Adequacy checking for the variance function in nonparametric regression

by

Jia Liang

B.S., Central University of Finance and Economics, 2012

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

2020

Abstract

Correctly specifying the parametric form of the variance function in regression models can help us make more efficient statistical inferences. Many existing Lack-of-fit testing procedures have already been proposed to decide the proper forms of the variance function, however, most of them are either checking the homoscedasticity, that is, to see if the variance function is a constant, or checking a pre-specified parametric forms of the variance function under the assumption of the mean regression function being known. In this report, we would like to construct some formal testing procedure to check the appropriateness of certain parametric forms for the variance function when the mean regression function is unknown.

The report consists of two parts. In the first part, we propose a minimum distance-based test to check the forms of the variance function. The test statistics is a modified L2-distance between a nonparametric estimate and a parametric estimate of the variance function under the null hypothesis. The Nadaraya-Watson kernel regression function estimator is used to estimate the regression function. The large sample properties, including the consistency and asymptotic normality, of the minimum distance estimate for the parameters in the variance function are discussed; the asymptotic distribution of the test statistics under the null hypothesis is established, as well as the consistency of the test and the power under local alternative hypotheses. Simulation studies, comparison studies, as well as some applications to the real data sets, are carried out to evaluate the finite sample performance of the proposed test.

In the second part, we proposed a computationally efficient test procedure for checking the parametric forms of the variance function. The test is based on an empirical smoothing of the fitted residuals by replacing the mean regression function with the Nadaraya-Watson estimator and a pre-obtained root-n consistent estimate of the parameter in the variance function. By multiplying the kernel density estimate at each individual sample points to the

fitted residual, we successfully remove the constraint of compact support for design variables assumed in some existing work. Large sample properties of the proposed test under the null hypothesis is discussed alongside with consistency of the test and the power under local alternatives. Finally, some simulation studies are carried out showing the performance of the test under finite population.

Adequacy checking for the variance function in nonparametric regression

by

Jia Liang

B.S., Central University of Finance and Economics, 2012

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

2020

Approved by:

Major Professor
Dr. Weixing Song

Copyright

© Jia Liang 2020.

Abstract

Correctly specifying the parametric form of the variance function in regression models can help us make more efficient statistical inferences. Many existing Lack-of-fit testing procedures have already been proposed to decide the proper forms of the variance function, however, most of them are either checking the homoscedasticity, that is, to see if the variance function is a constant, or checking a pre-specified parametric forms of the variance function under the assumption of the mean regression function being known. In this report, we would like to construct some formal testing procedure to check the appropriateness of certain parametric forms for the variance function when the mean regression function is unknown.

The report consists of two parts. In the first part, we propose a minimum distance-based test to check the forms of the variance function. The test statistics is a modified L2-distance between a nonparametric estimate and a parametric estimate of the variance function under the null hypothesis. The Nadaraya-Watson kernel regression function estimator is used to estimate the regression function. The large sample properties, including the consistency and asymptotic normality, of the minimum distance estimate for the parameters in the variance function are discussed; the asymptotic distribution of the test statistics under the null hypothesis is established, as well as the consistency of the test and the power under local alternative hypotheses. Simulation studies, comparison studies, as well as some applications to the real data sets, are carried out to evaluate the finite sample performance of the proposed test.

In the second part, we proposed a computationally efficient test procedure for checking the parametric forms of the variance function. The test is based on an empirical smoothing of the fitted residuals by replacing the mean regression function with the Nadaraya-Watson estimator and a pre-obtained root-n consistent estimate of the parameter in the variance function. By multiplying the kernel density estimate at each individual sample points to the

fitted residual, we successfully remove the constraint of compact support for design variables assumed in some existing work. Large sample properties of the proposed test under the null hypothesis is discussed alongside with consistency of the test and the power under local alternatives. Finally, some simulation studies are carried out showing the performance of the test under finite population.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Six years of learning and working as a Ph. D, I learned so many, gained so much. None of these would have been possible without the help of my major adviser Dr. Song. I know that both in my academic life and career path building, you have dedicated a lot, and that being said, there is nothing I can pay it back now, but I will remember it, I will keep all the endeavor, everything that you taught me, trained me will continue affects me through the rest of my life. I will always be a humble disciple and keep this learning attitude forever.

The work of my final dissertation is a combination and an essential reflection of my study in the department of statistics. I have learned the statistical knowledge, trained the application skill, and gained the persistence in statistical research. These many thanks goes to all my committee members, Dr. Du, Dr. Goh and Dr. Natarajan, who has been my teachers and/or mentors in different part of my Ph. D. A simple thank you would not be enough, but I wouldn't hesitate to express my sincerest gratefulness to all of you.

I would also like to thank Dr. Albin for his willingness to be my outside chair. Mathematics is the foundation of statistics, I hope that my current work will be satisfactory to the wise eye (to me, mathematicians are always the smartest people).

Finally, I would like to thank my wife Huan and my parents for their support. Six years is a lengthy time that could take a baby to be a child, a child to finish his/her elementary school, and a student to graduate from high school. And for me, it builds the most important part of my life. This, will be an appreciation to them as well.

Chapter 1

Introduction

1.1 Introduction

It is well known that in regression analysis, knowing the variance function could help us to construct efficient estimates for the regression coefficients. Although some graphical methods, such as checking the structure of the residual plots, are still very common in practice, however, developing formal test procedures to check the form of the variance functions has been long-lasting research interest among statisticians.

Generally speaking, there are two main research directions in checking the adequacy of the variance function forms in regression analysis, one is testing the homoscedasticity or checking whether the variance function is constant or not; the other is to check if the variance function has a specific parametric form. The quest for solution on the homoscedasticity test is long and fruitful. Based on an empirical residual process, Zhu et al. (2001) constructed a Cramér-von Mises type test statistic for checking the homoscedasticity when the mean regression function has a parametric form or simply unknown. In a nonparametric regression setup with fixed design, Dette et al. (1998) proposed a homoscedasticity test based on a L_2-approximation of the variance function, the null asymptotic distribution of the test statistic is asymptotically normal. In the fixed design case, Dette and Hetzler (2009) also constructed a simple test for checking the parametric form of the variance function based on an empirical process of pseudo-residuals, and in some special cases, the limiting process under the null hypothesis is a Brownian bridge. In the random design case, Dette et al. (2007) proposed two tests for checking the parametric form of the variance function. Based on the empirical residual processes, Dette et al. (2007) showed that the proposed Kolmogorov-Smirnov and Cramér-von Mises type tests asymptotically follow a complicated distribution related to a zero mean gaussian process under the null hypothesis. Due to their intractable asymptotic null distributions, Zhu et al. (2001) and Dette et al. (2007) suggested to use bootstrap algorithms to implement the tests. Inspired by Zheng (1996)'s consistent test for the mean function, for a fixed design nonparametric regression model, Dette (2002) proposed a homoscedasticity test based on pseudo-residuals constructed from differencing operators. Due to the smoothing parameter, the test in Dette (2002) can detect the local hypothesis converging to the null at the rate of $1/\sqrt{nh^{1/2}}$, where n is the sample size, and h is the bandwidth used in the test statistic, while the tests proposed in Dette et al. (1998), Dette and Hetzler (2009) and Dette et al. (2007) can detect the local alternatives which converge to the null model at the rate of $1/\sqrt{n}$. These tests mentioned above are developed for the cases where the predictor is univariate. When the predictor are multidimensional, also built on Zheng (1996)'s test, Zheng (2009) proposed two nonparametric tests for homoscedasticity by assuming the mean regression function is known or unknown. It is shown that these test can detect local alternative with rate $1/\sqrt{nh^{d/2}}$, where d is the dimension of predictors. Recently, by considering the L_2-distance between the characteristic functions of a nonparametric standard residual and a semi-parametric standard residual under the null hypothesis, Pardo-Fernández and Jiménez-Gamero (2019) proposed a specification test for the variance function in nonparametric regression models, which can detect $1/\sqrt{n}$-local alternatives.

When checking the parametric form for the mean regression functions, Koul and Ni (2004) developed a minimum distance test. Assuming the mean regression function is known, Samarakoon and Song (2011) extended the methodology in Koul and Ni (2004) to check the parametric form for the variance function. Similar to the existing work, the test is made possible by the fact that the variance function can be viewed as the regression function in which the squared residual is treated as the response variable, and the test is built upon a L_2-norm of a kernel weighted average of residuals, which is modified from the L_2 distance between a nonparametric fit and a parametric fit under the null hypothesis of the mean regression function. Based on Zheng (1996)'s work, Samarakoon and Song (2012) and Song and Du (2011) also construct some consistent tests for the forms of the variance function assuming the regression function is known.

In chapter 2, we will mainly focus on the discussion of test constructed with minimum distance method, all the detailed theoretical results are provided, and simulations are carried out at the end of the discussion. Two on going real examples are studied with further details to be added. In chapter 3, we will again focus our attention on the same hypothesis, but the test constructed using the empirical smoothing method over the conditional expectation of the modified residual function.

Chapter 2

Checking Adequacy of Variance Function with Unknown Mean Function

2.1 Introduction

It is well known that in regression analysis, knowing the variance function could help us to construct efficient estimates for the regression parameters. Although some graphical methods, such as checking the structure of the residual plots, are still very common in practice, developing formal tests to check the form of the variance functions has been long-lasting research interest among statisticians.

Generally speaking, there are two main research streams in checking the adequacy of the variance function forms in regression analysis. One is testing the homoscedasticity or constancy of the variance function; the other is to check if the variance function has a specific parametric form, although the former is nested in the latter. The quest for testing the homoscedasticity has been long and fruitful. In a nonparametric regression setup with fixed design, Dette et al. (1998) proposed a homoscedasticity test based on a L_2-approximation of the variance function and the null asymptotic distribution of the test statistic is normal.

Dette and Hetzler (2009) also constructed a simple test for checking the parametric form of the variance function in a fixed design regression model based on an empirical process of pseudo-residuals, but the limiting distribution under the null hypothesis is shown to be related to Brownian bridge. A Cramér-von Mises type test based on empirical residual process was proposed in Zhu et al. (2001). In the random design case, Dette et al. (2007) proposed two tests for checking the parametric form of the variance function based on the empirical residual processes, Dette et al. (2007) showed that the proposed Kolmogorov-Smirnov and Cramér-von Mises type tests asymptotically follow a complicated distribution of a zero mean gaussian process under the null hypothesis. Due to their intractable asymptotic null distributions, Zhu et al. (2001) and Dette et al. (2007) suggested to use bootstrap algorithms to implement their tests. Inspired by Zheng (1996)'s consistent test for the mean function, for a fixed design nonparametric regression model, Dette (2002) proposed a homoscedasticity test based on pseudo-residuals constructed from difference operators. Due to the smoothing parameter, the test in Dette (2002) can detect the local hypotheses converging to the null at the rate of $1/\sqrt{nh^{1/2}}$, where n is the sample size, and h is the bandwidth used in the test statistic, while the tests proposed in Dette et al. (1998), Dette and Hetzler (2009) and Dette et al. (2007) can detect the local alternatives which converge to the null model at the rate of $1/\sqrt{n}$. The above mentioned tests are developed for the cases where the predictor is univariate. When the predictor are multidimensional, Zheng (2009) proposed two nonparametric tests for homoscedasticity. It is shown that these test can detect local alternatives with rate $1/\sqrt{nh^{d/2}}$, where d is the dimension of predictors. Recently, by considering a L_2-distance between the characteristic functions of a nonparametric standardized residual and a semi-parametric standardized residual under the null hypothesis, Pardo-Fernández and Jiménez-Gamero (2019) proposed a specification test for the variance function in nonparametric regression models, which could detect $1/\sqrt{n}$-local alternatives.

Koul and Ni (2004) developed a minimum distance test for checking the parametric form for the mean regression functions. Assuming the mean regression function is known, Samarakoon and Song (2011) extended the methodology in Koul and Ni (2004) to check the parametric form of the variance function. The test is made possible by realizing that the variance function can be viewed as the regression function in a regression model in which the squared residual is treated as the response variable, and the test is built upon a L_2-norm of a kernel weighted average of residuals. Inspired by Zheng (1996)'s work, Samarakoon and Song (2012) and Song and Du (2011) construct some other consistent tests to verify the parametric forms of the variance function assuming the regression function is known.

Although the asymptotic theory and finite sample simulation studies show that the proposed test in Samarakoon and Song (2011) performs satisfactory, the assumption of known mean regression function restricts its applications. In this paper, we will try to relax this rigid assumption by extending Samarakoon and Song (2011)'s idea to the cases where the mean regression function does not have any specific parametric forms except for some smoothing conditions. Similar to Zheng (2009), Zhu et al. (2001), and Dette et al. (2007), we use the Nadaraya-Watson (N-W) estimator to estimate the unknown mean regression function.

The paper is organized as follows. The proposed test will be introduced in Section 2. The minimum distance estimator of the parameters in the variance function under the null hypothesis will be considered in Section 3, as well as its large sample properties such as consistency and asymptotic normality. In Section 4, the asymptotic null distribution of the proposed test statistic will be discussed, together with the consistency for fixed alternatives and the power for local alternatives. Simulation studies and real examples will be presented in Section 5, and all the proofs of the main theoretical results will be deferred to the appendix.

2.2 Minimum Distance Estimator and Test

To be specific, we consider the following regression model

$$Y = m(X) + \sqrt{v(X)}\varepsilon, \tag{2.2.1}$$

where X is a d-dimensional covariate, Y is a scalar response variable, $m(x)$ and $v(x) > 0$ are the mean regression function and variance function, respectively. The regression error ε satisfies the usual assumption $E(\varepsilon|X) = 0$ and $E(\varepsilon^2|X) = 1$. As we indicated earlier, the mean regression function $m(x)$ is unknown. The research of interest in this paper is to develop a test procedure to check if the variance function $\mathrm{Var}(Y|X = x) = v(x)$ has a pre-specified parametric form with parameters lies in a compact parametric space Θ. That is, we want to test

$$H_0 : v(x) = v(x;\theta) \quad \text{for some } \theta \in \Theta \subset \mathbb{R}^q \qquad \text{versus} \qquad H_1 : H_0 \text{ is not true}$$

Note that from (2.2.1), we have $E[(Y - m(X))^2|X] = v(X)$. Therefore, we can consider a new regression model

$$(Y - m(X))^2 = v(X) + \xi, \tag{2.2.2}$$

where $\xi = (Y - m(X))^2 - v(X)$. It is easy to see that $E(\xi|X) = 0$, ξ and $v(X)$ are uncorrelated. If $m(x)$ is known or unknown up to an Euclidean parameter, then the proposed tests in Samarakoon and Song (2011) can be used to check H_0. In the cases of unknown $m(x)$, a natural thought will be replacing $m(x)$ with some estimator. In this paper, we shall use the N-W estimator to estimate $m(x)$. To be specific, let K be a d-dimensional kernel density function, and h_n be a sequence of positive numbers tending to 0 as $n \to \infty$, then the NW estimator of $m(x)$ is defined by

$$\hat{m}(x) = \frac{1}{n\hat{f}_{h_n}(x)} \sum_{i=1}^{n} K_{h_n}(X_i - x)Y_i, \quad \hat{f}_{h_n}(x) = \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(X_i - x),$$

where $K_{h_n}(x) = {h_n}^{-d}K(x/h_n)$. Thus, mimic the test statistic introduced in Samarakoon and Song (2011), a L_2-distance between $\hat{m}(x)$ and the parametric variance function $v(x;\theta)$ under H_0 can be constructed as

$$\int_{\mathcal{C}} \left[\frac{\sum_{i=1}^{n} K_{h_n}(X_i - x)(Y_i - \hat{m}(X_i))^2}{\sum_{i=1}^{n} K_{w_n}(X_i - x)} - v(x,\theta)\right]^2 dG(x), \tag{2.2.3}$$

where G is a weighting function supported on a compact subset $\mathcal{C}$ of $\mathbb{R}^d$, and w_n is another sequence of positive numbers tending to 0 as $n \to \infty$. Note that the random denominator $\hat{f}_{h_n}(x)$ in $\hat{m}(x)$ would create some extra troubles in developing the theoretical properties of the test statistic, so we will use the following modified distance to construct the test statistic

$$T_n^*(\theta) := \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^n K_{h_n}(X_i - x)(Y_i - \hat{m}(X_i))^2 \hat{f}_{h_n}^2(X_i)}{\hat{f}_{w_n}^2(x) \sum_{i=1}^n K_{w_n}(X_i - x)} - v(x,\theta) \right]^2 dG(x)$$
$$= \int_{\mathcal{C}} \left[\frac{1}{n\hat{f}_{w_n}^3(x)} \sum_{i=1}^n K_{h_n}(X_i - x)(Y_i - \hat{m}(X_i))^2 \hat{f}_{h_n}^2(X_i) - v(x,\theta) \right]^2 dG(x).$$

Here we choose different bandwidths for the denominator and numerator in the above construction to avoid similar difficulty encountered in Koul and Ni (2004). Thus, we can estimate θ by the minimizer of $T_n^*(\theta)$, that is,

$$\hat{\theta}_n^* := \text{argmin}_{\theta \in \Theta} T_n^*(\theta),$$

and construct the test based on $T_n^*(\hat{\theta}_n^*)$. However, to avoid too much bias from nonparametric smoothing, instead of $T_n^*(\hat{\theta}_n^*)$, we construct a centered modification of $T_n^*(\theta)$,

$$T_n(\theta) := \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^n K_{h_n}(X_i - x)\{(Y_i - \hat{m}(X_i))^2 - v(X_i,\theta)\} \hat{f}_{h_n}^2(X_i) \right]^2 d\hat{\varphi}_w(x), \qquad (2.2.4)$$

and estimate θ by $\hat{\theta}_n := \text{argmin}_{\theta \in \Theta} T_n(\theta)$. Finally, we construct the test based on the statistic $T_n(\hat{\theta}_n)$, where, also throughout the paper, $d\hat{\varphi}_w(x)$ denotes $\hat{f}_w^{-6}(x)dG(x)$. The reason why we still consider $T_n^*(\theta)$ and $\hat{\theta}_n^*$ is that the proof of the consistency of $\hat{\theta}_n$ relies on the proof of consistency of $\hat{\theta}_n^*$.

2.3 Main Results

To state the main results, we start with a list of regularity conditions. For the variance function and the density function of X, we assume

(e1). $\{(X_i, Y_i);\ X_i \in R^d,\ Y_i \in R,\ i = 1, \ldots,\ n\}$ is a random sample of size n from model (2.2.1), and the variance function $v(x) = E((Y - m(X))^2 | X = x)$ satisfies $v \in L_2(G)$, where G is a σ-finite measure on R^d.

(e2). $E((Y - m(X))^2 - v(X))^2 < \infty$ and $\tau^2(x) := E(((Y - m(X))^2 - v(X))^2 | X = x)$ is a.s. G continuous on $\mathcal{C}$.

(e3). $E|(Y - m(X))^2 - v(X)|^{2+\delta} < \infty$, for some $\delta > 0$.

(e4). $E((Y - m(X))^2 - v(X))^4 < \infty$.

(f1). The design variable X has a uniformly continuous Lebesgue density f that is bounded from below on $\mathcal{C}$.

(f2). The density f is twice continuously differentiable with a compact support.

(g). There is a continuous function g such that $dG(x) = g(x)dx$.

For the kernel function K, we shall assume the following:

(k). K is a positive symmetric square integrable, and Lipschitz continuous density on $[-1, 1]^d$.

About the parametric function $v(x, \theta)$ to be fitted under the null hypothesis, we assume the following:

(v1). For each θ, $v(x, \theta)$ is a.s. continuous in x w.r.t. the weighting measure G.

(v2). The parametric family of models $v(x, \theta)$ is identifiable with respect to the parameter θ, i.e., if $v(x, \theta_1) = v(x, \theta_2)$, for almost all $x(G)$, then $\theta_1 = \theta_2$.

