delta(X_j)} + \sqrt{v(X_j)})^2}\epsilon_j^2\xi_k\hat{f}^{-1}(X_i)
\end{aligned}
$$

By taking $c_n^2 = 1/nh^{d/2}$,

$$
nh^{d/2}W_{n2} \leq \frac{1}{2}a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}\frac{\delta^2(X_j)}{v(X_j)}\epsilon_j^2\xi_k\hat{f}^{-1}(X_i)
$$

By the assumptions of ϵ, ξ and similar to the arguments in Theorem 3.3.2, one can show that the right hand side of the above is $o_p(1)$. Hence $nh^{d/2}W_{n2} = o_p(1)$. Using the facts of $\sqrt{n}-$consistency of $\hat{\beta}_n, c_n \to 0$ as $n \to \infty$, assumption (C4), and similar methods used in the previous part, one can show that $nh^{d/2}W_{n3} = nh^{d/2}W_{n4} = nh^{d/2}W_{n5} = o_p(1)$.

Next, we will consider $nh^{d/2}W_{n6}$.

$$
nh^{d/2}W_{n6} = 4nh^{d/2}a_n \sum_{i\neq j\neq k} K_{ij}K_{ik}c_n^2\delta(X_j)\delta(X_k)\epsilon_j^2\epsilon_k^2V'(X_j)V'(X_k)\hat{f}^{-1}(X_i)
$$

where

$$
V(X_i) = \frac{\sqrt{v(X_i)+c_n\delta(X_i)}}{(\sqrt{v(X_i)+c_n\delta(X_i)} + \sqrt{v(X_i)})}
$$

By adding and subtracting 1/2 from $V(X_j)$ and $V(X_k)$, $nh^{d/2}W_{n6}$ can be written as the sum of $W_{n61} + W_{n62} + W_{n63}$; where

$$W_{n61} = 4a_n \sum_{i \neq j \neq k} K_{ij} K_{ik} \delta(X_j) \delta(X_k) \left[V(X_j) - \frac{1}{2} \right] \left[V'(X_k) - \frac{1}{2} \right] \epsilon_j^2 \epsilon_k^2 \hat{f}^{-1}(X_i),$$

$$W_{n62} = 4a_n \sum_{i \neq j \neq k} K_{ij} K_{ik} \delta(X_j) \delta(X_k) \left[V'(X_j) - \frac{1}{2} \right] \epsilon_j^2 \epsilon_k^2 \hat{f}^{-1}(X_i),$$

$$W_{n63} = a_n \sum_{i \neq j \neq k} K_{ij} K_{ik} \delta(X_j) \delta(X_k) \epsilon_j^2 \epsilon_k^2 \hat{f}^{-1}(X_i).$$

Note that

$$\begin{aligned} V(X_i) - \frac{1}{2} &= \frac{\sqrt{v(X_i) + c_n \delta(X_i)}}{(\sqrt{v(X_i) + c_n \delta(X_i)} + \sqrt{v(X_i)})} - \frac{1}{2} \\ &= \frac{c_n \delta(X_i)}{2[\sqrt{v(X_i) + c_n \delta(X_i)} + \sqrt{v(X_i)}]^2} \\ &\leq \frac{c_n \delta(X_i)}{8v(X_i)} \end{aligned} \tag{3.35}$$

Since $v(x; \beta_0, \theta_0)$ is a continuous function and the design variable X has the compact support I, there is a constant $c > 0$ such that $v(x; \beta_0, \theta_0) \geq c$ for $\forall x \in I$. Hence by the inequality in 3.35, $V(X_i) - \frac{1}{2} \leq \frac{c_n \delta(X_i)}{8c}$.