(v3). For some positive continuous function l on $\mathcal{C}$ and for some $0 < \beta < 1$,

$$|v(x, \theta_1) - v(x, \theta_2)| \le \|\theta_2 - \theta_1\|^\beta l(x) \ \forall \theta_2, \theta_1 \in \Theta, x \in R^d$$

(v4). For every x, $v(x, \theta)$ is differentiable in θ in a neighborhood of θ_0, and for every $\varepsilon > 0$, $k < \infty$,

$$\lim_n P\left(\sup_{1 \le i \le n, (nh_n^d)^{1/2}\|\theta - \theta_0\| \le k} \frac{|v(X_i, \theta) - v(X_i, \theta_0) - (\theta - \theta_0)^T \dot{v}(X_i, \theta_0)|}{\|\theta - \theta_0\|} > \varepsilon \right) = 0,$$

where $\dot{v}(x,\theta)$ is the derivative of $v(x,\theta)$ with respect to θ.

(v5). As a function of x, $\dot{v}(x,\theta_0)$ is continuous in $x \in \mathcal{C}$ and for every $\varepsilon > 0$, such that for every $0 < k < \infty$,

$$\lim_{n\to\infty} P\left(\max_{1\le i\le n, (nh_n^d)^{1/2}\|\theta-\theta_0\|\le k} h_n^{-d/2} \|\dot{v}(X_i,\theta) - \dot{v}(X_i,\theta_0)\| \ge \varepsilon \right) = 0.$$

As for the bandwidths h_n, w_n, we shall make the following assumptions:

(h1). $h_n \to 0$ and $nh_n^{2d} \to \infty$, as $n \to \infty$.

(h2). $nh_n^8 \to 0$ as $n \to \infty$.

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

Finally, for the parameter space Θ and the true parameter θ_0 in the variance function under the null hypothesis, we shall assume

(s). The parameter space Θ is a compact subset in $\mathbb{R}^q$, and the true parameter value θ_0 is an interior point of Θ.

Conditions (e1), (e2), (f1), (k), (v1)-(v3), (h1), (h2) and (s) suffice for the proof of consistency of $\hat{\theta}_n$, while adding (e3), (f2), (v4), (v5) and (w), we can further prove the asymptotic normality of $\hat{\theta}_n$. The proof of the asymptotic null distribution of $T_n(\hat{\theta}_n)$ requires (e1), (e2), (e4), (f1)-(v5), and (w). Note that the concurrence of (h1) and (h2) implies that $d < 4$ which puts a severe restriction on the application of the proposed test. This is the cost we might have to pay for testing the variance function when the mean regression function has to be estimated in advance. How to relax this restriction deserves an independent study. Also note that (w) implies $w_n \to 0$, $nw_n^{2d} \to \infty$.

Under conditions (f1), (k), (h1), (h2), and (w), we have

$$\sup_{x\in\mathcal{C}} |\hat{f}_{h_n}(x) - f(x)| = o_p(1), \quad \sup_{x\in\mathcal{C}} |\hat{f}_{w_n}(x) - f(x)| = o_p(1), \quad \sup_{x\in\mathcal{C}} \left| \frac{f(x)}{\hat{f}_{w_n}(x)} - 1 \right| = o_p(1).$$

These results are from Mack and Silverman (1982), which will be frequently used in the subsequent theoretical development.

Throughout the following discussion, the integrals with respect to the G-measure are understood to be over the set $\mathcal{C}$. The convergence in distribution is denoted by $\stackrel{d}{\rightarrow}$ and $N_p(\mu, \Sigma)$ denotes a p-dimensional normal distribution with mean vector μ and covariance matrix Σ, $p \geq 1$. Also, for the sake of simplicity, throughout the paper, the bandwidths h_n and w_n will be written as h and w, respectively.

2.3.1 Large Sample Results of $\hat{\theta}_n^*$ and $\hat{\theta}_n$

In this section, we shall present the large sample results for the minimum distance estimators $\hat{\theta}_n^*$ and $\hat{\theta}_n$ defined in the previous section. The main step for showing the consistency of $\hat{\theta}_n^*$ and $\hat{\theta}_n$ relies on a lemma from Koul and Ni (2004) which is reproduced here for the sake of completeness. Let $L_2(G)$ denote a class of square-integrable real-valued functions on $\mathbb{R}^d$ w.r.t. G. Define

$$\rho(v_1, v_2) = \int_{\mathcal{C}} [v_1(x) - v_2(x)]^2 dG(x) \; v_1, \; v_2 \in L_2(G)$$

and for a parametric function $m_\theta(x) \in L_2(G)$, we also define a map

$$M(v) = \text{argmin}_{\theta \in \Theta} \rho(v, m_\theta), \; v \in L_2(G).$$

Then Koul and Ni (2004) shows that

Lemma 2.3.1. Let $m_\theta(x)$ be such that

(m1). For each θ, $m_\theta(x)$ is a.s. continuous in x, w.r.t. integrating measure G.

(m2). The parametric family of models $m_{\theta_1}(x) = m_{\theta_2}(x)$ a.s. (G), then $\theta_1 = \theta_2$.

(m3). There exists a positive continuous function l on $\mathcal{C}$ and for some $\beta > 0$,

$$|m_{\theta_2}(x) - m_{\theta_1}(x)| \leq \| \theta_2 - \theta_1 \|^\beta \, l(x), \quad for\ any\ \theta_1, \theta_2 \in \Theta, x \in \mathcal{C}.$$

Then the following hold

(a). $M(v)$ always exists, $v \in L_2(G)$.

(b). If $M(v)$ is unique, then M is continuous at v in the sense that for any sequence of $\{v_n\} \in L_2(G)$ converging to v in $L_2(G)$, $M(v_n) \to M(v)$, that is,

$$\rho(v_n, v) \to 0 \Longrightarrow M(v_n) \to M(v) \quad \text{as } n \to \infty.$$

(c). $M(m_\theta(\cdot)) = \theta$, uniquely for any $\theta \in \Theta$.

Based on the above lemma, we can show the consistency result for $\hat{\theta}_n^*$.

Theorem 2.3.1. Under H_0, conditions (e1), (e2), (f1), (h1), (h2), (w) and (v1)-(v3), we have, as $n \to \infty$, $\hat{\theta}_n^* \to \theta_0$ in probability.

The consistency of $\hat{\theta}_n$ is summarized in the following theorem.

Theorem 2.3.2. Under H_0, conditions (e1), (e2), (f1), (k), (v1)-(v3), (h1), and (w), we have $\hat{\theta}_n \to \theta_0$ in probability as $n \to \infty$.

Note that $\hat{\theta}_n$ will be used to construct the test statistic, so we only report the asymptotic normality result for $\hat{\theta}_n$. Define

$$\Sigma_0 = \int \dot{v}(x, \theta_0)\dot{v}^T(x, \theta_0)dG(x), \quad \Sigma = \int \frac{\dot{v}(x, \theta_0)\dot{v}^T(x, \theta_0)\tau^2(x)g^2(x)}{f(x)}dx. \tag{2.3.1}$$

The asymptotic normality of the minimum distance estimator $\hat{\theta}_n$ is summarized in the following theorem.

Theorem 2.3.3. Under the null hypothesis H_0, assume that the conditions (e1)-(e3), (f1), (f2), (g), (k), (v1)-(v5), (h1), (h2)and (w) hold, then we have $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N_q(0, \Sigma_0^{-1}\Sigma\Sigma_0^{-1})$, where Σ_0 and Σ are defined in (2.3.1).

From Theorem 2.3.3, we see that the asymptotic variance is a functional of the weight function g. Denote $W(g) = \Sigma_0^{-1}\Sigma\Sigma_0^{-1}$, then similar to the corresponding discussion in Koul and Ni (2004), we can see that in the case of $d = 1$, by Cauchy-Schwartz inequality,

$$\Sigma_0^2 \leq \int \frac{\dot{v}^2(x,\theta_0) g^2(x) \tau^2(x)}{f(x)} dx \cdot \int \frac{\dot{v}^2(x,\theta_0) f(x)}{\tau^2(x)} dx.$$

Therefore,

$$W(g) \geq \left[\int \frac{\dot{v}^2(x,\theta_0) f(x)}{\tau^2(x)} dx \right]^{-1}$$

with equality holds if and only if $g(x) \propto f(x)/\tau^2(x)$. Note that $W(g)$ is scale invariant, that is, $W(ag) = W(g)$ for any positive number a, so an optimal weight function $g(x)$ can be chosen as $f(x)/\tau^2(x)$. However, in the real application, estimates of $f(x)$, $\tau^2(x)$ should be used.

2.3.2 Minimum Distance Test

In this section, we shall establish the test for checking the null hypothesis H_0 in model (2.2.1) based on the statistic $T_n(\theta)$ defined in (2.2.4).

To present the asymptotic distribution of the test statistic under H_0, we shall use the following notations

$$\begin{aligned}
\hat{\xi}(X_i) &= (Y_i - \hat{m}(X_i))^2 - v(X_i, \hat{\theta}_n), \quad \xi(X_i) = (Y_i - m(X_i))^2 - v(X_i, \theta_0), \\
\hat{\Gamma}_n &= \frac{h^d}{n^2} \sum_{i \neq j} \left(\int_{\mathcal{C}} K_h(X_i - x) K_h(X_j - x) \hat{\xi}(X_i) \hat{\xi}(X_j) \hat{f}_h^2(X_i) \hat{f}_h^2(X_j) \hat{f}_w^{-6}(x) dG(x) \right)^2, \\
\hat{C}_n &= \frac{1}{n} \sum_{i=1}^n \int_{\mathcal{C}} K_h^2(X_i - x) \hat{\xi}^2(X_i) \hat{f}_h^4(X_i) d\hat{\varphi}_w(x).
\end{aligned}$$

The following theorem presents the asymptotic null distribution of $T_n(\hat{\theta}_n)$ under H_0.

Theorem 2.3.4. Under H_0, we assume that conditions (e1),(e2),(e4),(f1),(f2),(g),(h1), (h2), (w),(k), and (v1)-(v5) hold, then $nh^{d/2}\hat{\Gamma}_n^{-1/2}(T_n(\hat{\theta}_n) - \hat{C}_n) \xrightarrow{d} N(0,1)$ in distribution.

Accordingly, the test rejects H_0 with significance level α whenever

$$nh^{d/2}\hat{\Gamma}_n^{-1/2} |T_n(\hat{\theta}_n) - \hat{C}_n| > z_{\alpha/2}, \tag{2.3.2}$$

where $z_{\alpha/2}$ is the $100(1-\alpha/2)\%$ percentile of the standard normal distribution.

A reasonable test should be consistent in the sense that a consistent test can detect fixed departures from the null model with arbitrary large probability as long as the sample size is sufficiently large. We claim that under some regularity conditions, the proposed test (2.3.2) is consistent.

To be specific, consider the alternative hypothesis,

$$H_a : v(x) = v_a(x), \text{ for } x \text{ a.s } (G),$$

where $v_a(x) \notin \{v(x,\theta), \theta \in \Theta\}$. Under H_a, in general the minimum distance estimate $\hat{\theta}_n$ will not be a consistent estimator of θ_0 due to the misspecification of the variance function. However, under very general regularity conditions, we can still show that there exists a value $\theta_a \in \Theta$ such that $\sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1)$. We will not rigorously prove this fact in this paper, and instead, we direct the reader to Samarakoon and Song (2011) for more explanations.

To state the consistency result, we need the following condition which rephrases more precisely about $v_a(x) \notin \{v(x,\theta), \theta \in \Theta\}$:

$$\int_{\mathcal{C}} (v_a(x) - v(x,\theta_a))^2 dG(x) > 0. \tag{2.3.3}$$

Theorem 2.3.5. Suppose that all the conditions in Theorem 2.3.4 holds for θ_a instead of θ_0. Then under H_a and (2.3.3), for $0 < \alpha < 1$, the test that rejects H_0 whenever $|nh^{d/2}\Gamma_n^{-1/2}(T_n(\hat{\theta}_n) - \hat{C}_n)| \geq z_{\alpha/2}$ is consistent for H_a.

Finally, we investigate how the proposed test performs when checking some local alternative hypothesis which converges to the null hypothesis as the sample size $n \to \infty$. For this purpose, we consider the following local alternative hypothesis,

$$H_{\text{loc}} : v(x) = v(x,\theta_0) + c_n\delta(x),$$

where $\delta(x)$ is a continuous function on $\mathcal{C}$ such that $\int_{\mathcal{C}} \delta^2(x)dG(x) < \infty$. Since $v(x)$ is the variance function, so $\delta(x)$ should be some function to guarantee the non-negativeness of $v(x, \theta_0) + c_n\delta(x)$. c_n is a sequence of positive numbers tending to 0 as $n \to \infty$. The following theorem states that when the variance function under the local alternative hypothesis converges to the variance function under the null hypothesis at the rate of $1/\sqrt{nh^{d/2}}$, the proposed test still has some nontrivial power to detect the departure.

Theorem 2.3.6. *Suppose all the conditions in Theorem 2.3.4 hold. Then under the local alternative hypothesis H_{loc}, if $c_n = (nh^{d/2})^{-1/2}$, we have*

$$nh^{d/2}\Gamma_n^{-1/2}(T_n(\hat{\theta}_n) - \hat{C}_n) \xrightarrow{d} N\left(\Gamma^{-1/2}\int_{\mathcal{C}} \delta^2(x)dG(x), 1\right)$$

in distribution as $n \to \infty$, where

$$\Gamma = 2\int \frac{\tau^4(x)g^2(x)}{f^2(x)}dx \int \left(\int K(u)K(v+u)du\right)^2 dv.$$

From Theorem 2.3.6, we see that the optimal weight function G which maximizes the asymptotic local power of the proposed test is the one to maximize the mean of the asymptotic normal distribution, or $\Psi(g) := \Gamma^{-\frac{1}{2}} \int_{\mathcal{C}} \delta^2(x)dG(x)$. By Cauchy-Schwartz inequality, we have

$$\Psi(g) \leq \frac{1}{(2\int\left(\int K(u)K(v+u)du\right)^2 dv)^{1/2}} \cdot \left(\int \frac{\delta^4(x)f^2(x)}{\tau^4(x)}dx\right)^{1/2}$$

with equality if and only if $g(x) \propto \tau^{-4}(x)\delta^4(x)f^2(x)$ for all x. Note that the functional Ψ is also scale-invariant, that is $\Psi(ag) = \Psi(g)$ for all positive constant $a > 0$, we can choose the optimal weight function $g(x)$ to be $\tau^{-4}(x)\delta^4(x)f^2(x)$. However, due to many unknown quantities in this form, it simply cannot be used, therefore, in real application, these unknown quantities should be replaced by certain estimates.

2.4 Numerical Study

To evaluate the finite sample performance of the minimum distance test proposed in this paper, we conduct some simulation studies in this section, as well as two real data applications. To implement the proposed methods, the Epanechnikov kernel $K(x) = 0.75(1-x^2)I(|x| \leq 1)$ is used as the kernel function, the bandwidth h and w are chosen to be $an^{-1/4}$ and $[(\log n)/n]^{-1/5}$, respectively. Several values of a are selected to see the effect of the bandwidth on the performance of the minimum distance estimator and the test. The compact set $\mathcal{C}$ is chosen to be $[0,1]$ to avoid any computational instability of $\hat{f}_w(x)$ when x approaches the boundary. Accordingly, the weight function G is chosen to be a function supported on $\mathcal{C}$. The nominal significance level is chosen to be 0.05, so 1.96 serves as the critical value for the decision rule. In the simulation study, the sample sizes are chosen to be $n = 100, 200, 300, 400$ and 500, and in each scenario, the simulation is repeated 500 times, and the empirical power, calculated by the percentage of how time times the test statistic $\hat{\Gamma}_n^{-1/2} nh^{d/2}|T_n(\hat{\theta}_n) - \hat{C}_n|$ exceeds 1.96.

2.4.1 Simulation Study

First we generate the sample from the regression model $Y = m(X) + \sqrt{v(X)}\varepsilon$ with $m(x) = 1 + 2x + 3x^2$ and variance function $v(x, \theta) = \theta x^2$. Therefore, the null hypothesis H_0 is set to be $v(x, \theta) = 1 + \theta x^2$ with the true parameter $\theta_0 = 0.5$. The regression error $\varepsilon \sim N(0, 1)$ and $X \sim \text{Uniform}(0, 1)$. The bandwidth h is chosen to be $an^{-1/4}$ with four different adjustment values $a = 0.8, 1, 1.2, 1.5$.

To evaluate the power of the proposed test, we consider the following five fixed or local alternative variance functions

$$H_1 : v(X) = \exp(x+1), \quad H_2 : v(x) = \delta(x), \quad H_3 : v(X) = v(x, \theta_0) + n^{-4}\delta(x),$$
$$H_4 : v(X) = v(x, \theta_0) + (nh^{d/2})^{-1/2}\delta(x), \quad H_5 : v(X) = v(x, \theta_0) + (\ln n)^{-1}\delta(x).$$

where $\delta(x) = (x^2 + 0.1)^{-1}$. H_1 and H_2 are fixed alternative hypotheses, and the local alternative hypothesis H_3 converges to the null hypothesis faster than $c_n = (nh^{1/2})^{-1/2}$ specified in Theorem 2.3.6, H_4 converges to the null hypothesis at the exact rate c_n, and H_5 converges to the null hypothesis at a slower rate of $(\ln n)^{-1}$.

The parameter θ_0 in the variance function is estimated using $\hat{\theta}_n = \text{argmin}_\theta T_n(\theta)$ as described in Section 2, and two different weight function G are considered, the uniform density function and $g(x) = \hat{f}_w^6(x)$, both are supported over $\mathcal{C} = [0, 1]$. The latter one is intentionally chosen to eliminate the denominator inside the test statistics $T_n(\hat{\theta}_n)$ and the estimated variance $\hat{\Gamma}_n$.

The simulation results are summarized in Table 2.1, 2.2 and 2.3. Table 2.1 reports the simulation results for $G(x)$ being uniform. All simulations results show a good control of the empirical levels of the test, although somewhat conservative. As expected, the power of fixed alternative hypotheses tends to be bigger as the sample size increases, and the test has higher probability to detect the local alternatives with slower convergence rates towards the null hypothesis. It is also clear that the bandwidth selection does affect the performance of the test.

Table 2.2 presents the simulation results for $g(x) = \hat{f}_w^6(x)$. The overall performance of the test in this case are satisfying. Once can clearly see that the empirical levels are still conservative in most scenarios, while the empirical powers are maintained very well in detecting the alternatives. Also the simulation results are pretty stable for difference choices of a values.

Table 2.3 presents the simulation results based on the estimated optimal weight $g(x) = \tau^{-4}(x)\delta^4(x)f^2(x)$. The functions $\tau^4(x)$ and $f^2(x)$ are estimated using a modified N-W estimator and the classic kernel density estimator, respectively. The estimation of θ uses $\text{argmin}_\theta T_n(\theta)$ with $g(x) = \hat{f}_w^6(x)$. In particular, the modified N-W estimator estimation of $\tau^2(x) = E(\xi^2(X)|X = x)$ is defined as

$$\frac{n^{-1}\sum_{i=1}^n K_h(X_i - x)\xi^2(X_i)f_h^2(X_i)}{\hat{f}_h(x)\hat{f}_w^2(x)}.$$

The benefit of such modification is that this will, together with $\hat{f}_w^2(x)$ for $f^2(x)$, eliminate the

Table 2.1: Empirical Levels and Powers ($g(x) = I_{[0,1]}(x)$)

	n	100	200	300	400	500
$a = 1.5$	H_0	0.030	0.026	0.018	0.024	0.020
	H_1	0.264	0.556	0.810	0.930	0.986
	H_2	0.988	0.994	1.000	1.000	1.000
	H_3	0.080	0.012	0.016	0.008	0.014
	H_4	0.368	0.456	0.532	0.636	0.646
	H_5	0.710	0.950	0.996	1.000	1.000
$a = 1.2$	H_0	0.034	0.024	0.022	0.036	0.022
	H_1	0.286	0.580	0.788	0.882	0.960
	H_2	0.990	1.000	1.000	1.000	1.000
	H_3	0.012	0.012	0.014	0.016	0.020
	H_4	0.354	0.538	0.600	0.638	0.702
	H_5	0.750	0.954	0.998	1.000	0.996
$a = 1$	H_0	0.022	0.018	0.012	0.014	0.018
	H_1	0.298	0.534	0.708	0.848	0.928
	H_2	0.998	1.000	1.000	1.000	1.000
	H_3	0.012	0.014	0.014	0.016	0.014
	H_4	0.392	0.518	0.604	0.672	0.700
	H_5	0.748	0.936	0.988	1.000	0.998
$a = 0.8$	H_0	0.014	0.012	0.012	0.016	0.010
	H_1	0.282	0.496	0.698	0.800	0.916
	H_2	0.990	0.992	1.000	1.000	1.000
	H_3	0.024	0.018	0.010	0.016	0.024
	H_4	0.398	0.546	0.632	0.678	0.704
	H_5	0.722	0.942	0.994	1.000	1.000

term $\hat{f}_w^6(x)$ in the denominator of $T_n(\theta)$, therefore reduce the computation burden. However, since we have two extra terms to estimate, the computation complexity compared to both previous methods are significantly bigger.