Now $|W_{n61}|$ is bounded above by

$$\frac{c_n^2}{16c^2} a_n \sum_{i \neq j \neq k} K_{ij} K_{ik} \delta^2(X_j) \delta^2(X_k) \epsilon_j^2 \epsilon_k^2 \hat{f}^{-1}(X_i)$$

The continuity and boundedness of δ and using the similar arguments in Theorem 3.3.2, the second part of the above expression is $O_p(1)$. Since $c_n^2 = 1/nh^{d/2}$ which goes to 0 as $n \to \infty$, it is clear that $W_{n61} = o_p(1)$. Using the same arguments, we can show that $W_{n62} = o_p(1)$.

Again using the same methods used in proving Theorem 3.3.2, we can show that
$W_{n63} = E[(\delta(X))^2 f(X)] + o_p(1)$.
Hence, $nh^{d/2}W_{n6} \to E[(\delta(X))^2 f(X)]$ in probability. Hence as the summary,
$nh^{d/2}Z_n = nh^{d/2}W_{n1} + E[\delta^2(X)f(X)] + o_p(1)$, so $nh^{d/2}Z_n \to N(E[\delta^2(X)f(X)], \sigma^2)$ and it completes the proof of Theorem (3.3.3).

■

Chapter 4

Summary and Future Research

Two testing procedures; Minimum Distance test and Empirical Smoothing Lack-of-Fit test are developed in the thesis to assess the adequacy of fitting parametric variance function in heteroscedastic regression models. The asymptotic distribution of the test statistics are shown to be normal and the estimators of the parameters are $\sqrt{n}-$consistent. The asymptotic power of the proposed tests against some local nonparametric alternatives are also investigated. Numerical simulation studies are conducted to evaluate the finite sample performance of the tests. It reveals that the Empirical Smoothing Lack-of-Fit test is more powerful and computationally more efficient than some existing tests. Also the simulation studies show that the selection of bandwidths and the different distributions of the error terms have some effects on both the accuracy and the power of the test. Therefore, in the real world problems, it is better to perform the tests with several values of bandwidth and different error distributions to make a decision to reject or not to reject the null hypothesis.

Although the motivation of the Empirical Smoothing Lack-of-Fit test in Chapter 3 of the thesis is to construct a more precise test by modifying Wang and Zhou (2006)'s test, it turns out that there are some interesting connections with the method proposed in Chapter 2 of the thesis using minimum distance test. In fact, if we choose $w = h$, $dG(x) = \hat{f}_h(x)dF_n(x)$ in the minimum distance test statistic $T_n(\beta, \theta)$, where $F_n(x)$ is the empirical cumulative distribution function of X_i's, then after a slight and obvious modification, $T_n(\hat{\beta}_n, \hat{\theta}_n)$ is simply Z_n defined in Empirical Smoothing Lack-of-Fit test.

In both of the tests proposed in previous Chapters, one of the main assumptions is a known parametric form of the mean function. In the real world problems, this assumption may be violated. It can be relaxed by estimating the mean function using kernel-smoothing estimator. Consider the following regression model:

$$Y = m(X) + \sqrt{v(X)}\epsilon,$$

where Y is a one dimensional response variable, X is a d-dimensional explanatory variable, $m(.)$ is the mean function only assumed to be smooth, and $v(x)$ is the conditional variance function of Y given $X = x$. We want to test

$$H_0 : v(X) = v(X; \theta) \quad \text{for some} \quad \theta \in \Theta$$

i.e. whether the variance function $v(X)$ can be modeled parametrically. Let $\hat{m}(x)$ be the estimator of $m(x)$ using kernel-smoothing method.

Define the the test statistic similarly as in Chapter 2 using Minimum Distance method,

$$T_n(\theta) = \int_C \left[\frac{\sum_{i=1}^n K_h(x - X_i)[(Y_i - \hat{m}(X_i))^2 - v(X_i; \theta)]}{\sum_{i=1}^n K_w(x - X_i)} \right]^2 dG(x) \tag{4.1}$$

and the corresponding estimate of θ is

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} T_n(\theta).$$

Considering the Empirical Smoothing Lack-of-Fit test in Chapter 3, define $\hat{\xi}_i = (Y_i - \hat{m}(X_i))^2 - v(X_i; \hat{\theta}_n)$, where $\hat{\theta}_n$ is the estimator of θ. Under the null hypothesis, the test statistic can be written as

$$Z_n = \frac{1}{n(n-1)(n-2)h^{2d}} \sum_{i=1}^{n} \left[\sum_{j \neq i} \sum_{k \neq i,j} K_h(i,j) K_h(i,k) \hat{\xi}_j \hat{\xi}_k \right] \hat{f}^{-1}(x_i). \tag{4.2}$$