Table 2.3 reports the simulation result using the estimated optimal weight. Comparing to the results in Table 2.1 and 2.2, we can see that the power does increase by a significant amount. However, the test becomes more liberal as evidenced by the inflated empirical levels. Comparing the results presented in Table 2.1, 2.2 and Table 2.3, one can see that the performance of the test with the three chosen weight functions are satisfactory, therefore, in real applications we may focus more on those weight functions that can significantly reduce the computational burden.

From the simulation results, we can also see that the bandwidth selection has non-

Table 2.2: Empirical Levels and Powers ($g(x) = \hat{f}_w^6(x)$)

	n	100	200	300	400	500
$a = 1.5$	H_0	0.020	0.012	0.008	0.016	0.014
	H_1	0.384	0.458	0.764	0.908	0.970
	H_2	0.996	0.998	1.000	1.000	1.000
	H_3	0.018	0.010	0.016	0.020	0.012
	H_4	0.078	0.226	0.320	0.398	0.440
	H_5	0.372	0.768	0.922	0.974	1.000
$a = 1.2$	H_0	0.014	0.018	0.008	0.012	0.016
	H_1	0.374	0.526	0.820	0.870	0.974
	H_2	0.998	1.000	1.000	1.000	1.000
	H_3	0.018	0.020	0.012	0.016	0.018
	H_4	0.202	0.382	0.426	0.510	0.520
	H_5	0.530	0.856	0.96	0.988	0.996
$a = 1$	H_0	0.024	0.012	0.016	0.012	0.020
	H_1	0.256	0.570	0.798	0.900	0.960
	H_2	0.992	0.998	1.000	1.000	1.000
	H_3	0.024	0.028	0.032	0.042	0.030
	H_4	0.256	0.380	0.512	0.578	0.618
	H_5	0.564	0.856	0.954	0.994	0.998
$a = 0.8$	H_0	0.020	0.020	0.014	0.016	0.014
	H_1	0.260	0.570	0.800	0.888	0.954
	H_2	0.998	1.000	1.000	1.000	1.000
	H_3	0.030	0.036	0.022	0.018	0.028
	H_4	0.290	0.476	0.566	0.628	0.670
	H_5	0.568	0.886	0.958	0.986	0.998

negligible effect on the test. It is well known that in hypothesis testing setup how to choose smoothing parameters has been remaining a difficult research topic and not discussed thoroughly in literature. Although a profound discussion on this issue was conducted in Gao and Gijbels (2008), the implementation of such a bandwidth selector is far from trivial and we will not pursue it in this research. A further study on this matter is needed.

Next we conduct a comparison study with the test proposed in Dette et al. (2007). We follow exactly the same simulation set up as in Example 3 in Dette et al. (2007) where two

Table 2.3: Empirical Levels and Powers ($g(x) = \tau^{-4}(x)\delta^4(x)f^2(x)$)

	n	100	200	300	400	500
$a = 1.5$	H_0	0.038	0.034	0.042	0.060	0.044
	H_1	0.676	0.978	1.000	1.000	1.000
	H_2	0.998	1.000	1.000	1.000	1.000
	H_3	0.034	0.044	0.038	0.042	0.036
	H_4	0.340	0.586	0.706	0.790	0.790
	H_5	0.696	0.990	1.000	1.000	1.000
$a = 1.2$	H_0	0.070	0.074	0.062	0.060	0.052
	H_1	0.776	0.992	1.000	1.000	1.000
	H_2	0.996	1.000	1.000	1.000	1.000
	H_3	0.072	0.068	0.072	0.068	0.062
	H_4	0.400	0.686	0.770	0.826	0.852
	H_5	0.784	0.974	1.000	1.000	1.000
$a = 1$	H_0	0.092	0.104	0.084	0.092	0.072
	H_1	0.794	0.986	1.000	1.000	1.000
	H_2	0.996	1.000	1.000	1.000	1.000
	H_3	0.102	0.084	0.098	0.078	0.064
	H_4	0.444	0.700	0.774	0.820	0.868
	H_5	0.79	0.976	0.998	1.000	1.000
$a = 0.8$	H_0	0.106	0.096	0.084	0.102	0.068
	H_1	0.786	0.984	1.000	1.000	1.000
	H_2	0.992	1.000	1.000	1.000	1.000
	H_3	0.138	0.102	0.084	0.080	0.086
	H_4	0.500	0.702	0.804	0.852	0.848
	H_5	0.750	0.976	0.994	1.000	1.000

regression model were considered:

$$(\mathrm{M1}):\ m(x) = 1 + \sin(x), \quad v(x) = \sigma \exp(cx);$$

$$(\mathrm{M2}):\ m(x) = 1 + \sin(2\pi x), \quad v(x) = \sigma(1 + cx)^2,$$

where $X \sim$ Uniform$[0, 1]$ and $\sigma = 0.5$, and the null hypothesis H_0 in both (M1) and (M2) claims that $v(x)$ is a constant ($c = 0$). The variance functions in (M1) and (M2) are the alternative parametric forms to test, where c serves as an adjustment parameter controlling the deviation of the alternative variance function from the null constant variance function. In both models, we use H_0 to denote the null hypothesis $v(x) = \sigma$, corresponding to $c = 0$, H_1 to represent the alternative variance function with $c = 0.5$, and H_2 for $c = 1$. To check if the test is sensitive to the error distribution, same as in Dette et al. (2007), two distributions of ε are considered, $\varepsilon \sim N(0,1)$ and $\varepsilon \sim \sqrt{3}(2U-1)$, where U ~Uniform$[0,1]$.

To reduce the computational burden we choose $g(x) = \hat{f}_w^6(x)$ and the integration range to be $[0,1]$ matching with the support of X. The kernel function is chosen as the Epanechnikov kernel. The simulation results presented in Table 2.4 corresponding to $\varepsilon \sim N(0,1)$, and Table 2.5 corresponds to $\varepsilon \sim \sqrt{3}(2U-1)$, where $U \sim$ Uniform$[0,1]$. The values outside the parentheses are the empirical levels and powers from the proposed test and the ones within the parentheses are from the test in Dette et al. (2007).

Table 2.4: Simulation Results: $\varepsilon \sim N(0,1)$

	n=50			n=100			n=200		
	2.50%	5%	10%	2.50%	5%	10%	2.50%	5%	10%
				Model (M1)					
H_0	0.072	0.080	0.107	0.060	0.073	0.098	0.042	0.062	0.073
	(0.039)	(0.061)	(0.112)	(0.034)	(0.057)	(0.109)	(0.035)	(0.049)	(0.101)
H_1	0.332	0.381	0.460	0.728	0.776	0.822	0.981	0.991	0.993
	(0.157)	(0.195)	(0.284)	(0.335)	(0.388)	(0.484)	(0.459)	(0.558)	(0.684)
H_2	0.940	0.954	0.968	1.000	1.000	1.000	1.000	1.000	1.000
	(0.418)	(0.498)	(0.619)	(0.824)	(0.875)	(0.936)	(0.990)	(0.998)	(1.000)
				Model (M2)					
H_0	0.047	0.070	0.105	0.062	0.080	0.113	0.047	0.076	0.105
	(0.042)	(0.063)	(0.107)	(0.038)	(0.059)	(0.109)	(0.032)	(0.057)	(0.106)
H_1	0.945	0.959	0.972	0.994	0.997	0.999	1.000	1.000	1.000
	(0.288)	(0.366)	(0.501)	(0.635)	(0.713)	(0.818)	(0.934)	(0.971)	(0.978)
H_2	0.998	0.996	1.000	0.999	1.000	1.000	1.000	1.000	1.000
	(0.612)	(0.707)	(0.795)	(0.964)	(0.974)	(0.988)	(1.000)	(1.000)	(1.000)

From Table 2.4 and Table 2.5, we can see that, our test in general performs better in detecting alternative, however, the type I error, as we have observed previously, is not well controlled under the null hypothesis.

We also conduct a comparison study using the setup in Example 4 of Dette et al. (2007). The regression function $m(x) = 1 + \sin(2\pi x)$ and the variance function is taken to be $v(x,\theta) = 1 + \theta x^2 + 2.5c\sin(2\pi x)$. We use (M3) to refer this model. The null hypothesis corresponds to $c = 0$, and the choices of $c = 0.5, 1$ correspond to the alternative variance functions. Again, we use H_1 denote the alternative model with $c = 0.5$, and H_2 with $c = 1$, and the true value of θ is 3. The simulation results are shown in Table 2.6, and one can see

Table 2.5: Simulation Results: $\varepsilon \sim \sqrt{3}(2U-1)$, where $U \sim$Uniform$[0,1]$

	n=50			n=100			n=200		
	2.50%	5%	10%	2.50%	5%	10%	2.50%	5%	10%
	Model (M1)								
H_0	0.012	0.044	0.092	0.031	0.043	0.087	0.044	0.053	0.081
	(0.044)	(0.063)	(0.106)	(0.035)	(0.056)	(0.110)	(0.031)	(0.058)	(0.109)
H_1	0.683	0.726	0.772	0.971	0.984	0.989	1.000	1.000	1.000
	(0.189)	(0.236)	(0.337)	(0.457)	(0.521)	(0.645)	(0.804)	(0.849)	(0.925)
H_2	0.994	0.996	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	(0.532)	(0.609)	(0.738)	(0.970)	(0.987)	(0.995)	(1.000)	(1.000)	(1.000)
	Model (M2)								
H_0	0.032	0.050	0.087	0.047	0.070	0.127	0.065	0.083	0.140
	(0.036)	(0.056	(0.100)	(0.026)	(0.048)	(0.105)	(0.029)	(0.053)	(0.107)
H_1	0.991	0.993	0.996	1.000	1.000	1.000	1.000	1.000	1.000
	(0.474)	(0.558)	(0.663)	(0.901)	(0.923)	(0.968)	(1.000)	(1.000)	(1.000)
H_2	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(0.781)	(0.850)	(0.912)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)	(1.000)

Table 2.6: Simulation Results for Model (M3)

	n=50			n=100			n=200		
	2.50%	5%	10%	2.50%	5%	10%	2.50%	5%	10%
H_0	0.004	0.017	0.028	0.013	0.016	0.018	0.009	0.012	0.015
	(0.042)	(0.067)	(0.119)	(0.038)	(0.061)	(0.109)	(0.036)	(0.052)	(0.095)
H_1	0.054	0.075	0.113	0.258	0.315	0.388	0.753	0.802	0.845
	(0.073)	(0.109)	(0.158)	(0.117)	(0.181)	(0.286)	(0.376)	(0.446)	(0.585)
H_2	0.267	0.347	0.441	0.877	0.913	0.939	1.000	1.000	1.000
	(0.213)	(0.301)	(0.421)	(0.426)	(0.536)	(0.693)	(0.861)	(0.907)	(0.966)

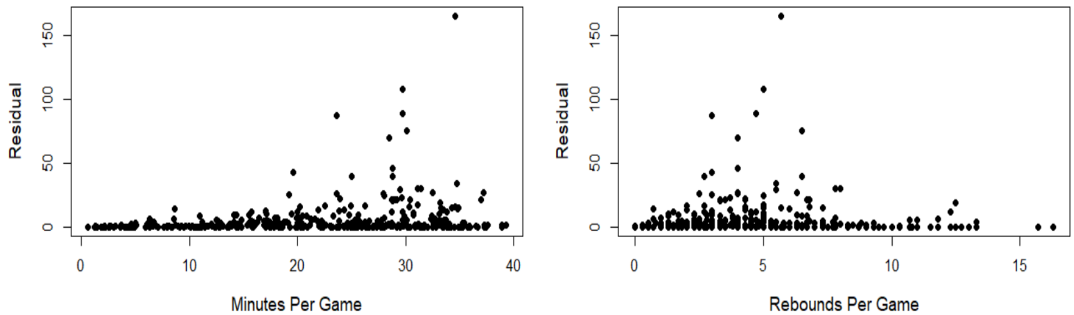
that the minimum distance test proposed in this paper is more conservative than the test from Dette et al. (2007), but it has higher power in detecting the alternatives.

2.4.2 Real Data Application

Now we apply the proposed test procedure to two real data sets.

NBA Player Evaluation: The performance of the Basketball players is widely discussed by both the general audience and the experts. Many statistical models are tried to analyze the relationship between the points each player scored and all other possible metrics. We obtained a data set from the website https://www.nbastuffer.com/2019-2020-nba-player-

Table 2.1: NBA Player Evaluation: Residual Plots



stats which consists of observations from 377 athletes with three variables, including the point earned per game, minutes per game, and rebounds per game from October 22, 2019 to October 30, 2019. In this study, we are particularly interested in how player's performance oscillates with respect to the minutes and rebounds they played per game. For this purpose, we fit a regression model with the point scored as the response variable Y and the minutes per game X_m and rebounds X_r per game as the predictors. We first use the bivariate N-W kernel estimator to estimate the mean regression function, thus obtain the fitted residuals. The following plots show the patterns of the residuals versus each variable X_m and X_r.

Both figures clearly show that the regression model is not homoscedastic, the variability increases as the values of both predictors are getting larger. We would like to check which of the following two parametric forms of the variance function fit the data well:

$$H_{01}: \ v(x) = \beta_1 X_m + \beta_2 X_m^2 + \beta_4 X_r + \beta_5 X_r^2$$

$$H_{02}: \ v(x) = \beta_0 + \beta_1 \sqrt{X_m} + \beta_2 X_r^4.$$

We choose the weight function to be $g(x) = \hat{f}_w^6(x)$ to alleviate computational burden. As indicated in the previous simulation studies, the bandwidth has non-negligible effect on the test. Therefore, instead of selecting a particular one, we implement the test using $h = an^{-1/4}$ for a sequence of a-values from $[0.7, 3]$ and try to make a decision based on the overall conclusions from the test. This technique is also known as "trace plot" in Hart

(2013). Other quantities are chosen as the same as in the previous simulation study. The values for the test statistic versus the bandwidth h are plotted in Figure 2.2 and 2.3 for H_{01} and H_{02}, respectively. From Figure 2.2, we see that all the values of the test statistic are below the 1.96 threshold, thus we do not have enough evidence to reject H_{01}. However, the test statistic values in 2.3 are easily getting over 1.96, hence the variance function in H_{02} is rejected.

Table 2.2: NBA Player Evaluation: Trace Plot for Testing H_{01}

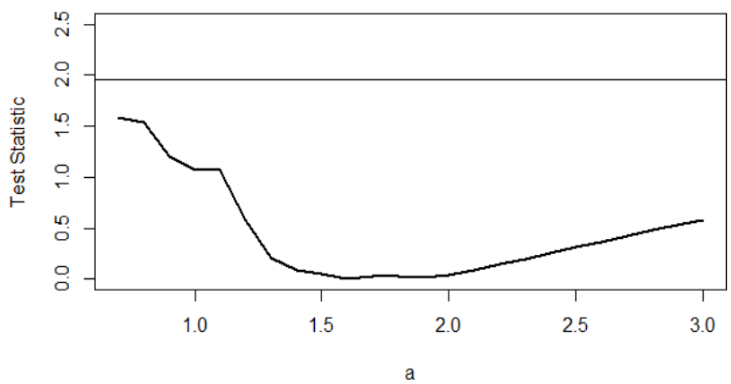
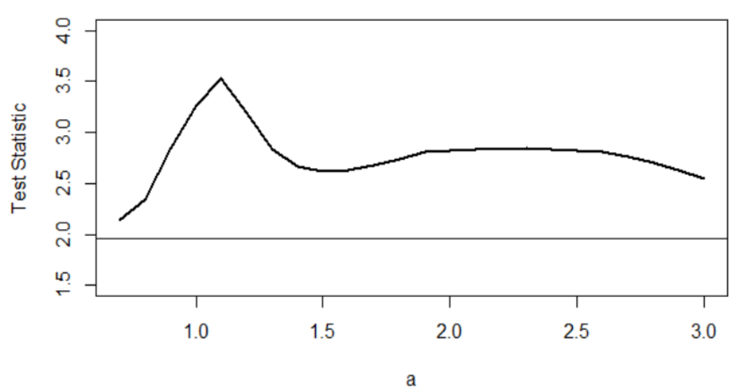


Table 2.3: NBA Player Evaluation: Trace Plot for Testing H_{02}



Note that the proposed test procedure does not exclude its application to some models with discrete response variables. In the following, we apply the proposed test to a real data set with dichotomous response variable.

Traffic Accident Analysis: In a recent research by Newmark. (2020), a data set on vehicle crashes related to police officers in Kansas is analyzed. The data set consists of 2292 crashes data collected from January 1st 2013 to December 31st, 2018 in Kansas (Source: Kansas Crash and Analysis Reporting System). The main research question is to investigate the relationship between damages caused by vehicle crashes related to police officers and the road condition, weather condition, age of police officers, number of traffic units (including motor vehicles, bikes, and pedestrians) etc.. A logistic regression model is used based on $n = 1831$ complete crash records in which the response variable is denoted by Y with $Y = 1$ if the crash causes injury, and $Y = 0$ if property damage only. To see if the logistic regression model is indeed a reasonable choice, we apply the proposed test to check if the conditional variance function has the right form. The predictors are chosen to be the number of traffic units X_1 and the age of the police officers X_2. To be specific, the hypothesis we are testing is

$$H_0 : v(x_1, x_2) = \frac{\exp(\theta_0 + \theta_1 x_1 + \theta_2 x_2)}{[1 + \exp(\theta_0 + \theta_1 x_1 + \theta_2 x_2)]^2}, \quad \text{for some } \theta_0, \theta_1, \theta_2 \in \mathbb{R}, \tag{2.4.1}$$

versus $H_a : H_0$ is not true. The regression function is again estimated with the N-W estimator. The test statistic reports a value of 1.008544, accordingly the p-value is around 0.313. So we cannot reject the null hypothesis. Therefore using the logistic regression to fit the data might be a reasonable choice if the mean function also has the right form.

2.5 Appendix: Proofs of Main Results

In this section, we will present the proofs of all the main results from Section 3. For the sake of brevity, we denote $K_{hi}(x) = K_h(X_i - x), K_{hij} = K_h(X_i - X_j)$, $v_0(x) = v(x, \theta_0)$, $K_{h1,2}(x) = K_{h1}(x)K_{h2}(x)$.

The proof of Theorem 2.3.1. Define

$$\hat{v}_n(x) = \frac{1}{n\hat{f}_w^3(x)} \sum_{i=1}^{n} K_{hi}(x)(Y_i - \hat{m}(X_i))^2 \hat{f}_h^2(X_i), \quad v_{\theta_0}(x) = v_0(x).$$

Then $T_n^*(\theta) = \rho(\hat{v}_n, v_{\theta_0})$, and $\hat{\theta}_n^* = M(\hat{v}_n)$. Based on Lemma 2.3.1, to show the consistency of $\hat{\theta}_n^*$, it suffices to show $T_n^*(\theta_0) \to 0$ as $n \to \infty$.

Adding and subtracting $m(X_i)$ from $Y_i - \hat{m}(X_i)$, $T_n^*(\theta_0)$ can be written as the sum of six terms, and the first three terms are

$$\begin{aligned}
T_{n1}^*(\theta_0) &= \int_{\mathcal{C}} \left\{ \frac{1}{n\hat{f}_w^3(x)} \sum_{i=1}^{n} K_{hi}(x)(Y_i - m(X_i))^2 \hat{f}_h^2(X_i) - v_0(x) \right\}^2 dG(x) \\
T_{n2}^* &= \int_{\mathcal{C}} \left\{ \frac{1}{n\hat{f}_w^3(x)} \sum_{i=1}^{n} K_{hi}(x)(Y_i - m(X_i))(m(X_i) - \hat{m}(X_i))\hat{f}_h^2(X_i) \right\}^2 dG(x) \\
T_{n3}^* &= \int_{\mathcal{C}} \left\{ \frac{1}{n\hat{f}_w^3(x)} \sum_{i=1}^{n} K_{hi}(x)(m(X_i) - \hat{m}(X_i))^2 \hat{f}_h^2(X_i) \right\}^2 dG(x)
\end{aligned}$$

and the last three terms, by Cauchy-Schwartz inequality, are bounded above by $\sqrt{T_{n1}^*(\theta_0)T_{n2}^*}$, $\sqrt{T_{n1}^*(\theta_0)T_{n3}^*}$, and $\sqrt{T_{n2}^*T_{n3}^*}$, respectively. Thus, to show $T_n^*(\theta_0) \to 0$ in probability, it is sufficient to show $T_{n1}^*(\theta_0)$, T_{n2}^* and T_{n3}^* converge to 0.

Adding and subtracting $f^2(X_i)$ from $\hat{f}_h^2(X_i)$, $T_{n1}^*(\theta_0)$ can be further written as a sum of three terms, and the first two terms are

$$\begin{aligned}
T_{n11}^*(\theta_0) &= \int_{\mathcal{C}} \left\{ \frac{1}{n\hat{f}_w^3(x)} \sum_{i=1}^{n} K_{hi}(x)(Y_i - m(X_i))^2 f^2(X_i) - v_0(x) \right\}^2 dG(x) \\
T_{n12}^* &= \int_{\mathcal{C}} \left\{ \frac{1}{n\hat{f}_w^3(x)} \sum_{i=1}^{n} K_{hi}(x)(Y_i - m(X_i))^2 (\hat{f}_h^2(X_i) - f^2(X_i)) \right\}^2 dG(x),
\end{aligned}$$

and the last term, by Cauchy-Schwartz inequality, are bounded above by $\sqrt{T_{n11}^*(\theta_0)T_{n12}^*}$. Thus, to show $T_{n1}^*(\theta_0) \to 0$ in probability, it is sufficient to show $T_{n11}^*(\theta_0)$ and T_{n12}^* converge to 0. Note that $T_{n11}^*(\theta_0)$ can be further bounded above by

$$2\int_{\mathcal{C}}\left\{\frac{1}{n\hat{f}_w(x)}\sum_{i=1}^{n}K_{hi}(x)(Y_i-m(X_i))^2f^2(X_i)-f^2(x)v_0(x)\right\}^2\hat{f}_w^{-4}(x)dG(x)$$
$$+2\int_{\mathcal{C}}\left\{[\hat{f}_w^2(x)-f^2(x)]v_0(x)\right\}^2\hat{f}_w^{-4}(x)dG(x).$$

By treating $(Y-m(X))^2f^2(X)$ as the response variable, similar to the proof of the Corollary 3.1 in Koul and Ni (2004), we can easily show that the first term on the right hand side of the above inequality is the order of $o_p(1)$. The details are omitted here for the sake of brevity. Denote the second term on the right hand side of the above inequality as T^*_{n13}.

Next we show $T^*_{n12}\to 0$ in probability. Recall the notation $\hat{\varphi}_w$, we have

$$\int_{\mathcal{C}}\left\{\frac{1}{n\hat{f}_w^3(x)}\sum_{i=1}^{n}K_{hi}(x)(Y_i-m(X_i))^2(\hat{f}_h^2(X_i)-f^2(X_i))\right\}^2dG(x)$$
$$\le\sup_{x\in\mathcal{C}}\frac{f^6(x)}{\hat{f}_w^6(x)}\cdot\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)(Y_i-m(X_i))^2(\hat{f}_h^2(X_i)-f^2(X_i))\right\}^2d\varphi(x).$$

Furthermore, we have

$$\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)(Y_i-m(X_i))^2(\hat{f}_h^2(X_i)-f^2(X_i))\right\}^2d\varphi(x)$$
$$\le 2\int\left\{\hat{f}_h^2(X_i)\varepsilon_i^2v_0(X_i)\left(\hat{f}_h(X_i)-f(X_i)\right)^2\right\}^2d\varphi(x)$$
$$+8\int\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\varepsilon_i^2v_0(X_i)f(X_i)\left(\hat{f}_h(X_i)-f(X_i)\right)\right\}^2d\varphi(x).$$

Denote the right hand side of the above inequality as $2A_{n1}+8A_{n2}$. For A_{n1}, we have

$$EA_{n1}=\frac{1}{n}\int EK_{h1}^2(x)\varepsilon_1^4v_0^2(X_1)\left(\frac{1}{nh^d}K(0)+\frac{1}{n}\sum_{j=2}^{n}K_{h1j}-f(X_1)\right)^4d\varphi(x)$$
$$+\frac{n-1}{n}\int EK_{h1,2}(x)v_0(X_1)v_0(X_2)\left(\hat{f}_h(X_1)-f(X_1)\right)^2\left(\hat{f}_h(X_2)-f(X_2)\right)^2d\varphi(x).$$

For the first term on the right hand side, we can show that it is bounded above by

$$
\begin{aligned}
&\frac{8K^4(0)}{n^5h^{4d}}\int EK_{h1}^2(x)\varepsilon_1^4v_0^2(X_1)d\varphi(x)+\frac{8}{n}\int EK_{h1}^2(x)\varepsilon_1^4v_0^2(X_1)\left(\frac{1}{n}\sum_{j=2}^{n}K_{h1j}-f(X_1)\right)^4 d\varphi(x)\\
&\le O\left(\frac{1}{n^5h^{5d}}\right)+\frac{8}{n}\int EK_{h1}^2(x)\varepsilon_1^4v_0^2(X_1)\left(\frac{1}{n}\sum_{j=2}^{n}\left(K_{h1j}-E[K_{h12}|X_1]\right)\right)^4 d\varphi(x)\\
&\quad+\frac{8}{n}\int EK_{h1}^2(x)\varepsilon_1^4v_0^2(X_1)\left(E[K_{h12}|X_1]-f(X_1)\right)^4 d\varphi(x)\\
&\le O\left(\frac{1}{n^5h^{5d}}\right)+\frac{8(n-1)}{n^5}\int EK_{h1}^2(x)\varepsilon_1^4v_0^2(X_1)\left(K_{h12}-E[K_{h12}|X_1]\right)^4 d\varphi(x)\\
&\quad+O\left(\frac{1}{n^3}\right)\int EK_{h1}^2(x)\varepsilon_1^4v_0^2(X_1)\left(K_{h12}-E[K_{h12}|X_1]\right)^2\left(K_{h13}-E[K_{h12}|X_1]\right)^2 d\varphi(x)\\
&\quad+\frac{8}{n}\int EK_{h1}^2(x)\varepsilon_1^4v_0^2(X_1)\left(E[K_{h12}|X_1]-f(X_1)\right)^4 d\varphi(x)\\
&=O\left(\frac{1}{n^3h^{3d}}\right)+O\left(\frac{1}{nh^{d-8}}\right).
\end{aligned}
$$

For the second term in EA_{n1}, denote $L_{12h}(x)=K_{h1,2}(x)v_0(X_1)v_0(X_2)$ for the sake of brevity,

$$
\begin{aligned}
&\int EL_{12h}(x)\left(\hat{f}_h(X_1)-f(X_1)\right)^2\left(\hat{f}_h(X_2)-f(X_2)\right)^2 d\varphi(x)\\
&=\int EL_{12h}(x)\left(\frac{1}{nh^d}K(0)+\frac{1}{n}K_{h12}+\frac{1}{n}\sum_{j=3}^{n}K_{h1j}-f(X_1)\right)^2\\
&\qquad\qquad\left(\frac{1}{nh^d}K(0)+\frac{1}{n}K_{h12}+\frac{1}{n}\sum_{j=1,3}^{n}K_{h2j}-f(X_2)\right)^2 d\varphi(x)\\
&\le O\left(\frac{1}{n^4h^{4d}}\right)+O\left(\frac{1}{n^2h^{2d}}\right)\int EL_{12h}(x)\left(\frac{1}{n}\sum_{j=3}^{n}K_{h1j}-f(X_1)\right)^2 d\phi(x)\\
&\quad+4\int EL_{12h}(x)\left(\frac{1}{n}\sum_{j=3}^{n}K_{h1j}-f(X_1)\right)^2\left(\frac{1}{n}\sum_{j=1,3}^{n}K_{h2j}-f(X_2)\right)^2 d\varphi(x)\\
&\le O\left(\frac{1}{n^2h^{2d}}\right)\int EL_{12h}(x)\left(\frac{1}{n}\sum_{j=3}^{n}[K_{h1j}-E(K_{h1j}|X_1)]\right)^2 d\phi(x)+O\left(\frac{1}{n^4h^{4d}}\right)\\
&\quad+O\left(\frac{1}{n^2h^{2d}}\right)\int EL_{12h}(x)\left(\frac{n-2}{n}E(K_{h1j|X_1})-f(X_1)]\right)^2 d\phi(x)
\end{aligned}
$$

$$+16\int EL_{12h}(x)\left(\frac{1}{n}\sum_{j=3}^{n}[K_{h1j}-E(K_{h1j}|X_1)]\right)^2\left(\frac{1}{n}\sum_{j=1,3}^{n}[K_{h2j}-E(K_{h2j}|X_2)]\right)^2 d\varphi(x)$$

$$+16\int EL_{12h}(x)\left(\frac{n-2}{n}E(K_{h1j}|X_1)-f(X_1)\right)^2\left(\frac{1}{n}\sum_{j=1,3}^{n}[K_{h2j}-E(K_{h2j}|X_2)]\right)^2 d\varphi(x)$$

$$+16\int EL_{12h}(x)\left(\frac{n-2}{n}E(K_{h1j}|X_1)-f(X_1)\right)^2\left(\frac{n-2}{n}E(K_{h2j}|X_2)-f(X_2)\right)^2 d\varphi(x)$$

$$=O\left(\frac{1}{n^4h^{4d}}\right)+O\left(\frac{1}{n^2h^{2d-8}}\right)+O(h^8).$$

Therefore,

$$A_{n1}=O_p\left(\frac{1}{n^3h^{3d}}\right)+O_p\left(\frac{1}{nh^{d-8}}\right)+O_p(h^8).$$

Similarly, one can show that

$$A_{n2}=O_p\left(\frac{1}{n^2h^{2d}}\right)+O_p\left(\frac{1}{nh^{d-4}}\right)+O_p(h^4).$$

Hence

$$T^*_{n12}=O_p\left(\frac{1}{n^2h^{2d}}\right)+O_p\left(\frac{1}{nh^{d-4}}\right)+O_p(h^4)=o_p(1).$$

For T^*_{n13}, we have

$$T^*_{n13}=\int_{\mathcal{C}}(\hat{f}^2_w(x)-f^2(x))^2v_0(x)dG(x)=\int_{\mathcal{C}}([\hat{f}_w(x)-f(x)+f(x)]^2-f^2(x))^2v_0(x)dG(x)$$

$$\le 2\int_{\mathcal{C}}(\hat{f}_w(x)-f(x))^4v_0(x)dG(x)+8\int_{\mathcal{C}}(\hat{f}_w(x)-f(x))^2f^2(x)v_0(x)dG(x)$$

$$=O_p\left(\frac{1}{nw^d}\right)+O_p(w^4).$$

In summary, we get $T^*_{n11}=o_p(1)$, hence $T^*_{n1}=o_p(1)$. To show $T^*_{n2}=o_p(1)$, first we note that T^*_{n2} is bounded above by

$$\sup_{x\in\mathcal{C}}\frac{f^6(x)}{\hat{f}^6_w(x)}\cdot\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)(Y_i-m(X_i))(m(X_i)-\hat{m}(X_i))\hat{f}^2_h(X_i)\right\}^2 d\varphi(x).$$

Note that the integral can be written as

$$
\begin{aligned}
&E\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\varepsilon_i\sqrt{v_0(X_i)}(m(X_i)\hat{f}_h(X_i)-\hat{m}(X_i)\hat{f}_h(X_i))\hat{f}_h(X_i)\right\}^2 d\varphi(x)\\
&=\frac{1}{n}\int E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)(m(X_1)\hat{f}_h(X_1)-\hat{m}(X_1)\hat{f}_h(X_1))^2\hat{f}_h^2(X_1)\right]d\varphi(x)\\
&+\frac{(n-1)}{n}\int E\Big[K_{h1}(x)\varepsilon_1\sqrt{v_0(X_1)}(m(X_1)\hat{f}_h(X_1)-\hat{m}(X_1)\hat{f}_h(X_1))\hat{f}_h(X_1)\\
&\qquad K_{h2}(x)\varepsilon_2\sqrt{v_0(X_2)}(m(X_2)\hat{f}_h(X_2)-\hat{m}(X_2)\hat{f}_h(X_2))\hat{f}_h(X_2)\Big]d\varphi(x),
\end{aligned}
$$

and we have

$$
\begin{aligned}
&E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)(m(X_1)\hat{f}_h(X_1)-\hat{m}(X_1)\hat{f}_h(X_1))^2\hat{f}_h^2(X_1)\right]\\
&=E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left(\frac{1}{n}\sum_{j=1}^{n}K_{hj1}[m(X_1)-Y_j]\right)^2\left[\frac{1}{n}\sum_{j=1}^{n}K_{hj1}\right]^2\right]\\
&\le 4E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left(\frac{1}{nh^d}K(0)[m(X_1)-Y_1]\right)^2\left[\frac{1}{nh^d}K(0)\right]^2\right]\\
&+4E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left(\frac{1}{nh^d}K(0)[m(X_1)-Y_1]\right)^2\left[\frac{1}{n}\sum_{j=2}^{n}K_{hj1}\right]^2\right]\\
&+4E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left(\frac{1}{n}\sum_{j=2}^{n}K_{hj1}[m(X_1)-Y_j]\right)^2\left[\frac{1}{nh^d}K(0)\right]^2\right]\\
&+4E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left(\frac{1}{n}\sum_{j=2}^{n}K_{hj1}[m(X_1)-Y_j]\right)^2\left[\frac{1}{n}\sum_{j=2}^{n}K_{hj1}\right]^2\right].
\end{aligned}
$$

Following the above expansion, we may have the following derivation,

$$
\begin{aligned}
&E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left(\frac{1}{nh^d}K(0)[m(X_1)-Y_1]\right)^2\left[\frac{1}{nh^d}K(0)\right]^2\right]=O\left(\frac{1}{n^4h^{5d}}\right),\\
&E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left(\frac{1}{nh^d}K(0)[m(X_1)-Y_1]\right)^2\left[\frac{1}{n}\sum_{j=2}^{n}K_{hj1}\right]^2\right]\\
&=\frac{4K^2(0)}{n^4h^{2d}}\sum_{j=2}^{n}E\left[K_{h1}^2(x)\varepsilon_1^4 v_0^2(X_1)K_{hj1}^2\right]+\frac{4K^2(0)}{n^4h^{2d}}\sum_{j\neq k=2}^{n}E\left[K_{h1}^2(x)\varepsilon_1^4 v_0^2(X_1)K_{hj1}K_{hk1}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{4(n-1)K^2(0)}{n^4h^{2d}} E\left[K_{h1}^2(x)\varepsilon_1^4 v_0^2(X_1) K_{h21}^2\right] + O\left(\frac{1}{n^2h^{2d}}\right) E\left[K_{h1}^2(x)\varepsilon_1^4 v_0^2(X_1) K_{h21} K_{h31}\right] \\
&= O\left(\frac{1}{n^3h^{4d}}\right) + O\left(\frac{1}{n^2h^{3d}}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left(\frac{1}{n}\sum_{j=2}^n K_{hj1}[m(X_1)-Y_j]\right)^2\left[\frac{1}{nh^d}K(0)\right]^2\right] \\
&= \frac{K^2(0)}{n^2h^{2d}} E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left[\frac{1}{n}\sum_{j=2}^n K_{hj1}[m(X_1)-m(X_j)]\right]^2\right] \\
&\quad + \frac{K^2(0)}{n^2h^{2d}} E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left[\frac{1}{n}\sum_{j=2}^n K_{hj1}\sqrt{v(X_j)}\varepsilon_j\right]^2\right] \\
&= \frac{(n-1)K^2(0)}{n^4h^{2d}} E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1) K_{h21}^2 [m(X_1)-m(X_2)]^2\right] \\
&\quad + \frac{(n-1)(n-2)K^2(0)}{n^4h^{2d}} E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1) K_{h21} K_{h21} [m(X_1)-m(X_2)][m(X_1)-m(X_3)]\right] \\
&\quad + \frac{(n-1)K^2(0)}{n^4h^{2d}} E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1) K_{h21}^2 v_0(X_2)\varepsilon_2^2\right] \\
&= O\left(\frac{1}{n^2h^{3d-4}}\right) + O\left(\frac{1}{n^3h^{4d}}\right) = o(1).
\end{aligned}
$$

Similarly, we can show the following results,

$$
\begin{aligned}
&E\left[K_{h1}^2(x)\varepsilon_1^2 v_0(X_1)\left(\frac{1}{n}\sum_{j=2}^n K_{hj1}[m(X_1)-Y_j]\right)^2\left[\frac{1}{n}\sum_{j=2}^n K_{hj1}\right]^2\right] = O\left(\frac{1}{h^{d-4}}\right). \\
&E[K_{h1}(x)\varepsilon_1\sqrt{v_0(X_1)}(m(X_1)\hat{f}_h(X_1) - \hat{m}(X_1)\hat{f}_h(X_1))\hat{f}_h(X_1)\cdot \\
&\qquad K_{h2}(x)\varepsilon_2\sqrt{v_0(X_2)}(m(X_2)\hat{f}_h(X_2) - \hat{m}(X_2)\hat{f}_h(X_2))\hat{f}_h(X_2)] \\
&= E\Big[K_{h1}(x)\varepsilon_1\sqrt{v_0(X_1)}\Big[\frac{1}{n}K_h(0)\varepsilon_1\sqrt{v_0(X_1)} + \frac{1}{n}K_{h21}(Y_2 - m(X_1))\Big]\Big]\hat{f}_h(X_1) \\
&\quad K_{h2}(x)\varepsilon_2\sqrt{v_0(X_2)}\left[\frac{1}{nh^d}K(0)\varepsilon_2\sqrt{v_0(X_2)} + \frac{1}{n}K_{h21}[Y_1 - m(X_2)]\right]\hat{f}_h(X_2)]
\end{aligned}
$$

$$= \frac{K^2(0)}{n^2h^{2d}} E\left[K_{h1}(x)K_{h2}v_0(X_1)v_0(X_2)\hat{f}_h(X_1)\hat{f}_h(X_2)\right]$$
$$+\frac{1}{n^2}E\left[K_{h1,2}(x)K_{h12}^2v_0(X_1)v_0(X_2)\hat{f}_h(X_1)\hat{f}_h(X_2)\right]$$

which has the order of $O\left(\frac{1}{n^2h^{2d}}\right)$. Therefore, we eventually get

$$\sup_{x\in\mathcal{C}} \frac{f^6(x)}{\hat{f}_w^6(x)} \cdot \int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)(Y_i - m(X_i))(m(X_i) - \hat{m}(X_i))\hat{f}_h^2(X_i) \right\}^2 d\varphi(x)$$
$$= O\left(\frac{1}{nh^{d-4}}\right) + O\left(\frac{1}{n^2h^{2d}}\right),$$

and this implies $T_{n2}^* = o_p(1)$.