Under the above circumstances, the asymptotic distributions of the test statistics of 4.1 and 4.2 under the null hypothesis and consistency, asymptotic power under some local alternatives can be discussed as a future work.

Bibliography

Beran, R. J. (1977). Minimum hellinger distance estimates for parametric models. *Ann. Statist. 5*, 445–463.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes*, Volume 2nd Ed. New York: Springer.

Breusch, T. S. and A. R. Pagan (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica 47*, 1287–1294.

Cook, R. D. and S. Weisberg (1983). Diagnostic for heteroscedastic in regression. *Biometrika 48*, 1–10.

Dette, H. (2002). A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *Journal of Statistical Planning and Inference. 103*, 311–329.

Dette, H. and A. Munk (1998). Testing heteroscedasticity in nonparametric regression. *J.R. Statist. Soc. Ser. B26*, 693–708.

Dette, H., N. Neumeyer, and I. Keilegom (2007). A new test for the parametric form of the variance function in non-parametric regression. *J.R. Statist. Soc. Ser. B69*, 903–917.

Diblasi, A. and A. Bowman (1997). Testing for constant variance in a linear models. *Statistics and Probability Letters 33*, 95–103.

Drossos, C. A. and A. N. Philippou (1980). A note on minimum distance estimates. *Ann. Math. Statist. 32*, 121–123.

Eubank, R. L. and W. Thomas (1993). Detecting heteroscedasticity in nonparametric regression. *J.R. Statist. Soc. Ser. 55*, 145–155.

Hardle, W. and E. Mamman (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist. 21*, 1926–1947.

Harrison, M. J. and B. P. M. McCabe (1979). A test of heteroscedasticity based on least squares residuals. *J. Amer. Statist. Assoc. 74*, 494–500.

Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist. 40*, 633–643.

Koener, R. and G. Basset (1981). Robest test for heteroscedasticity based on regression quantiles. *Econometrica 50*, 43–61.

Koul, H. L. and P. Ni (2004). Minimum distance regression model checking. *Journal of Statistical Planning and Inference. 119*, 109–141.

Koul, K. L. and W. Song (2009). Minimum distance regression model checking with berkson measurement errors. *Journal of Statistical Planning and Inference. 138*, 1615–1628.

Kutner, M. H., C. J. Nachtsheim, J. Neter, and W. Li (2004). *Applied linear statistical models*, Volume 5th Ed. McGraw-Hill.

Liero, H. (2003). Testing homoscedasticity in nonparametric regression. *Nonparametric Statistics. 15(1)*, 31–51.

Mack, Y. P. and B. W. Siverman (1982). Weak and strong uniform consistency of kernel regression estimates. *Probability Theory and Related Fields 61*, 405–415.

Muller, H. G. and P. L. Zhao (1995). On a semi-parametric variance function model and a test for heteroscedasticity. *Ann. Statist. 23*, 946–967.

Song, W. and J. Du (2011). A note on testing the regression functions via nonparametric smoothing. *Canadian Journal of Statistics 39*, 579–600108–125.

Wang, L. and X. A. Zhou (2006). Assessing the adequacy of variance function in heteroscedastic regression models. *Biometrics 63*, 1218–1225.

White, H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test of heteroscedasticity. *Econometrica 48*, 817–838.

White, H. (1981). Consequences and detection of misspecified nonlinear regression models. *J. Amer. Statist. Assoc. 76*, 419–433.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*, 1–25.