Finally, let's show that $T_{n3}^* = o_p(1)$. Note that

$$T_{n3}^* \leq \sup_{x\in\mathcal{C}} \frac{f^6(x)}{\hat{f}_w^6(x)} \cdot \int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)[m(X_i) - \hat{m}(X_i)]^2 \hat{f}_h^2(X_i) \right\}^2 d\varphi(x).$$

We also have

$$E\int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)[m(X_i) - \hat{m}(X_i)]^2 \hat{f}_h^2(X_i) \right\}^2 d\varphi(x)$$
$$= E\int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \left[\frac{1}{n} \sum_{j=1}^{n} K_{hji}[m(X_i) - Y_j] \right]^2 \right\}^2 d\varphi(x)$$
$$= \frac{1}{n} \int_{\mathcal{C}} E\left\{ K_{h1}^2(x) \left[\frac{1}{n} \sum_{j=1}^{n} K_{hj1}[m(X_1) - Y_j] \right]^4 \right\} d\varphi(x)$$
$$+\frac{n-1}{n} \int_{\mathcal{C}} E\left\{ K_{h1,2}(x) \left[\frac{1}{n} \sum_{l=1}^{n} K_{hl1}[m(X_1) - Y_l] \right]^2 \left[\frac{1}{n} \sum_{l=1}^{n} K_{hl2}[m(X_2) - Y_l] \right]^2 \right\} d\varphi(x).$$

Note that with C-R inequality

$$
\begin{aligned}
&\frac{1}{n}\int_{\mathcal{C}} E\left\{K_{h1}^2(x)\left[\frac{1}{n}\sum_{j=1}^n K_{hj1}[m(X_1)-Y_j]\right]^4\right\} d\varphi(x) \\
\leq &\frac{8}{n}\int_{\mathcal{C}} E\left\{K_{h1}^2(x)\left[\frac{1}{n}\sum_{j=1}^n K_{hj1}\varepsilon_j\sqrt{v(X_j)}\right]^4\right\} d\varphi(x) \\
&+\frac{8}{n}\int_{\mathcal{C}} E\left\{K_{h1}^2(x)\left[\frac{1}{n}\sum_{j=2}^n K_{hj1}[m(X_1)-m(X_j)]\right]^4\right\} d\varphi(x) = O\left(\frac{1}{n^5h^{5d}}\right)+O\left(\frac{1}{nh^{d-8}}\right),
\end{aligned}
$$

$$
\begin{aligned}
&\int_{\mathcal{C}} E\left\{K_{h1,2}(x)\left[\frac{1}{n}\sum_{l=1}^n K_{hl1}[m(X_1)-Y_l]\right]^2\left[\frac{1}{n}\sum_{l=1}^n K_{hl2}[m(X_2)-Y_l]\right]^2\right\} d\varphi(x) \\
\leq\ &4\int_{\mathcal{C}} E\left\{K_{h1,2}(x)\left[\frac{1}{n}\sum_{l=1}^n K_{hl1}[m(X_1)-m(X_l)]\right]^2\left[\frac{1}{n}\sum_{l=1}^n K_{hl2}[m(X_2)-m(X_l)]\right]^2\right\} d\varphi(x) \\
&+4\int_{\mathcal{C}} E\left\{K_{h1,2}(x)\left[\frac{1}{n}\sum_{l=1}^n K_{hl1}\varepsilon_l\sqrt{v(X_l)}\right]^2\left[\frac{1}{n}\sum_{l=1}^n K_{hl2}[m(X_2)-m(X_l)]\right]^2\right\} d\varphi(x) \\
&+4\int_{\mathcal{C}} E\left\{K_{h1,2}(x)\left[\frac{1}{n}\sum_{l=1}^n K_{hl1}[m(X_1)-m(X_l)]\right]^2\left[\frac{1}{n}\sum_{l=1}^n K_{hl2}\varepsilon_l\sqrt{v(X_l)}\right]^2\right\} d\varphi(x) \\
&+4\int_{\mathcal{C}} E\left\{K_{h1,2}(x)\left[\frac{1}{n}\sum_{l=1}^n K_{hl1}\varepsilon_l\sqrt{v(X_l)}\right]^2\left[\frac{1}{n}\sum_{l=1}^n K_{hl2}\varepsilon_l\sqrt{v(X_l)}\right]^2\right\} d\varphi(x) \\
=\ &O(h^8)+O\left(\frac{1}{n^2h^{2d-4}}\right)+O\left(\frac{1}{n^4h^{4d}}\right).
\end{aligned}
$$

Hence

$$
T_{n3}^* = O_p\left(\frac{1}{n^5h^{5d}}\right)+O_p\left(\frac{1}{nh^{d-8}}\right)+O_p(h^8)+O_p\left(\frac{1}{n^2h^{2d-4}}\right)+O_p\left(\frac{1}{n^4h^{4d}}\right)=o_p(1).
$$

□

The proof of Theorem 2.3.2. Similar to Koul and Ni (2004), it suffices to show

$$\sup_\theta |T_n(\theta) - T_n^*(\theta)| = o_p(1).$$

Adding and subtracting $v(X_i,\theta)$ from $(Y_i - \hat{m}(X_i))^2$ inside the integrand of $T_n^*(\theta)$, $T_n^*(\theta)$ can be written as the sum of $T_n(\theta)$, $C_n(\theta)$ which is defined by

$$C_n(\theta) = \int_{\mathcal{C}} \left[\frac{\sum_{i=1}^n K_{hi}(x) v(X_i,\theta) \hat{f}_h^2(X_i)}{n \hat{f}_w^3(x)} - v(x,\theta) \right]^2 dG(x),$$

and another term which is bounded above in absolute value by $2(C_n(\theta)T_n(\theta))^{1/2}$. Therefore,

$$\sup_\theta |T_n^*(\theta) - T_n(\theta)| \le \sup_\theta C_n(\theta) + 2\sup_\theta (C_n(\theta)T_n(\theta))^{1/2}$$

It thus suffices to show that $\sup_\theta C_n(\theta) = o_p(1)$, and $\sup_\theta T_n(\theta) = O_p(1)$. Note that

$$\begin{aligned}
C_n(\theta) &= \int_{\mathcal{C}} \left[\frac{n^{-1}\sum_{i=1}^n K_{hi}(x) v(X_i,\theta) \hat{f}_h^2(X_i)}{\hat{f}_w^3(x)} - v(x,\theta) \right]^2 dG(x) \\
&= \int_{\mathcal{C}} \left[\frac{n^{-1}\sum_{i=1}^n K_{hi}(x) v(X_i,\theta) \hat{f}_h^2(X_i)}{\hat{f}_w(x)} - (\hat{f}_w^2(x) - f^2(x) + f^2(x)) v(x,\theta) \right]^2 \hat{f}_w^{-4}(x) dG(x) \\
&\le 4 \int_{\mathcal{C}} \left[\frac{n^{-1}\sum_{i=1}^n K_{hi}(x) v(X_i,\theta) f^2(X_i)}{\hat{f}_w(x)} - f^2(x) v(x,\theta) \right]^2 \hat{f}_w^{-4}(x) dG(x) \\
&\quad +4 \int_{\mathcal{C}} \left[\frac{n^{-1}\sum_{i=1}^n K_{hi}(x) v(X_i,\theta) (\hat{f}_h^2(X_i) - f^2(X_i))}{\hat{f}_w(x)} \right]^2 \hat{f}_w^{-4}(x) dG(x) \\
&\quad +2 \int_{\mathcal{C}} [(\hat{f}_w^2(x) - f^2(x)) v(x,\theta)]^2 \hat{f}_w^{-4}(x) dG(x) \\
&:= 4C_{n1}(\theta) + 4C_{n2}(\theta) + 2C_{n3}(\theta)
\end{aligned}$$

For $C_{n1}(\theta)$, we have

$$C_{n1}(\theta) = \int_{\mathcal{C}} \left[\frac{\frac{1}{n}\sum_{i=1}^n K_{hi}(x) v(X_i,\theta) f^2(X_i)}{\hat{f}_w(x)} - f^2(x) v(x,\theta) \right]^2 \hat{f}_w^{-4}(x) dG(x)$$

$$
\begin{aligned}
&\leq \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) v(X_i, \theta) f^2(X_i) - \hat{f}_w(x) f^2(x) v(x, \theta) \right]^2 \hat{f}_w^{-6}(x) dG(x) \\
&\leq \sup_{x \in \mathcal{C}} \frac{f^6(x)}{\hat{f}_w^6(x)} \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) v(X_i, \theta) f^2(X_i) - \hat{f}_w(x) f^2(x) v(x, \theta) \right]^2 d\varphi(x).
\end{aligned}
$$

Treating $f^2(x)v(x,\theta)$ as $m(x,\theta)$ and using the same arguments as in Koul and Ni (2014), one can show that $\sup_\theta C_{n1}(\theta) = o_p(1)$.

To show that $\sup_\theta C_{n2}(\theta) = o_p(1)$, first we note that for any fixed $\theta \in \Theta$, $C_{n2}(\theta) = o_p(1)$ which can be proved in the same manner as showing $T^*_{n12} = o_p(1)$ in the proof of Theorem 2.3.1, so it suffices to show that $C_{n2}(\theta)$ is uniformly continuous in $\theta \in \Theta$. Adding and subtracting $v(X_i, \theta_2)$, we have

$$
\begin{aligned}
C_{n2}(\theta_1) &= \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) v(X_i, \theta_1) (\hat{f}_h^2(X_i) - f^2(X_i)) \right]^2 \hat{f}_w^{-6}(x) dG(x) \\
&= \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) (v(X_i, \theta_1) - v(X_i, \theta_2) + v(X_i, \theta_2)) (\hat{f}_h^2(X_i) - f^2(X_i)) \right]^2 d\hat{\varphi}_w(x).
\end{aligned}
$$

Thus $|C_{n2}(\theta_1) - C_{n2}(\theta_2)|$ is bounded above by

$$
\begin{aligned}
&\sup_{x \in \mathcal{C}} \left| \frac{f^6(x)}{\hat{f}_w^6(x)} \right| \left| \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) (v(X_i, \theta_1) - v(X_i, \theta_2)) (\hat{f}_h^2(X_i) - f^2(X_i)) \right]^2 d\varphi(x) \right. \\
&+2 \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) (v(X_i, \theta_1) - v(X_i, \theta_2)) (\hat{f}_h^2(X_i) - f^2(X_i)) \right] \\
&\qquad\qquad \left. \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) v(X_i, \theta_2) (\hat{f}_h^2(X_i) - f^2(X_i)) \right] d\varphi(x) \right|
\end{aligned}
$$

By assumption (v3), $|C_{n2}(\theta_1) - C_{n2}(\theta_2)|$ is further bounded above by

$$
\|\theta_2 - \theta_1\|^2 \sup_{x \in \mathcal{C}} \left| \frac{f^6(x)}{\hat{f}_w^6(x)} \right| \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) |\hat{f}_h^2(X_i) - f^2(X_i)| l(X_i) \right]^2 d\varphi(x)
$$

$$+\|\theta_2-\theta_1\|\sup_{x\in\mathcal{C}}\left|\frac{f^6(x)}{\hat{f}_w^6(x)}\right|\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)|\hat{f}_h^2(X_i)-f^2(X_i)|l(X_i)\right]$$
$$\cdot\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)|\hat{f}_h^2(X_i)-f^2(X_i)|v(X_i,\theta_2)\right]d\varphi(x) \qquad (2.5.1)$$

Note that for a fixed $\theta_3\in\Theta$ which is compact, there exist a non-negative function L such that

$$|v(\theta_2,X_i)-v(\theta_3,X_i)+v(\theta_3,X_i)|\le\|\theta_2-\theta_3\|l(X_i)+v(\theta_3,X_i)\le L(X_i)$$

With similar argument in T^*_{n2}, we know the integration on the right hand side (RHS) of (2.5.1) is the order of $o_p(1)$. Thus it suffices to show that

$$\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)|\hat{f}_h^2(X_i)-f^2(X_i)|l(X_i)\right]\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)|\hat{f}_h^2(X_i)-f^2(X_i)|L(X_i)\right]d\varphi(x)$$

has the order of $o_p(1)$. Taking expectation, we have

$$E\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)|\hat{f}_h^2(X_i)-f^2(X_i)|l(X_i)\right]\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)|\hat{f}_h^2(X_i)-f^2(X_i)|L(X_i)\right]d\varphi(x)$$
$$=E\int_{\mathcal{C}}\left[\frac{1}{n^2}\sum_{i=1}^{n}K_{hi}^2(x)|\hat{f}_h^2(X_i)-f^2(X_i)|^2l(X_i)L(X_i)\right]$$
$$+\left[\frac{2}{n^2}\sum_{i\le j}K_{hi,j}(x)L(X_i)l(X_j)|\hat{f}_h^2(X_i)-f^2(X_i)||\hat{f}_h^2(X_j)-f^2(X_j)|\right]d\varphi(x):=C_1+C_2$$

Note that

$$C_1=\frac{1}{n}\int_{\mathcal{C}}E\left[K_{h1}^2(x)|\hat{f}_h^2(X_1)-f^2(X_1)|^2l(X_1)L(X_1)\right]d\varphi(x).$$

Recall A_{n1}, with exact same technique, adding and subtracting $f(x)$ inside $\hat{f}_h^2(x)$, and then adding and subtracting $E(K_h(X_2-X_1)|X_1)$ in the sum, while functions $K,\ l,\ L,\ f$ are continuous and bounded, we get $C_1=o(1)$

Next by Cauchy-Schwartz Inequality, C_2 is bounded above by

$$c\int_{\mathcal{C}}\sqrt{EK_{h1}^2(x)L^2(X_1)(\hat{f}_h^2(X_1)-f^2(X_1))^2}\sqrt{EK_{h2}^2(x)L^2(X_2)(\hat{f}_h^2(X_2)-f^2(X_2))^2}d\varphi(x)$$

for a constant c. Again, with similar technique applied to A_{n1}, and $nh^{2d}\to\infty$, $C_2=o(1)$. Together we know that the second part is $o_p(1)$, and this proves uniform continuity of $C_{n2}(\theta)$.

Since $v(x,\theta)$ is bounded on the compact set $\mathcal{C}\times\Theta$, so $\sup_{\theta\in\Theta}C_{n3}(\theta)=o_p(1)$ is straightforward.

Next, adding and subtracting $v_0(X_i)$ in $T_n(\theta)$, we obtain

$$\begin{aligned}T_n(\theta)\le 2\sup_{x\in\mathcal{C}}\frac{f^6(x)}{\hat{f}_w^6(x)}\Bigg\{&\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)[v_0(X_i)-v(X_i,\theta)]\hat{f}_h^2(X_i)\right]^2 d\varphi(x)\\&+\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)[(Y_i-\hat{m}(X_i))^2-v_0(X_i)]\hat{f}_h^2(X_i)\right]^2 d\varphi(x)\Bigg\}\end{aligned}$$

Note that the second term on the RHS of the above expression is indeed $T_n(\theta_0)$ which is $o_p(1)$. Finally we can show that for a positive constant c,

$$\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)[v_0(X_i)-v(X_i,\theta)]\hat{f}_h^2(X_i)\right]^2 d\varphi(x)\le c\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)l(X_i)\hat{f}_h^2(X_i)\right]^2 d\varphi(x)$$

which is $O_p(1)$ by the compactness of Θ. This concludes proof of $\sup_{\theta\in\Theta}T_n(\theta)=O_p(1)$, hence the proof of Theorem 2.3.2. □

To show the asymptotic normality of $\hat{\theta}_n$, we introduce the following notations, for the sake of convenience, and they maybe used in the context that follows.

$$\dot{\mu}_n(x,\theta):=\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\theta)f^2(X_i),\ \dot{\mu}_n(x,\theta_0):=\dot{\mu}_n(x)$$
$$\dot{\mu}_h(x):=E\dot{\mu}_n(x,\theta_0)=EK_h(x-X)\dot{v}(\theta_0,X)f^2(X)$$
$$S_n:=\int U_n(x)\dot{\mu}_n(x)d\varphi(x),\ \Sigma_n(b)=\int[b^T\dot{\mu}_n(x,\theta_0)]^2d\varphi(x)$$

$$
\begin{aligned}
U_n(x,\theta) &= \frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)[(Y_i-\hat{m}(X_i))^2 - v(X_i,\theta)]^2 \hat{f}_h^2(X_i),\ U_n(x) := U_n(x,\theta_0)\\
D_n(\theta) &:= \int Z_n^2(x,\theta)d\hat{\varphi}_w(x) := \int \Big[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)(v(X_i,\theta_0)-v(X_i,\theta))\hat{f}_h^2(X_i)\Big]^2 d\hat{\varphi}_w(x)\\
\xi(X_i) &:= (Y_i - m(X_i))^2 - v_0(X_i).
\end{aligned}
$$

We will use $\Delta_{ig} := \hat{g}(X_i) - g(X_i)$ to denote any difference between the function at sample point X_i and its approximation at point X_i. The approximation methods are either kernel smoothing estimates for densities and N-W estimates for regression functions.

The proof of Theorem 2.3.3. First show $nh^d\|\hat{\theta}_n - \theta_0\|^2 = O_P(1)$. To proceed, we could show that $nh^d D_n(\hat{\theta}_n) = O_p(1)$. Note that by adding and subtracting $(Y_i - \hat{m}(X_i))^2$, we can see that $D_n(\hat{\theta}_n)$ is bounded above by $2[T_n(\theta_0) + T_n(\hat{\theta}_n)]$, where

$$
T_n(\theta) = \int \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)((Y_i-\hat{m}(X_i))^2 - v(X_i,\theta))\hat{f}_h^2(X_i)\right]^2 d\hat{\varphi}_w(x).
$$

It suffices to show that $nh^d T_n(\theta_0) = O_p(1)$, together by definition, $\hat{\theta}_n = \operatorname{argmin}_{\theta\in\Theta} T_n(\theta)$, we know that $nh^d T_n(\hat{\theta}_n) = O_p(1)$, and therefore $nh^d\|\hat{\theta}_n - \theta_0\|^2 = O_P(1)$. Adding and subtracting $m(X_i)$

$$
\begin{aligned}
nh^d T_n(\theta_0) &= nh^d \int \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)((Y_i-\hat{m}(X_i))^2 - v_0(X_i))\hat{f}_h^2(X_i)\right]^2 d\hat{\varphi}_w(x)\\
&\le 3nh^d \int \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)((Y_i-m(X_i))^2 - v_0(X_i))\hat{f}_h^2(X_i)\right]^2 d\hat{\varphi}_w(x)\\
&\quad +6nh^d \int \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)((Y_i-m(X_i))(m(X_i)-\hat{m}(X_i))\hat{f}_h^2(X_i)\right]^2 d\hat{\varphi}_w(x)\\
&\quad +3nh^d \int \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)(m(X_i)-\hat{m}(X_i))^2)\hat{f}_h^2(X_i)\right]^2 d\hat{\varphi}_w(x)\\
&:= 3nh^d T_{n1}(\theta_0) + 6nh^d T_{n2}(\theta_0) + 3nh^d T_{n3}(\theta_0)
\end{aligned}
$$

Note that:

$$
\begin{aligned}
T_{n1}(\theta_0) &= \int \left[\frac{1}{n^2} \sum_{i=1}^{n} K_{hi}(x)(\varepsilon_i^2 - 1) v_0(X_i) \hat{f}_h^2(X_i) \right]^2 d\hat{\varphi}_w(x) \\
&= \int \frac{1}{n^2} \sum_{i=1}^{n} K_{hi}^2(x)(\varepsilon_i^2 - 1)^2 v_0^2(X_i) \hat{f}_h^4(X_i) d\hat{\varphi}_w(x) \\
&\quad + \int \frac{1}{n^2} \sum_{i \neq j}^{n} K_{hi}(x) K_{hj}(x)(\varepsilon_i^2 - 1)(\varepsilon_j^2 - 1) v_0(X_i) v_0(X_j) \hat{f}_h^2(X_i) \hat{f}_h^2(X_j) d\hat{\varphi}_w(x).
\end{aligned}
$$

Therefore, by Fubini Theorem,

$$
\begin{aligned}
&ET_{n1}(\theta_0) = \int \frac{1}{n} E[K_{h1}^2(x) E((\varepsilon_1^2 - 1)^2 | X_1) v_0^2(X_1) \hat{f}_h^4(X_1)] d\hat{\varphi}_w(x) \\
&+ \int \frac{n(n-1)}{n^2} E[K_{h1,2}(x) E((\varepsilon_1^2 - 1)|X_1) E((\varepsilon_2^2 - 1)|X_2) v_0(X_1) v_0(X_2) \hat{f}_h^2(X_1) \hat{f}_h^2(X_2)] d\hat{\varphi}_w(x)
\end{aligned}
$$

which is the order of $O\left(\frac{1}{nh^d}\right)$. The second term equals to zero under null assumption and the independence between any paired $X_i's$, therefore we have $nh^d T_{n1}(\theta_0) = O_p(1)$. Also it is easy to see that $T_{n2}(\theta_0) = T_{n2}^*(\theta_0)$, and $T_{n3}(\theta_0) = T_{n3}^*(\theta_0)$. Utilizing the previous result, with only the restriction that $d < 8$, which satisfies assumption (h1) and (h2), we could see that $nh^d T_n(\theta_0) = O_p(1)$, thus leading to $nh^d D_n(\hat{\theta}_n) = O_p(1)$.