Wolfowitz, J. (1953). Estimation by the minimum distance method. *Annals of the Institute of Statistical Mathematics 5*, 9–23.

Wolfowitz, J. (1954). Estimation by the minimum distance method in nonparametric stochastic difference equations. *Ann. Math. Statist. 25*, 203–217.

Wolfowitz, J. (1957). The minimum distance method. *Ann. Math. Statist. 28*, 75–88.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics 75*, 263–289.

Zhu, L. X., Y. Fujikoshi, and K. Naito (2001). Heteroscedasticity checks for regression models. *Science in China Series A: Mathematics. 44*, 1236–1252.

Appendix A

R Codes

A.1 Minimum Distance Test - One Dimensional

```
#   "Minimum Distance Conditional Variance Function
#   Checking in Heteroscedastic Regression Models"
#   Using Bootstrap Method (One Dimensional)

      set.seed(5637)

   a=1; # constant in bandwidth: 1, 0.8, 0.5

total=500;   #  Simulation runs
power=matrix(rep(0,28),nrow=4)

k1=1;
for(b in c(0, 0.5, 0.8, 1))
 {
  k2=1;
  for(n in c(100, 200, 300, 400, 500, 800,1000))
   {
    h=a*n^{-1/3}    # Bandwidth

    Mx=0; Sx=1;       # Mean and Stdev of design variable
    Me=0; Se=1;       # Mean and Stdev of error
```

```
 bt1=1; bt2=2; th1=2; th2=0.1;      #  True valus of parameters

 K=function(u){3*(1-u^2)*(abs(u)<=1)/4};  # Kernel Function

# variables to store MD estimate adn MD test statistic

Tn=Est.theta1=Est.theta2=rep(0,total);

  freq=0;

   for(i in seq(total))
     {

   ##### Generating Sample #####

    repeat
    {
     x=runif(n,-3,3);
     e=rnorm(n,Me,Se);
     y=bt1+bt2*x+sqrt(th1+th2*x+b*x^2)*e;

     #  LSE for the regression parameter

     myreg1=lm(y~x);

     ###### Minimum Distance Estimate #####

      ngrid=300;
      xgrid=seq(-3,3,length=ngrid);
      dgrid=xgrid[2]-xgrid[1];
      xdiff=kronecker(xgrid,rep(1,n))-kronecker(rep(1,ngrid),x);
      Kh=K(xdiff/h)/h;
      mKh=matrix(Kh,nrow=ngrid,byrow=T);

     y2=(myreg1$residual)^2;
     yT=mKh%*%y2;
     x1T=apply(mKh,1,sum);
     x2T=mKh%*%x;
     myreg2=lm(yT~x1T+x2T-1);
     theta1=myreg2$coefficient[1];
     theta2=myreg2$coefficient[2];
     Est.theta1[i]=theta1;
     Est.theta2[i]=theta2;
```

```
     if(all((theta1+theta2*x)>0)) break;
         }

          ###### Bootstrap step ########
    res=myreg1$residual/sqrt(theta1+theta2*x);
    res=(res-mean(res))/sd(res); # standardization of residuals #

        TTn=rep(0,400);       #Bootstrap sample size=400
          for(j in seq(400))
            {
             bres=sample(res,replace=T)
             bY=myreg1$fitted+sqrt(theta1+theta2*x)*bres;
             myreg3=lm(bY~x);
             yy2=(myreg3$residual)^2;
             yT=mKh%*%yy2;
             x1T=apply(mKh,1,sum);
             x2T=mKh%*%x;
             myreg4=lm(yT~x1T+x2T-1);
             xi=yy2-(myreg4$coefficient[1]+myreg4$coefficient[2]*x);
             TTn[j]=sum((mKh%*%xi/n)^2)*dgrid;
            }
      cval=TTn[order(TTn)][380]    # Bootstrap critical value
      xi=(myreg1$residual)^2-(myreg2$coefficient[1]+myreg2$coefficient[2]*x);
      Tn[i]=sum((mKh%*%xi/n)^2)*dgrid;
      freq=freq+(Tn[i]>=cval)
     }
      power[k1,k2]=freq/total;   #Power of the test
     k2=k2+1;
    }
    k1=k1+1;
  }
dimnames(power)=list(c("M0","M1","M2","M3"), c(100,200,300,400,500,800,1000))
 power
```