Next we shall show that $\forall\ 0 < a < \infty$, $\exists N_a$, such that

$$
P\left(\frac{D_n(\hat{\theta}_n)}{\|\hat{\theta}_n - \theta_0\|^2} \geq a + \inf_{\|b\|=1} b^T \Sigma_0 b \right) > 1 - a \tag{2.5.2}
$$

The claim will follow with the $nh^d D_n(\hat{\theta}_n) = O_p(1)$, and the fact that Σ_0 is positive definite. Let $u_n := (\hat{\theta}_n - \theta_0), d_{ni} := v(\hat{\theta}_n, X_i) - v(X_i, \theta_0) - u_n^T \dot{v}(X_i, \theta_0), i = 1, 2, \ldots, n$. Then $D_n(\hat{\theta}_n)/\|\hat{\theta}_n - \theta_0\|^2$ can be written as

$$
\begin{aligned}
&\int \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \left(\frac{\dot{v}(X_i, \theta_0) - \dot{v}(\hat{\theta}_n, X_i) + u_n^T \dot{v}(X_i, \theta_0) - u_n^T \dot{v}(X_i, \theta_0)}{\|\hat{\theta}_n - \theta_0\|} \right) \hat{f}_h^2(X_i) \right]^2 d\hat{\varphi}_w(x) \\
&= \int \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \left(\frac{\dot{v}(X_i, \theta_0) - \dot{v}(\hat{\theta}_n, X_i) - u_n^T \dot{v}(X_i, \theta_0)}{\|\hat{\theta}_n - \theta_0\|} \right) \hat{f}_h^2(X_i) \right]^2 d\hat{\varphi}_w(x)
\end{aligned}
$$

$$
\begin{aligned}
&+\int \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \left(\frac{u_n^T \dot{v}(X_i, \theta_0)}{\|\hat{\theta}_n - \theta_0\|} \right) \hat{f}_h^2(X_i) \right]^2 d\hat{\varphi}_w(x) \\
&+2 \int \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \left(\frac{d_{ni}}{\|u_n\|} \right) \hat{f}_h^2(X_i) \right] \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \left(\frac{u_n^T \dot{v}(X_i, \theta_0)}{\|u_n\|} \right) \hat{f}_h^2(X_i) \right] d\hat{\varphi}_w(x) \\
:= & \int \bar{D}_{n1}^2 d\hat{\varphi}_w(x) + \int \bar{D}_{n2}^2 d\hat{\varphi}_w(x) + 2 \int \bar{D}_{n1} \bar{D}_{n2} d\hat{\varphi}_w(x) \\
\geq & \int \bar{D}_{n1}^2 d\hat{\varphi}_w(x) + \int \bar{D}_{n2}^2 d\hat{\varphi}_w(x) - 2 \sqrt{\int \bar{D}_{n1}^2 d\hat{\varphi}_w(x) \int \bar{D}_{n2}^2 d\hat{\varphi}_w(x)} \\
:= & D_{n1} + D_{n2} - 2\sqrt{D_{n1} D_{n2}}.
\end{aligned}
$$

From the above derivation, it suffices to show that D_{n2} is bounded above, bounded below by $\inf_{\|b\|=1} \Sigma_n(b)$, and $D_{n1} = o_p(1)$. By Cauchy-Schwartz inequality,

$$
D_{n2} = \int \left[\frac{u_n^T \dot{\mu}_n(x, \theta_0)}{\|u_n\|} \right]^2 d\hat{\varphi}_w(x) \leq \int \dot{\mu}_n^T(x, \theta_0) \dot{\mu}_n(x, \theta_0) d\hat{\varphi}_w(x).
$$

With K, $\dot{v}$, f bounded above, we know that D_{n2} is bounded above. By assumption (v4), and the fact that $\hat{\theta}_n$ is a consistent estimator of θ, we know $D_{n1} = o_p(1)$.

Next we show $D_{n2} \geq \inf_{\|b\|=1} \Sigma_n(b)$, and $\Sigma_n(b) \to b^T \Sigma b$, for any q dimension vector b.

$$
\begin{aligned}
& \int \left[b^T \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \dot{v}(X_i, \theta_0) \hat{f}_h^2(X_i) \right]^2 d\hat{\varphi}_w(x) \\
= & b^T \int \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \dot{v}(X_i, \theta_0) \hat{f}_h^2(X_i) \right] \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \dot{v}(X_i, \theta_0) \hat{f}_h^2(X_i) \right]^T d\hat{\varphi}_w(x) b.
\end{aligned}
$$

Plus and minus $f^2(X_i)$ after $\hat{f}_h^2(X_i)$ gives the following,

$$
\begin{aligned}
& \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \dot{v}(X_i, \theta_0) \hat{f}_h^2(X_i) \right] \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \dot{v}(X_i, \theta_0) \hat{f}_h^2(X_i) \right]^T \\
= & \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \dot{v}(X_i, \theta_0) (\hat{f}_h^2(X_i) - f^2(X_i)) \right] \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \dot{v}(X_i, \theta_0) (\hat{f}_h^2(X_i) - f^2(X_i)) \right]^T
\end{aligned}
$$

$$+2\left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\dot{v}(X_i,\theta_0)(\hat{f}_h^2(X_i)-f^2(X_i))\right]\left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\dot{v}(X_i,\theta_0)f^2(X_i)\right]^T$$
$$+\left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\dot{v}(X_i,\theta_0)f^2(X_i)\right]\left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\dot{v}(X_i,\theta_0)f^2(X_i)\right]^T.$$

As proved in A_{n1} and A_{n2}, with $\dot{v}$ continuous and bounded, we know the first two term on the RHS goes to zero in probability. Also note that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\dot{v}(X_i,\theta_0)f^2(X_i) = f^3(x)\dot{v}(x,\theta)$$

and the fact that $\sup_x |f^6(x)\hat{f}_w^{-6}(x)-1| = o_p(1)$, this concludes $\Sigma_n(b)\to b^T\Sigma b$.

Now we will show that $\Sigma_n(b)$ is uniform continuous in b. For any $\delta>0$ and any two unit vectors b_1, b_2, such that $\|b_1-b_2\|\le\delta$, we see that $|\Sigma_n(b_1)-\Sigma_n(b_2)|$ is bounded above by

$$\sup_{x\in\mathcal{C}}\left|\frac{f^6(x)}{\hat{f}_w^6(x)}\right|\left|\int_{\mathcal{C}}[b_1^T\dot{\mu}_n(x,\theta_0)]^2 d\varphi(x) - \int_{\mathcal{C}}[b_2^T\dot{\mu}_n(x,\theta_0)]^2 d\varphi(x)\right|$$
$$=\sup_{x\in\mathcal{C}}\left|\frac{f^6(x)}{\hat{f}_w^6(x)}\right|\left|\int_{\mathcal{C}}[(b_1-b_2)^T\dot{\mu}_n(x,\theta_0)]^2 d\varphi(x) + 2\int_{\mathcal{C}}[(b_1-b_2)^T\dot{\mu}_n(x,\theta_0)][b_2^T\dot{\mu}_n(x,\theta_0)]d\varphi(x)\right|$$
$$\le\delta(\delta+2)\sup_{x\in\mathcal{C}}\left|\frac{f^6(x)}{\hat{f}_w^6(x)}\right|\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\dot{v}(X_i,\theta_0)\hat{f}_h^2(X_i)\right]^2 d\varphi(x).$$

Again, add and subtract $f^2(X_i)$ after $\hat{f}_h^2(X_i)$, similarly we can show that $\sup_{\|b\|=1}|\Sigma_n(b_1) - b^T\Sigma_0 b| = o_p(1)$. This, together with the fact that $D_{n2}\ge\inf_{\|b\|=1}\Sigma_n(b)$, implies that $nh^d\|\hat{\theta}_n - \theta_0\|^2 = O_P(1)$ by routine arguments.

Note that $\hat{\theta}_n$ satisfies the equation $\partial T_n(\theta)/\partial\theta|_{\theta=\hat{\theta}_n} = 0$. Adding and subtracting $v_0(X_i)$ inside $U_n(x,\theta)$, we obtain

$$\int_{\mathcal{C}} U_n(x)\dot{\mu}_n(x,\hat{\theta}_n)d\hat{\varphi}_w(x) = \int_{\mathcal{C}} Z_n(x,\hat{\theta}_n)\dot{\mu}_n(x,\hat{\theta}_n)d\hat{\varphi}_w(x) \quad (2.5.3)$$

We may discover that the left hand side (LHS) of (2.5.3) is equivalent to

$$\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[(Y_i-m(X_i)+m(X_i)-\hat{m}(X_i))^2-v_0(X_i)][\hat{f}_h^2(X_i)-f^2(X_i)+f^2(X_i)]\right]$$

$$\cdot\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)[\hat{f}_h^2(X_i)-f^2(X_i)+f^2(X_i)]\right]d\hat{\varphi}_w(x)$$

$$=\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)+(m(X_i)-\hat{m}(X_i))^2+2\varepsilon_i\sqrt{v_0(X_i)}(m(X_i)-\hat{m}(X_i))]\right.$$

$$\left.\cdot[\hat{f}_h^2(X_i)-f^2(X_i)+f^2(X_i)]\right]\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)[\hat{f}_h(X_i)-f(X_i)+f(X_i)]^2\right]d\hat{\varphi}_w(x)$$

$$:=\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)+\Delta_{im}^2+2\varepsilon_i\sqrt{v(X_i,\theta)}\Delta_{im}][\hat{f}_h^2(X_i)-f^2(X_i)+f^2(X_i)]\right]$$

$$\cdot\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}_0(X_i)[\Delta_{if}^2+f^2(X_i)+2\Delta_{if}f(X_i)]\right]d\hat{\varphi}_w(x)$$

$$=\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)f^2(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f^2(X_i)\right\}d\hat{\varphi}_w(x)$$

$$+2\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)f(X_i)\Delta_{if}]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f^2(X_i)\right\}d\hat{\varphi}_w(x)$$

$$+\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)\Delta_{if}^2]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f^2(X_i)\right\}d\hat{\varphi}_w(x)$$

$$+2\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\varepsilon_i\sqrt{v_0(X_i)}\Delta_{im}\hat{f}_h(X_i)f(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f^2(X_i)\right\}d\hat{\varphi}_w(x)$$

$$+2\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\varepsilon_i\sqrt{v_0(X_i)}\Delta_{im}\hat{f}_h(X_i)\Delta_{if}]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f^2(X_i)\right\}d\hat{\varphi}_w(x)$$

$$+\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\Delta_{im}^2\hat{f}_h^2(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f^2(X_i)\right\}d\hat{\varphi}_w(x)$$

$$+\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)f^2(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)\Delta_{if}^2\right\}d\hat{\varphi}_w(x)$$

$$+2\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)f(X_i)\Delta_{if}]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)\Delta_{if}^2\right\}d\hat{\varphi}_w(x)$$

$$+\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)\Delta_{if}^2]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)\Delta_{if}^2\right\}d\hat{\varphi}_w(x)$$
$$+2\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\varepsilon_i\sqrt{v_0(X_i)}\Delta_{im}\hat{f}_h(X_i)f(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)\Delta_{if}^2\right\}d\hat{\varphi}_w(x)$$
$$+2\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\varepsilon_i\sqrt{v_0(X_i)}\Delta_{im}\hat{f}_h(X_i)\Delta_{if}]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)\Delta_{if}^2\right\}d\hat{\varphi}_w(x)$$
$$+\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\Delta_{im}^2\hat{f}_h^2(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)\Delta_{if}^2\right\}d\hat{\varphi}_w(x)$$
$$+2\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)f^2(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}d\hat{\varphi}_w(x)$$
$$+4\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)f(X_i)\Delta_{if}]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}d\hat{\varphi}_w(x)$$
$$+2\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\xi(X_i)\Delta_{if}^2]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}d\hat{\varphi}_w(x)$$
$$+4\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\varepsilon_i\sqrt{v_0(X_i)}\Delta_{im}\hat{f}_h(X_i)f(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}d\hat{\varphi}_w(x)$$
$$+4\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\varepsilon_i\sqrt{v_0(X_i)}\Delta_{im}\hat{f}_h(X_i)\Delta_{if}]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}d\hat{\varphi}_w(x)$$
$$+2\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)[\Delta_{im}^2\hat{f}_h^2(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}d\hat{\varphi}_w(x).$$

For the sake of convenience, we will denote these 18 terms as $S_{n1},\cdots,S_{n18}$.

Next, we would, on each term S_{n1} to S_{n6}, add and subtract $\dot{\mu}_n(x,\theta)$, $\dot{\mu}_n(x,\theta_0)$ after $\dot{\mu}_h(x,\theta_0)$, add and subtract $f^{-6}(x)$ to $\hat{\varphi}_w(x)$. The main structure after such derivation are denoted with $\tilde{S}$. For the rest, we could show that after multiplying by $\sqrt{n}$, they converges to zero fast enough so that no component is left for variance. First note that,

$$n\tilde{S}_{n2}^2 = n\left(\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\xi(X_i)f(X_i)\Delta_{if}\right]\dot{\mu}_h(x,\theta_0)d\varphi(x)\right)^2$$
$$= n\int_{\mathcal{C}}\left[\frac{1}{n^2}\sum_{i=1}^{n}K_{hi}(x)K_{hi}(y)\xi^2(X_i)f^2(X_i)(\hat{f}_h(X_i)-f(X_i))^2\right]\dot{\mu}_h(x,\theta_0)\dot{\mu}_h(y,\theta_0)d\varphi(x)d\varphi(y)$$

$$+n\int_{\mathcal{C}}\left[\frac{1}{n^2}\sum_{i\neq j}^{n}K_{hi}(x)K_{hi}(y)\xi(X_i)\xi(X_j)f(X_i)f(X_j)(\hat{f}_h(X_i)-f(X_i))(\hat{f}_h(X_j)-f(X_j))\right]$$
$$\cdot\dot{\mu}_h(x,\theta_0)\dot{\mu}_h(y,\theta_0)d\varphi(x)d\varphi(y).$$

Note that by taking conditional expectation on X_i, and X_j respectively, the cross product term equals to zero under the Null hypothesis. Note that every element inside the integral with random variable $X_i's$ are positive, by Fubini's theorem, adding and subtracting $E[K_{hi1}|X_1]$, we see that $En\tilde{S}_{n2}^2$ equals

$$n\int_{\mathcal{C}}E\left[\frac{1}{n^2}\sum_{i=1}^{n}K_{hi}(x)K_{hi}(y)\xi^2(X_i)f^2(X_i)(\hat{f}_h(X_i)-f(X_i))^2\right]\dot{\mu}_h(x,\theta_0)\dot{\mu}_h(y,\theta_0)d\varphi(x)d\varphi(y)$$
$$=\int_{\mathcal{C}}E\Bigg[K_{h1}(x)K_{h1}(y)\xi_1^2f^2(X_1)\left(\frac{1}{n}\sum_{i=2}^{n}K_{hi1}-E(K_{hi1}|X_1)\right)^2$$
$$+2K_{h1}(x)K_{h1}(y)\xi^2(X_1)f^2(X_1)\left(\frac{1}{n}\sum_{i=2}^{n}K_{hi1}-E(K_{hi1}|X_1)\right)\left(E(K_{hi1}|X_1)-f(X_1)\right)$$
$$+K_{h1}(x)K_{h1}(y)\xi^2(X_1)f^2(X_1)\left(E(K_{hi1}|X_1)-f(X_1)\right)^2\Bigg]\dot{\mu}_h(x,\theta_0)\dot{\mu}_h(y,\theta_0)d\varphi(x)d\varphi(y)+O\Big(\frac{1}{nh^d}\Big)$$
$$=O\Big(\frac{1}{nh^{2d}}\Big)+O\Big(\frac{1}{n^2h^{3d}}\Big)+O(h^4)$$

where the last equality comes from $\dot{\mu}_h(x,\theta_0)$ and $\dot{\mu}_h(y,\theta_0)$ is finite on $\mathcal{C}$. For $S_{n3}^2, S_{n4}^2, S_{n5}^2$, we may observe the similar pattern where the expectation of cross product term equals zero when conditioned on X_i, X_j. So, for simplicity, we will show the form of $nS_{n3}^2, nS_{n4}^2, nS_{n5}^2$ where they all converges to zero.

Next, for nS_{n3}^2, add and subtract $E(K_h(X_j-X_i)|X_1)$ in $(\hat{f}_{hij})^4$ of the $En\tilde{S}_{n3}^2$ will gives us the following derivation.

$$n\tilde{S}_{n3}^2=n\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(x)\xi(X_i)(\hat{f}_h(X_i)-f(X_i))^2\right\}\left\{\frac{1}{n}\sum_{i=1}^{n}K_{hi}(y)\xi(X_i)(\hat{f}_h(X_i)-f(X_i))^2\right\}$$
$$\cdot\dot{\mu}_h(x,\theta)\dot{\mu}_h(y,\theta)d\varphi(x)d\varphi(y)$$

Therefore, we may calculate the expectation,

$$\begin{aligned}
En\tilde{S}_{n3}^2 &= \int_{\mathcal{C}} EK_{h1}(x)K_{h1}(y)\xi^2(X_1)(\hat{f}_h(X_i) - f(X_1))^4\dot{\mu}_h(x,\theta_0)\dot{\mu}_h(x,\theta_0)d\varphi(x)d\varphi(y) \\
&\le 8\int_{\mathcal{C}} EK_{h1}(x)K_{h1}(y)\xi^2(X_1)(\frac{1}{n}\sum_{j=1}^{n} K_{hj1} - E(K_{hj1}|X_1)^4\dot{\mu}_h(x,\theta_0)\dot{\mu}_h(x,\theta_0)d\varphi(x)d\varphi(y) \\
&\quad +8\int_{\mathcal{C}} EK_{h1}(x)K_{h1}(y)\xi^2(X_1)(E(K_{hj1}|X_1) - f(X_1))^4\dot{\mu}_h(x,\theta_0)\dot{\mu}_h(x,\theta_0)d\varphi(x)d\varphi(y).
\end{aligned}$$

Then similar to A_{n1}, we know that $nS_{n3}^2 = O_p((n^3h^{3d})^{-1}) + O_p((nh^{d-8})^{-1})$.

$$\begin{aligned}
En\tilde{S}_{n4}^2 = \int_{\mathcal{C}} E\Bigg[&K_{hi}(x)K_{hi}(y)\varepsilon_1^2 v_0(X_1)(\frac{1}{n}\sum_{j=1}^{n} K_{hj1}(Y_j - m(X_1)))^2 f^2(X_i)\Bigg] \\
&\cdot\dot{\mu}_h(x,\theta_0)\dot{\mu}_h(x,\theta_0)d\varphi(x)d\varphi(y)
\end{aligned}$$

And by technique similar to that of T_{n2}^*, we may show that the order is $O(h^4)+O(1/(n^2h^{2d-4}))$.

$$\begin{aligned}
En\tilde{S}_{n5}^2 &= \int_{\mathcal{C}} EK_{hi}(x)K_{hi}(y)\varepsilon_i^2 v_0(X_i)\Big(\frac{1}{n}\sum_{i=1}^{n} K_{hj1}(m(X_j) - m(X_1) + \varepsilon_j\sqrt{v_0(X_j)})\Big)^2 \\
&\qquad (\hat{f}_h(X_i) - f(X_i))^2\dot{\mu}_h(x,\theta_0)\dot{\mu}_h(x,\theta_0)d\varphi(x)d\varphi(y) \\
&= O(h^8) + O\Big(\frac{1}{nh^{d-4}}\Big) + O\Big(\frac{1}{n^2n^{2d}}\Big)
\end{aligned}$$

Also note that $En\tilde{S}_{n6}^2$ equals

$$\begin{aligned}
&n\int_{\mathcal{C}} \left\{\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)[\Delta_{im}^2\hat{f}_h^2(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n} K_{hi}(y)[\Delta_{im}^2\hat{f}_h^2(X_i)]\right\}\dot{\mu}(x,\theta_0)\dot{\mu}(y,\theta_0)d\varphi(x)d\varphi(y) \\
&= n\int_{\mathcal{C}} \left\{\frac{1}{n}\sum_{i=1}^{n} K_{hi}(y)\left[\frac{1}{n}\sum_{j=1}^{n} K_{hij}(Y_j - m(X_i))\right]^2\right\} \\
&\qquad \left\{\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\left[\frac{1}{n}\sum_{j=1}^{n} K_{hij}(Y_j - m(X_i))\right]^2\right\}\dot{\mu}(x,\theta_0)\dot{\mu}(y,\theta_0)d\varphi(x)d\varphi(y) \\
&= n\int_{\mathcal{C}} \left\{\frac{1}{n^2}\sum_{i=1}^{n} K_{hi}(x)K_{hi}(y)\left[\frac{1}{n}\sum_{j=1}^{n} K_{hij}(Y_j - m(X_i))\right]^4\right\}\dot{\mu}(x,\theta_0)\dot{\mu}(y,\theta_0)d\varphi(x)d\varphi(y)
\end{aligned}$$

$$+n\int_{\mathcal{C}}\left\{\frac{1}{n^2}\sum_{i\neq k}^{n}K_{hi}(x)K_{hj}(y)\left[\frac{1}{n}\sum_{j=1}^{n}K_{hij}(Y_j-m(X_i))\right]^2\right.$$

$$\left.\cdot\left[\frac{1}{n}\sum_{k=1}^{n}K_{hki}(Y_k-m(X_i))\right]^2\right\}\dot{\mu}(x,\theta_0)\dot{\mu}(y,\theta_0)d\varphi(x)d\varphi(y):=nS_{n61}+nS_{n62}$$

First we have

$$\begin{aligned}EnS_{n61} &= \int_{\mathcal{C}}E\left\{K_{h1}(x)K_{h1}(y)\left[\frac{1}{n}\sum_{j=1}^{n}K_{hj1}(Y_j-m(X_1))\right]^4\right\}\dot{\mu}(x,\theta_0)\dot{\mu}(y,\theta_0)d\varphi(x)d\varphi(y)\\ &\leq 8\int_{\mathcal{C}}E\left\{K_{h1}^2(x)\left[\frac{1}{n}\sum_{j=1}^{n}K_{hj1}(m(X_j)-m(X_1))\right]^4\right\}\dot{\mu}^2(x,\theta_0)d\varphi(x)\\ &\quad+8\int_{\mathcal{C}}E\left\{K_{h1}^2(x)\left[\frac{1}{n}\sum_{j=1}^{n}K_{hj1}(\varepsilon_j\sqrt{v_0(X_j)})\right]^4\right\}\dot{\mu}^2(x,\theta_0)d\varphi(x)\\ &:= O\Big(\frac{1}{n^3h^{3d}}\Big)+O(h^8).\end{aligned}$$

With similar technique, we also have

$$\begin{aligned}EnS_{n62} = n\int_{\mathcal{C}}E\Bigg\{&\frac{1}{n^2}\sum_{i\neq k}^{n}K_{hi}(x)K_{hj}(y)\left[\frac{1}{n}\sum_{j=1}^{n}K_{hij}(Y_j-m(X_i))\right]^2\\ &\cdot\left[\frac{1}{n}\sum_{k=1}^{n}K_{hki}(Y_k-m(X_i))\right]^2\Bigg\}\dot{\mu}(x,\theta_0)\dot{\mu}(y,\theta_0)d\varphi(x)d\varphi(y)\\ = O(nh^8)&+O\left(\frac{1}{nh^{2d-4}}\right)+O\left(\frac{1}{n^3h^{4d}}\right).\end{aligned}$$

Next, we focus on $n\tilde{S}_{n8}^2$ to $n\tilde{S}_{n12}^2$, and $n\tilde{S}_{n14}^2$ to $n\tilde{S}_{n18}^2$ first. Since we have shown that $n\tilde{S}_{n2}^2$ to $n\tilde{S}_{n6}^2$ are $o_p(1)$, it would be easy to show that all those terms are bounded above by $n\tilde{S}_{n2}^2$ to $n\tilde{S}_{n6}^2$, respectively, and then it suffices to show that $n\tilde{S}_{n7}^2$ and $n\tilde{S}_{n13}^2$ are $o_p(1)$. $n\tilde{S}_{n1}^2$ will be of order $O_p(1)$.

$$
\begin{aligned}
\sqrt{n}\tilde{S}_{n8} &= \int_{\mathcal{C}} \sqrt{n}\left\{\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)[\xi(X_i)f(X_i)\Delta_{if}]\right\}\left\{\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\dot{v}(X_i,\theta)\Delta_{if}^2\right\}d\hat{\varphi}_w(x)\\
&= \int_{\mathcal{C}} \sqrt{n}\left\{\frac{1}{n^2}\sum_{i=1}^{n} K_{hi}^2(x)\xi(X_i)f(X_i)\Delta_{if}^3\dot{v}(X_i,\theta_0)\right.\\
&\quad \left. +\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hi}(x)K_{hj}(x)\xi(X_i)f(X_i)\dot{v}(X_j,\theta_0)\Delta_{if}^2\Delta_{jf}\right\}d\hat{\varphi}_w(x) := S_{n81}+S_{n82}.
\end{aligned}
$$

Note that $ES_{n81} = ES_{n82} = 0$, so it suffices to show $E(S_{n81}^2) = o(1)$ and $E(S_{n82}^2) = o(1)$. By Fubini's Theorem, and the fact that $E(\xi(X_1)|X_1) = 0$, $E(S_{n81}^2)$ equals

$$
\begin{aligned}
&\frac{1}{n^3}\int_{\mathcal{C}} E\Big(\sum_{i=1}^{n} K_{hi}^2(x)\xi(X_i)f(X_i)\dot{v}(X_i,\theta)\Big(\frac{1}{n}\sum_{j=1}^{n} K_{hij} - f(X_i)\Big)^3\Big)^2 d\hat{\varphi}_w(x)\\
&= \frac{1}{n^3}\int_{\mathcal{C}} nE\Big(K_{h1}^4(x)\xi^2(X_i)f^2(X_1)\dot{v}^2(X_i,\theta)\Big(\frac{1}{n}\sum_{j=1}^{n} K_{hij} - f(X_i)\Big)^6\Big)d\hat{\varphi}_w(x)\\
&= \frac{1}{n^2}\int_{\mathcal{C}} EK_{h1}^4(x)\xi^2(X_i)f^2(X_1)\dot{v}^2(X_i,\theta)\Big(\frac{1}{n}\sum_{j=1}^{n} K_{hij} - E(K_{h1j}|X_1)\\
&\qquad\qquad +E(K_{h1j}|X_1) - f(X_i)\Big)^6 d\hat{\varphi}_w(x)\\
&\le \frac{32}{n^2}\int_{\mathcal{C}} EK_{h1}^4(x)\xi^2(X_i)f^2(X_1)\dot{v}^2(X_i,\theta)\Big(\frac{1}{n}\sum_{j=1}^{n} K_{hij} - E(K_{h1j}|X_1)\Big)^6 d\hat{\varphi}_w(x)\\
&\qquad\qquad +\frac{32}{n^2}\int_{\mathcal{C}} EK_{h1}^4(x)\xi^2(X_i)f^2(X_1)\dot{v}^2(X_i,\theta)(E(K_{h1j}|X_1) - f(X_i))^6 d\hat{\varphi}_w(x)\\
&= \frac{32}{n^8}\int_{\mathcal{C}} EK_{h1}^4(x)\xi^2(X_i)f^2(X_1)\dot{v}^2(X_i,\theta)\Big(\sum_{j=1}^{n}(K_{hij} - E(K_{h1j}|X_1))^6\\
&\quad + \sum_{j\neq k\neq l}^{n}(K_{h1j} - E(K_{h1j}|X_1))^2(K_{h1k} - E(K_{h1k}|X_1))^2(K_{h1l} - E(K_{h1l}|X_1))^2\\
&\quad + \sum_{j\neq k}^{n}(K_{h1j} - E(K_{h1j}|X_1))^4(K_{h1k} - E(K_{h1k}|X_1))^2\Big)d\hat{\varphi}_w(x) + O(h^{12}/n^2)\\
&= O\Big(\frac{1}{n^7h^{7d}}\Big) + O\Big(\frac{1}{n^6h^{6d}}\Big) + O\Big(\frac{1}{n^5h^{5d}}\Big) + O\Big(\frac{h^{12}}{n^2}\Big).
\end{aligned}
$$

For S_{n82}, we have

$$\begin{aligned}
E(S_{n82}^2) &= \frac{1}{n^3}\int_{\mathcal{C}} E\Big(\sum_{i=1}^n\sum_{j\neq i}^n K_{hi}(x)K_{hj}(x)\xi(X_i)f(X_i)\dot{v}(X_i,\theta)\Delta_{if}\Delta_{jf}^2\Big)^2 d\hat{\varphi}_w(x)\\
&= \frac{1}{n^3}\int_{\mathcal{C}} E\Big(\sum_{i=1}^n\sum_{j\neq i}^n K_{hi}^2(x)K_{hj}^2(x)\xi^2(X_i)f^2(X_i)\dot{v}^2(X_i,\theta)\Delta_{if}^2\Delta_{jf}^4\\
&\quad+\sum_{i=1}^n\sum_{j\neq i}^n\sum_{k\neq j\neq i}^n K_{hi}^2(x)K_{hj}(x)K_{hk}(x)\xi^2(X_i)f^2(X_i)\dot{v}(X_j,\theta)\dot{v}(X_k,\theta)\Delta_{if}^2\Delta_{jf}^2\Delta_{kf}^2\Big)d\hat{\varphi}_w(x)\\
&= O\Big(\frac{1}{n^4h^{4d}}\Big)+O\Big(\frac{1}{n^3h^{3d}}\Big)+O\Big(\frac{h^{12}}{n}\Big)+O(h^{12}).
\end{aligned}$$

Thus we have,

$$\left(\sqrt{n}\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\xi(X_i)f(X_i)\Delta_{if}\right\}\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\theta)\Delta_{if}f(X_i)\right\}d\hat{\varphi}_w(x)\right)^2 = o_p(1).$$

Next we will show $n\tilde{S}_{n13}^2$ converges to zero, and $n\tilde{S}_{n7}^2 = o_p(1)$ will follow similarly. Note that

$$\begin{aligned}
n\tilde{S}_{n13}^2 &= n\left(\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(x)[\xi(X_i)f^2(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}d\varphi(x)\right)^2\\
&= n\int_{\mathcal{C}}\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(x)[\xi(X_i)f^2(X_i)]\right\}\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(y)[\xi(X_i)f^2(X_i)]\right\}\\
&\quad\cdot\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(y)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}d\varphi(x)d\varphi(y)\\
&= n\int_{\mathcal{C}}\left\{\frac{1}{n^2}\sum_{i=1}^n K_{hi}(x)K_{hi}(y)\xi^2(X_i)f^4(X_i)]+\frac{1}{n^2}\sum_{i\neq j}^n K_{hi}(x)K_{hj}(y)\xi(X_i)\xi(X_j)f^2(X_i)f^2(X_j)\right\}\\
&\quad\cdot\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(y)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}d\varphi(x)d\varphi(y).\\
En\tilde{S}_{n13}^2 &= \int_{\mathcal{C}} EK_{h1}(x)K_{h1}(y)\xi^2(X_1)f^4(X_1)\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}\\
&\qquad\cdot\left\{\frac{1}{n}\sum_{i=1}^n K_{hi}(y)\dot{v}(X_i,\theta)f(X_i)\Delta_{if}\right\}d\varphi(x)d\varphi(y) = o(1).
\end{aligned}$$

In the following we show that $\sqrt{n}S_{n1}$ is asymptotically normal. For the sake of brevity, we only prove the result for $d = 1$. The multidimensional version can be easily derived by using Wold technique. Rewrite S_{n1},

$$
\begin{aligned}
S_{n1} &= \int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)\xi(X_i)f^2(X_i) \right\} \{\dot{\mu}_n(x,\theta) - \dot{\mu}_n(x,\theta_0) + \dot{\mu}_n(x,\theta_0) - \dot{\mu}_h(x,\theta_0) + \dot{\mu}_h(x,\theta_0)\} \\
&\qquad \{\hat{f}_w^{-6}(x) - f^{-6}(x) + f^{-6}(x)\} dG(x) \\
&= \int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)\xi(X_i)f^2(X_i) \right\} \dot{\mu}_h(x,\theta_0) f^{-6}(x) dG(x) \\
&\quad + \int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)\xi(X_i)f^2(X_i) \right\} \{\dot{\mu}_n(x,\theta) - \dot{\mu}_n(x,\theta_0)\} f^{-6}(x) dG(x) \\
&\quad + \int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)\xi(X_i)f^2(X_i) \right\} \{\dot{\mu}_n(x,\theta_0) - \dot{\mu}_h(x,\theta_0)\} f^{-6}(x) dG(x) \\
&\quad + \int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)\xi(X_i)f^2(X_i) \right\} \dot{\mu}_h(x,\theta_0) \{\hat{f}_w^{-6}(x) - f^{-6}(x)\} dG(x) \\
&\quad + \int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)\xi(X_i)f^2(X_i) \right\} \{\dot{\mu}_n(x,\theta) - \dot{\mu}_n(x,\theta_0)\} \{\hat{f}_w^{-6}(x) - f^{-6}(x)\} dG(x) \\
&\quad + \int_{\mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)\xi(X_i)f^2(X_i) \right\} \{\dot{\mu}_n(x,\theta_0) - \dot{\mu}_h(x,\theta_0)\} \{\hat{f}_w^{-6}(x) - f^{-6}(x)\} dG(x) \\
&:= S_{n1,1} + S_{n1,2} + S_{n1,3} + S_{n1,4} + S_{n1,5} + S_{n1,6}.
\end{aligned}
$$

To proceed, we need the following lemmas:

Lemma 2.5.1. Under H_0, suppose conditions (e1), (e2), (f1), (k), (v1) to (v5) hold:

(a). If additionally, (g) holds, $\sqrt{n}S_{n1,1} \to N(0, \Sigma)$.

(b). If (f2) and (w) hold, the $\sqrt{n}|S_{n1,4}| = o_p(1)$.

Lemma 2.5.2. Under H_0, suppose all the conditions in Lemma 2.5.1, Furthermore, assume (f2) and (w) hold, then $\sqrt{n}S_{n1,k} = o_p(1)$ holds for $k = 2, 3, 5, 6$.

We also need the following lemma from Bosq (2012).

Lemma 2.5.3. Let $\hat{f}_w$ be the kernel estimate associate with a kernel K which satisfies a

Lipschitz condition. If (f2) holds and $w_n = a_n(\log n/n)^{1/(d+4)}$, where $a_n \to a_0 > 0$, then

$$(\log_k n)^{-1}(n/\log n)^{2/(d+4)} \sup_{x\in\mathcal{C}} |\hat{f}_w(x) - f(x)| \to 0,\ a.s.,\ \forall k > 0, k \in N$$

To prove part (a) of Lemma 2.5.1, let $s_{n11i} := \int_{\mathcal{C}} K_{hi}(x)\xi(X_i)f^2(X_i)\dot{\mu}_h(x,\theta_0)d\varphi(x)$, we have

$$\sqrt{n}S_{n1,1} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s_{n11i}$$

Note that $\{s_{n11i}, 1 \le i \le n\}$ are i.i.d. centered random variables for each n. By Central Limit Theorem with Lindeberg Feller condition, it suffices to show that as $n \to \infty$,

$$Es^2_{n111} \to \Sigma \tag{2.5.4}$$

$$E\{s^2_{n111}I(|s_{n111}| \ge n^{1/2}\eta)\} \to 0,\ \forall \eta \ge 0 \tag{2.5.5}$$

By the definition of $\dot{\mu}_h(x,\theta)$, Fubini's theorem, change of variables by $\frac{u-x}{h} = u_1$, and $\frac{u-y}{h} = u_2$ and the continuity of functions $K, v, f, \dot{v}$, together with bounded convergence theorem, we have $\lim_{n\to\infty} \dot{\mu}_h(x,\theta) = \dot{v}_0(x)f^3(x)$, thus

$$\begin{aligned}
Es^2_{n111} &= \int\int\int K_h(u-x)K_h(u-y)E(\xi^2|X=u)f^4(u)f(u)du\dot{\mu}_h(x,\theta_0)\dot{\mu}_h(y,\theta_0) \\
&\qquad\qquad \cdot f^{-6}(x)f^{-6}(y)g(x)g(y)dxdy \\
Es^2_{n111} &= \int\int\int K(u_1)K(u_2)\dot{\mu}_h(u+u_1h,\theta_0)\dot{\mu}_h(u+u_2h,\theta_0)f^{-6}(u+u_1h)f^{-6}(u+u_2h) \\
&\qquad \cdot g(u+u_1h)g(u+u_2h)du_1du_2f^5(u)E(\xi^2|X=u)du \\
&\to \int \dot{v}^2(u,\theta_0)f^6(u)f^{-6}(u)f^{-6}(u)f^5(u)g^2(u)E(\xi^2|X=u)du \\
&= \int \frac{\dot{v}^2(u,\theta_0)E(\xi^2|X=u)g^2(u)}{f(u)}du = \Sigma
\end{aligned}$$

To prove the next claim, first note that by Holder's inequality,

$$E\{s^2_{n11}I(|s_{n11}| \ge n^{1/2}\eta)\} = E(s^{2+\delta}_{n11})\frac{1}{s^{\delta}_{n11}}I(|s_{n11}| \ge n^{1/2}\eta) \le E\Big(s^{2+\delta}_{n11}\frac{1}{n^{\delta/2}\eta^{\delta}_d}\Big)$$

$$= n^{-\frac{\delta}{2}}\eta_d^{-\delta}E\Bigg[\Bigg(\int(K_{h1}(x)f^2(X_1)\dot{\mu}_h(x))^{(2+\delta)/2}d\varphi(x)\Bigg)^2|\xi|^{2+\delta}\Bigg] = O(nh^d)^{-\delta/2} = o(1).$$

To prove part (b) of the Lemma 2.5.1, we need to use Lemma 2.5.3, and the Cauchy-Schwartz inequality, the boundedness of $\dot{\mu}_h(x)$, we have,

$$\begin{aligned}
nS_{n1,1}^2 &\le Mn\int(\frac{1}{n}\sum_{i=1}^n K_{hi}(x)f^2(X_i)\xi(X_i)\dot{\mu}_h(x))^2d\varphi(x)\sup_{x\in\mathcal{C}}|f^6(x)/\hat{f}_w^6(x)-1|^2\\
&= nO_p((nh^d)^{-1})O_p((\log_k n)^2(\log n/n)^{4/(4+d)})\\
&= O_p((\log_k n)^2(\log n)^{4/(4+d)}n^{ad-4/(4+d)}) = o_p(1).
\end{aligned}$$

And this concludes the proof of Lemma 2.5.1. Next, to show Lemma 2.5.2 we have, by Cauchy-Schwartz inequality,

$$||\sqrt{n}S_{n1,3}||^2 \le \Bigg(\sqrt{n}\int\Big(\frac{1}{n}\sum_{i=1}^n K_{hi}(x)f^2(X_i)\xi(X_i)\Big)^2d\varphi(x)\Bigg)\Bigg(\sqrt{n}\int||\dot{\mu}_n(x,\theta_0)-\dot{\mu}_h(x,\theta_0)||^2d\varphi(x)\Bigg)$$

and it is again, easy to show that

$$E\sqrt{n}\int\Big(\frac{1}{n}\sum_{i=1}^n K_{hi}(x)f^2(X_i)\xi(X_i)\Big)^2d\varphi(x) = O\Big(\frac{1}{n^{1/2}h^d}\Big) = o(1)$$

As for the second term, since $\dot{\mu}_h(x,\theta_0)$ is the mean of $\dot{\mu}_n(x,\theta_0)$, then the expectation of the second term would be bounded above by the second order of $\dot{\mu}_n(x,\theta_0)$, that is

$$\sqrt{n}\int E||K_{h1}(x)\dot{v}_0(x)f^2(X_1)||^2d\varphi(x) = O\Big(\frac{1}{n^{1/2}h^d}\Big) = o(1)$$

Similarly we can ge the proof of $\sqrt{n}S_{n1,6}$.

$$||\sqrt{n}S_{n1,2}||^2 \le n\int\Big(\frac{1}{n}\sum_{i=1}^n K_{hi}(x)f^2(X_i)\xi(X_i)\Big)^2d\varphi(x)\int||\dot{\mu}_n(x,\hat{\theta}_n)-\dot{\mu}_n(x,\theta_0)||^2d\varphi(x)$$

by assumption (v5), we know the second part is bounded above by

$$\max_{1\le i\le n} ||\dot{v}(X_i,\theta_0)-\dot{v}(\theta_n,X_i)||^2\int(\hat{f}_h(x)f^2(x))^2d\varphi(x)=o_p(h^d)O_p(1)$$

Together, we can show that $\sqrt{n}S_{n1,2}=o_p(1)$, and $\sqrt{n}S_{n1,5}=o_p(1)$ follows similarly. The proof of $\sqrt{n}S_{n1,6}$ follows $\sqrt{n}S_{n1,6}$ and $\sqrt{n}S_{n1,2}$, which complete the proof of both lemmas, and therefore the left hand side of equation (2.5.3).