A.2 Minimum Distance Test - Two Dimensional

```
#   "Minimum Distance Conditional Variance Function
#   Checking in Heteroscedastic Regression Models"
#   Using Bootstrap Method (Two Dimensional)
```

```
set.seed(9999)

 a=1;          # constant in bandwidth: 1, 0.8, 0.5

 total=500;     # Simulation runs
power=matrix(rep(0,28),nrow=4)

k1=1;
for(b in c(0,0.5,0.8,1))

 {

  k2=1;

  for(n in c(100,200,300,400,500,800,1000))

   {

    h=a*n^{-1/3}    # Bandwidth

bt0=1; bt1=2; bt2=1;   # True values of parameters
th0=2; th1=0.1; th2=0.1;
    K=function(u,v){9*(1-u^2)*(1-v^2)*(abs(u)<=1)*(abs(v)<=1)/16};# Kernel Function
# Variables to store MD estimate and MD test statistic

   Tn=Est.theta0=Est.theta1=Est.theta2=rep(0,total);

  freq=0;
      for(i in seq(total))
      {

         ####### Generating Sample ####

       repeat

         {

          b0=bt0;

          b1=bt1;

          b2=bt2;
```

```
  th0=th0;

  th1=th1;

  th2=th2;

  x1=rnorm(n,0,1);

  x2=rnorm(n,0,1);

  e=runif(n,-1.732,1.732);

  y=b0+b1*x1+b2*x2+sqrt(th0+th1*x1+th2*x2+b*x1^2+b*x2^2)*e;

# LSE for the regression parameter

  myreg1=lm(y~x1+x2);

####### Minimum Distance Estimate ######

 x1diff=kronecker(x1,rep(1,n))-kronecker(rep(1,n),x1)

 x2diff=kronecker(x2,rep(1,n))-kronecker(rep(1,n),x2)
 Kh=K(x1diff/h,x2diff/h)/(h^2);
 mKh=matrix(Kh,nrow=n,byrow=T)
 y2=(myreg1$residual)^2;
 yT=mKh%*%y2;
 x0T=apply(mKh,1,sum);
 x1T=mKh%*%x1;
 x2T=mKh%*%x2;
 myreg2=lm(yT~x0T+x1T+x2T-1);
 theta0=myreg2$coefficient[1];
 theta1=myreg2$coefficient[2];
 theta2=myreg2$coefficient[3];
 Est.theta0[i]=theta0;
 Est.theta1[i]=theta1;
 Est.theta2[i]=theta2;
 if(all((theta0+theta1*x1+theta2*x2)>0)) break;

 }
   ##### Bootstrap step  #######

  res=myreg1$residual/sqrt(theta0+theta1*x1+theta2*x2);
```

```
                res=(res-mean(res))/sd(res); # standardization of residuals

                TTn=rep(0,400);    # Bootstrap sample size=400

                for(j in seq(400))

                  {

                   bres=sample(res,replace=T)

                   bY=myreg1$fitted+sqrt(theta0+theta1*x1+theta2*x2)*bres;

                   myreg3=lm(bY~x1+x2);

                   yy2=(myreg3$residual)^2;

                   yT=mKh%*%yy2;

                   x0T=apply(mKh,1,sum);

                   x1T=mKh%*%x1;

                   x2T=mKh%*%x2;

                   myreg4=lm(yT~x0T+x1T+x2T-1);

xi=yy2-
(myreg4$coefficient[1]+myreg4$coefficient[2]*x1+myreg4$coefficient[3]*x2);

                   TTn[j]=sum((mKh%*%xi/n)^2)/n;

                  }
        cval=TTn[order(TTn)][380]    # Bootstrap critical value

xi=(myreg1$residual)^2-
(myreg2$coefficient[1]+myreg2$coefficient[2]*x1+myreg2$coefficient[3]*x2);

                 Tn[i]=sum((mKh%*%xi/n)^2)/n;

                 freq=freq+(Tn[i]>=cval)

          }
```

```
      power[k1,k2]=freq/total;   # Power of the test

     k2=k2+1;

     cat("b=",b,"n=",n,"\n")

    }

    k1=k1+1;

  }

dimnames(power)=list(c("M0","M1","M2","M3"),c(100,200,300,400,500,800,1000))

 power
```

A.3 Empirical Smoothing Lack-of-Fit Test

```
##   "Empirical Smoothing Lack-of-Fit Tests for Variance Function "#
##   Using Bootstrap Method  ######
    rm(list=ls())

    set.seed(5637)

      a=1;   # constant in bandwidth: 1, 0.8, 0.5
      total=500;     #  Simulation runs
      power=matrix(rep(0,28),nrow=4)
      k1=1;
      for(b in c( 0, 0.5, 0.8, 1))
       {
        k2=1;
        for(n in c(50,100,200,300,400,500, 800))
         {
          h=a*n^{-1/3}    # Bandwidth

          Mx=0; Sx=1;  #Mean and Stdev of design variable
          Me=0; Se=1;  # Mean and Stdev of error
          bt1=1; bt2=2;  #True values of parameters
          th1=2; th2=0.1;
          K=function(u){dnorm(u)};  # Kernel Function
```

```
### variables to store MD estimate and MD test statistic  ###

      Tn=Est.theta1=Est.theta2=rep(0,total)
    Zn=rep(0,total)
    Sigma=rep(0,total)
      freq=0;
      for(i in seq(total))
        {

          ######### Generating Sample ########

         repeat
           {
            x=runif(n,-3,3);
            e=rnorm(n,0,1);
            y=bt1+bt2*x+sqrt(th1+th2*x+b*x^2)*e;
            #  LSE for the regression parameter
            myreg1=lm(y~x);

            ######### Minimum Distance Estimate ########

            ngrid=200;
            xgrid=seq(-3,3,length=ngrid);
            dgrid=xgrid[2]-xgrid[1];

            xdiff=kronecker(xgrid,rep(1,n))-kronecker(rep(1,ngrid),x)

            Kh=K(xdiff/h);

            mKh=matrix(Kh,nrow=ngrid,byrow=T)

            y2=(myreg1$residual)^2;
            yT=mKh%*%y2;
            x1T=apply(mKh,1,sum);
            x2T=mKh%*%x;

            myreg2=lm(yT~x1T+x2T-1);
            theta1=myreg2$coefficient[1];
            theta2=myreg2$coefficient[2];
            if(all((theta1+theta2*x)>0)) break;
           }
```

```
              ######### Bootstrap step ########

           res=myreg1$residual/sqrt(theta1+theta2*x);
            res=(res-mean(res))/sd(res);# standardization of residuals #

             TTn=rep(0,400);    # Bootstrap sample size=400
             for(j in seq(400))
               {
                bres=sample(res,replace=T)
                bY=myreg1$fitted+sqrt(theta1+theta2*x)*bres;
                myreg3=lm(bY~x);
                yy2=(myreg3$residual)^2;
                yT=mKh%*%yy2;
                x1T=apply(mKh,1,sum);
                x2T=mKh%*%x
                myreg4=lm(yT~x1T+x2T-1)
                xi=yy2-(myreg4$coefficient[1]+myreg4$coefficient[2]*x)
  xdiff=kronecker(x,x, FUN="-")
            xKh=matrix(xdiff,nrow=n) # xij for i ne j
                Kh1=K(xKh/h)/h
                fx=apply(Kh1, 1, mean)*n/(n-1)-1/((n-1)*sqrt(2*pi))
 A1=Kh1%*%xi-diag(Kh1)*xi
  A2=Kh1^2%*%xi^2-diag(Kh1^2)*xi^2
     Zn=(sum(A1^2/fx)-sum(A2/fx))/(n*(n-1)*(n-2))
            H= function(u){dnorm(u, 0, sqrt(2))}
            Hh=H(xKh/h)/h
                Sigma= 2*h*(t(xi^2)%*%(Hh^2)%*%(xi^2)-diag(Hh^2)%*%(xi^4))/(n*(n-1))
              TTn[j]=n*sqrt(h)*abs(Zn)/sqrt(Sigma)
         }
        cval=TTn[order(TTn)][380] # Bootstrap critical Value

  xi=(myreg1$residual)^2-(myreg2$coefficient[1]+myreg2$coefficient[2]*x)
A1=Kh1%*%xi-diag(Kh1)*xi
A2=Kh1^2%*%xi^2-diag(Kh1^2)*xi^2
Zn=(sum(A1^2/fx)-sum(A2/fx))/(n*(n-1)*(n-2))
    Sigma= 2*h*(t(xi^2)%*%(Hh^2)%*%(xi^2)-diag(Hh^2)%*%(xi^4))/(n*(n-1))
Tn[i]=n*sqrt(h)*abs(Zn)/sqrt(Sigma)
    freq=freq+(Tn[i]>=cval)

    cat(b, n, i, " ", theta1, " ", theta2, " ", freq/total,"\n")

       }
        power[k1,k2]=freq/total;   # Power of the test
```

```
          k2=k2+1;
         }
         k1=k1+1;
       }
     dimnames(power)=list(c("M0", "M1", "M2", "M3"),c(50,100,200,300,400,500, 800))
     power

cat("Calculation took", proc.time()[1], "seconds.\n")
```

A.4 Plots of Empirical Sizes and Powers of the Tests

```
data1<-read.table("D:/Academic/simulation/RPOLT/power0unif1.txt", header = T)
data1
# Range of X and Y

x1range <- range(data1$Size)
y1range <- range(data1$pvalue)
data2<-read.table("D:/Academic/simulation/RPOLT/power0norm1.txt", header = T)
data2

# Range of X and Y

x2range <- range(data2$Size)
y2range <- range(data2$pvalue)

par(mfrow=c(1,2))

##### set up the plot ####

plot(x1range, y1range, type="n", xlab="Sample Size",
   ylab="Frequency of Rejecting the Null", ylim=c(0,0.1))
colors <- rainbow(3)
linetype <- c(1:3)
plotchar <- seq(18,22,1)
abline(h=0.05)
abline(h=1)

# add lines

for (i in 1:3) {
```

```
  method1 <- subset(data1, Method==i)
  lines(method1$Size, method1$pvalue, type="b", lwd=1.5,
    lty=linetype[i], col=colors[i], pch=plotchar[i])
}
# add a title and subtitle

title("Empirical Sizes  with Uniform Error (a = 1)", font.main= 1)

# add a legend

legend(600, 0.02, c("MD","ES","WZ"), cex=0.8, col=colors,
   pch=plotchar, lty=linetype, title="Method")
##############################################
# set up the plot
plot(x2range, y2range, type="n", xlab="Sample Size",
   ylab="Frequency of Rejecting the Null", ylim=c(0,0.1))
colors <- rainbow(3)
linetype <- c(1:3)
plotchar <- seq(18,22,1)
abline(h=0.05)
abline(h=1)

# add lines

for (i in 1:3) {
  method2 <- subset(data2, Method==i)
  lines(method2$Size, method2$pvalue, type="b", lwd=1.5,
    lty=linetype[i], col=colors[i], pch=plotchar[i])
}

# add a title and subtitle

title("Empirical Sizes with Normal Error (a = 1)", font.main=1)

# add a legend

legend(600, 0.02, c("MD","ES","WZ"), cex=0.8, col=colors,
   pch=plotchar, lty=linetype, title="Method")
```