Similarly, for the RHS of equation (2.5.3), plus and minus the true density at sample point X_i, we have the following,

$$\begin{aligned}
&\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)(\dot{v}_0(X_i)-\dot{v}(X_i,\hat{\theta}_n))\hat{f}_h^2(X_i)\right]\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)\hat{f}_h^2(X_i)\right]d\hat{\varphi}_w(x)\\
=&\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)(\dot{v}_0(X_i)-\dot{v}(X_i,\hat{\theta}_n))f^2(X_i)\right]\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)f^2(X_i)\right]d\hat{\varphi}_w(x)\\
&+2\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)(\dot{v}_0(X_i)-\dot{v}(X_i,\hat{\theta}_n))f^2(X_i)\right]\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)\Delta_{if}f(X_i)\right]d\hat{\varphi}_w(x)\\
&+\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)(\dot{v}_0(X_i)-\dot{v}(X_i,\hat{\theta}_n))f^2(X_i)\right]\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)\Delta_{if}^2\right]d\hat{\varphi}_w(x)\\
&+2\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)(\dot{v}_0(X_i)-\dot{v}(X_i,\hat{\theta}_n))\Delta_{if}f(X_i)\right]\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)f^2(X_i)\right]d\hat{\varphi}_w(x)\\
&+4\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)(\dot{v}_0(X_i)-\dot{v}(X_i,\hat{\theta}_n))\Delta_{if}f(X_i)\right]\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)\Delta_{if}f(X_i)\right]d\hat{\varphi}_w(x)\\
&+2\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)(\dot{v}_0(X_i)-\dot{v}(X_i,\hat{\theta}_n))\Delta_{if}f(X_i)\right]\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)\Delta_{if}^2\right]d\hat{\varphi}_w(x)\\
&+\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)(\dot{v}_0(X_i)-\dot{v}(X_i,\hat{\theta}_n))\Delta_{if}^2\right]\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)\Delta_{if}^2\right]d\hat{\varphi}_w(x)\\
&+2\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)(\dot{v}_0(X_i)-\dot{v}(X_i,\hat{\theta}_n))\Delta_{if}^2\right]\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)\Delta_{if}f(X_i)\right]d\hat{\varphi}_w(x)\\
&+\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)(\dot{v}_0(X_i)-\dot{v}(X_i,\hat{\theta}_n))\Delta_{if}^2\right]\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)\Delta_{if}^2\right]d\hat{\varphi}_w(x).
\end{aligned}$$

Note that with the previous derivation, we may conclude that except the first one, the remaining eight terms are $o_p(1)$. So the first term may have the following expansion,

$$
\begin{aligned}
&\int_{\mathcal{C}} \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)(\dot{v}_0(X_i) - \dot{v}(X_i,\hat{\theta}_n))f(X_i)^2\right]\left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\dot{v}(X_i,\hat{\theta}_n)f(X_i)^2\right] d\hat{\varphi}_w(x) \\
&= \int_{\mathcal{C}} \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)(\dot{v}_0(X_i) - \dot{v}(X_i,\hat{\theta}_n) - u_n^T\dot{v}_0(X_i) + u_n^T\dot{v}_0(X_i))f^2(X_i)\right]\dot{\mu}_n(x,\hat{\theta}_n)d\hat{\varphi}_w(x) \\
&= \int_{\mathcal{C}} \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\Big(\frac{\dot{v}_0(X_i) - \dot{v}(X_i,\hat{\theta}_n) - u_n^T\dot{v}_0(X_i)}{||\hat{\theta}_n - \theta_0||}\Big)f^2(X_i)\right]\dot{\mu}_n(x,\hat{\theta}_n)d\hat{\varphi}_w(x)||u_n|| \\
&\qquad\qquad + \int_{\mathcal{C}} \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)(u_n^T\dot{v}_0(X_i))f^2(X_i)\right]\dot{\mu}_n(x,\hat{\theta}_n)d\hat{\varphi}_w(x) \\
&= \int_{\mathcal{C}} \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\frac{d_{ni}}{||u_n||}f^2(X_i)\right]\dot{\mu}_n(x,\hat{\theta}_n)d\hat{\varphi}_w(x)||u_n|| + \int_{\mathcal{C}} \dot{\mu}_n(x,\theta_0)\dot{\mu}_n^T(x,\hat{\theta}_n)d\hat{\varphi}_w(x)u_n \\
&:= V_n||u_n|| + L_n u_n.
\end{aligned}
$$

By assumptions (v4) and (v5),

$$
\begin{aligned}
||V_n|| &\le \max_{1\le i\le n}\frac{d_{ni}}{||u_n||}\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)f^2(X_i)\right]||\dot{\mu}_n(x,\hat{\theta}_n) - \dot{\mu}_h(x,\theta_0) + \dot{\mu}_h(x,\theta_0)||d\hat{\varphi}_w(x) \\
&\le \max_{1\le i\le n}\frac{d_{ni}}{||u_n||}||\dot{v}(X_i,\hat{\theta}_n) - \dot{v}_0(X_i)||\int_{\mathcal{C}}\hat{f}_h(x)f^2(x)d\hat{\varphi}_w(x) + \int_{\mathcal{C}}\hat{f}_h(x)f^2(x)||\dot{\mu}_h(x,\theta_0)||d\hat{\varphi}_w(x) \\
&= o_p(1)(o_p(1) + O_p(1)) = o_p(1),
\end{aligned}
$$

this concludes the proof of $V_n = o_p(1)$.

$$
\begin{aligned}
L_n &= \int_{\mathcal{C}} \dot{\mu}_n(x,\theta_0)(\dot{\mu}_n(x,\hat{\theta}_n) - \dot{\mu}_h(x,\theta_0) + \dot{\mu}_h(x,\theta_0))^T d\hat{\varphi}_w(x) \\
&= \int_{\mathcal{C}} \dot{\mu}_n(x,\theta_0)(\dot{\mu}_n(x,\hat{\theta}_n) - \dot{\mu}_h(x,\theta_0))^T d\hat{\varphi}_w(x) + \int_{\mathcal{C}} \dot{\mu}_n(x,\theta_0)\dot{\mu}_h(x,\theta_0)^T d\hat{\varphi}_w(x) \\
&:= L_{n1} + L_{n2}.
\end{aligned}
$$

And it is easy to see that $||L_{n1}|| = o_p(1)$ by assumption (v5). Similarly we have,

$$
||L_{n2} - \int_{\mathcal{C}} \dot{\mu}_h^T(x,\theta_0)\dot{\mu}_h(x,\theta_0)d\hat{\varphi}_w(x)|| \le \int_{\mathcal{C}} ||\dot{\mu}_n(x,\theta_0) - \dot{\mu}_h(x,\theta_0)||^2 d\hat{\varphi}_w(x)
$$

$$+2\int_{\mathcal{C}} ||\dot{\mu}_n(x,\theta_0) - \dot{\mu}_h(x,\theta_0)|| \cdot ||\dot{\mu}_h(x,\theta_0)|| d\hat{\varphi}_w(x) = o_p(1).$$

Note that

$$\int_{\mathcal{C}} \dot{\mu}_h^T(x,\theta_0)\dot{\mu}_h(x,\theta_0)d\hat{\varphi}_w(x) = \int_{\mathcal{C}} \dot{v}(\theta_0)f^3(x)\dot{v}^T(\theta_0)f^3(x)d\hat{\varphi}_w(x) + o_p(1).$$

and $\int_{\mathcal{C}} \dot{v}(\theta_0)f^3(x)\dot{v}^T(\theta_0)f^3(x)d\hat{\varphi}_w(x)$ equals

$$\int_{\mathcal{C}} \dot{v}(\theta_0)f^3(x)\dot{v}^T(\theta_0)f^3(x)\Big\{\frac{1}{\hat{f}_w^6(x)} - \frac{1}{f^6(x)}\Big\}dG(x) + \int_{\mathcal{C}} \dot{v}(\theta_0)\dot{v}^T(\theta_0)dG(x) = o_p(1) + \Sigma_0,$$

this concludes the proof. □

To prove Theorem 2.3.4, we need the Theorem 1 from Hall (1984) which is reproduced here for the sake of completeness.

Lemma 2.5.4. Let $\tilde{X}_i$, $1 \le i \le n$, be i.i.d. random vectors, and define

$$U_n := \sum_{1\le i<j\le n} H_n(\tilde{X}_i, \tilde{X}_j), \quad G_n(x,y) = EH_n(\tilde{X}_1, x)H_n(\tilde{X}_1, y),$$

where H_n is a sequence of measurable functions symmetric under permutation, with

$$E(H_n(\tilde{X}_1, \tilde{X}_2)|\tilde{X}_1) = 0 \ a.s., \quad EH_n^2(\tilde{X}_1, \tilde{X}_2) < \infty$$

for each $n \ge 1$. If

$$\frac{[EG_n^2(\tilde{X}_1, \tilde{X}_2) + n^{-1}EH_n^4(\tilde{X}_1, \tilde{X}_2)]}{[EH_n^2(\tilde{X}_1, \tilde{X}_2)]^2} \to 0,$$

then U_n is asymptotically normal with mean 0 and variance $\lim_{n\to\infty} n^2 EH_n^2(\tilde{X}_1, \tilde{X}_2)/2$.

The proof of Theorem 2.3.4 is facilitated by the following lemmas.

Lemma 2.5.5. Under H_0, if (e1), (e2), (e4), (f1), (g), (h1), (h2), and (k) hold, then

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

in distribution.

Proof. First, note that by the similar derivation as in the proof of asymptotic normality of $\hat{\theta}_n$, we have

$$nh^{d/2}T_n(\theta_0) = nh^{d/2} \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)[(Y_i - m(X_i) + m(X_i) - \hat{m}(X_i))^2 - v_0(X_i)] \right. \\ \left. \cdot (\hat{f}_h(X_i) - f(X_i) + f(X_i))^2 \right]^2 \left(\frac{1}{\hat{f}_w^6(x)} - \frac{1}{f^6(x)} + \frac{1}{f^6(x)} \right) dG(x)$$

From the previous results, under similar assumptions, together with the dimension restriction $d < 4$, it would be easy to see that the dominating term on the right hand side is $nh^{d/2} \int_{\mathcal{C}} [n^{-1} \sum_{i=1}^{n} K_{hi}(x)[(Y_i - m(X_i))^2 - v_0(X_i)] f^2(X_i)]^2 d\varphi(x)$, denoted as $\tilde{T}_n(\theta_0)$.

It can be seen that

$$\begin{aligned} & nh^{d/2} \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)[(Y_i - m(X_i))^2 - v_0(X_i)] f^2(X_i) \right]^2 d\varphi(x) \\ = & \ nh^{d/2} \int_{\mathcal{C}} \frac{1}{n^2} \sum_{i=1}^{n} K_{hi}^2(x) \xi^2(X_i) f^4(X_i) d\varphi(x) \\ & + \int_{\mathcal{C}} \frac{1}{n^2} \sum_{i \neq j}^{n} K_{hi}(x) K_{hj}(x) \xi(X_i) \xi(X_j) f^2(X_i) f^2(X_j) d\varphi(x) \\ := & \ nh^{d/2}(\tilde{C}_n + M_n). \end{aligned}$$

We define the following quantities,

$$\begin{aligned} \Gamma_n &:= 2h^d \frac{n-1}{n} \int \int [EK_{h1}(x) K_{h1}(y) \xi^2(X_i) f^4(X_1)]^2 d\varphi(x) d\varphi(y) \\ H_n(\tilde{X}_i, \tilde{X}_j) &:= \frac{h^{d/2}}{n} \int K_{hi}(x) K_{hj}(x) \xi(X_i) \xi(X_j) f^2(X_i) f^2(X_j) d\varphi(x), \end{aligned}$$

where $\tilde{X}_i = (X_i^T, \xi(X_i))^T$. So $nh^{d/2}M_n/2 = \sum_{i<j} H_n(\tilde{X}_i, \tilde{X}_j)$, and we will have,

$$EH_n^2(\tilde{X}_1,\tilde{X}_2) = \frac{h^d}{n^2}\iint EK_{h1,2}(x)K_{h1,2}(y)\xi^2(X_1)\xi^2(X_2)f^4(X_1)f^4(X_2)d\varphi(x)d\varphi(y)$$
$$= \frac{h^d}{n^2}\iint [EK_{h1}(x)K_{h1}(y)\xi^2(X_1)f^4(X_1)]^2 d\varphi(x)d\varphi(y)$$
$$= \frac{h^d}{n^2}\iint \left(\int \frac{1}{h^{2d}}K\left(\frac{u-x}{h}\right)K\left(\frac{u-y}{h}\right)\xi^2(u)f^4(u)f(u)du\right)^2 d\varphi(x)d\varphi(y)$$
$$= \frac{1}{n^2h^d}\iint \left(K(u)K\left(\frac{x-y}{h}+u\right)\xi^2(x+uh)f^5(x+uh)du\right)^2 d\varphi(x)d\varphi(y)$$

Next, we will show

$$\frac{EG^2(\tilde{X}_1,\tilde{X}_2)}{[EH_n^2(\tilde{X}_1,\tilde{X}_2)]^2} = o(1), \quad \frac{EH^4(\tilde{X}_1,\tilde{X}_2)}{n[EH_n^2(\tilde{X}_1,\tilde{X}_2)]^2} = o(1),$$

which implies the desired result by Lemma 2.5.4. In fact, the above results can be obtained if we can show

$$EG^2(\tilde{X}_1,\tilde{X}_2) = O\Big(\frac{h^d}{n^4}\Big), \quad EH_n^2(\tilde{X}_1,\tilde{X}_2) = O\Big(\frac{1}{n^2}\Big), \quad EH^4(\tilde{X}_1,\tilde{X}_2) = O\Big(\frac{1}{n^4h^{2d}}\Big).$$

Note that $G_n(x,y) = EH_n(\tilde{X}_1,x)H_n(\tilde{X}_1,y)$, left $t^T = (t_1^T, t_2)$, and $s^T = (s_1^T, s_2)$, where t_1 and s_1 are $n \times 1$ vector, where t_2 and s_2 are scalar. So, we can get

$$G_n(t,s) = EH_n(\tilde{X}_1,t)H_n(\tilde{X}_1,s)$$
$$= h^d n^{-2} E\int K_h(\tilde{X}_1 - x)K_h(t_1-x)\xi(X_1)t_2f^2(X_1)f^2(t_1)K_h(\tilde{X}_1-z)K_h(s_1-z)\xi(X_1)s_2$$
$$\cdot f^2(X_1)f^2(s_1)d\varphi(x)d\varphi(y)$$
$$= h^d n^{-2}\int K_h(t_1-x)t_2s_2f^2(t_1)K_h(s_1-z)f^2(s_1)$$
$$\cdot E(K_h(\tilde{X}_1-x)K_h(\tilde{X}_1-z)\xi^2(X_1)f^4(X_1))d\varphi(x)d\varphi(y).$$
$$EK_{h1}(x)K_{h1}(z)\xi^2(X_1,\hat{\theta}_n)f^4(X_1) = \int K_h(u-x)K_h(u-z)\xi^2(u)f^5(u)du$$
$$= \frac{1}{h^d}\int K(u)K\Big(\frac{x-z}{h}+u\Big)\xi^2(x+uh)f^5(x+uh)du := B_h(x-z)$$

$$EG_n^2(\tilde{X}_1, \tilde{X}_2) = \frac{h^{2d}}{n^4} \int B_h(x-z)B_h(x-w)B_h(z-v)B(v-w)d\varphi(x)d\varphi(z)d\varphi(w)d\varphi(v)$$

which is the order of $n^{-4}h^d$, where the last equality hold by change of variables on $(x-z)/h = u$, $(x-w)/h = v$, $(z-v)/h = w$, and the continuity of functions K, ξ over compact set, we have

$$EH_n^4(\tilde{X}_1, \tilde{X}_2) = \frac{h^{2d}}{n^4} E\left(\int K_{h1,2}(x)\xi(X_1)\xi(X_2)f^2(X_1)f^2(X_2)d\varphi(x)\right)^4$$
$$= \frac{h^{2d}}{n^4} \iiiint (EK_{h1}(x)K_{h1}(y)K_{h1}(z)K_{h1}(t)\xi(X_1)^4 f^8(X_1))^2 d\varphi(x)d\varphi(y)d\varphi(z)d\varphi(t),$$

which is the order of $O(n^{-4}h^{-2d})$. We also have

$$EH_n^2(\tilde{X}_1, \tilde{X}_2) = \frac{h^d}{n^2} \iint EK_{h1,2}(x)K_{h1}(y)K_{h2}(y)\xi^2(X_1)\xi^2(X_2)f^4(X_1)f^4(X_2)d\varphi(x)d\varphi(y)$$
$$= \frac{1}{n^2h^{3d}} \iiiint K\left(\frac{x-u}{h}\right)K\left(\frac{y-u}{h}\right)K\left(\frac{x-v}{h}\right)K\left(\frac{y-v}{h}\right)\xi^2(u)\xi^2(v)$$
$$\cdot f^5(u)f^5(v)dudvd\varphi(x)d\varphi(y),$$

which is $O(n^{-2})$.

So far we have shown that the conditions for the CLT are satisfied, and next it suffices to show the form of variance. In fact, we have

$$\frac{1}{2}n^2EH_n^2(\tilde{X}_1, \tilde{X}_2) = \frac{n\Gamma_n}{4(n-1)}$$
$$= \frac{1}{2}h^d \int\int \left(\int K(u)\frac{1}{h^d}K\left(\frac{y-x}{h}+u\right)\tau^2(x-uh)f^5(x-uh)du\right)^2 d\varphi(x)d\varphi(y)$$
$$\to \frac{1}{2}\int (\tau^2(x))^2 f^{10}(x)f^{-12}(x)g^2(x)dx \int \left(\int K(u)K(v+u)du\right)^2 dv$$
$$= \frac{1}{2}\int \frac{(\tau^2(x))^2g^2(x)}{f^2(x)}dx \int \left(\int K(u)K(v+u)du\right)^2 dv = \frac{\Gamma}{4},$$

by the Bounded Convergence theorem, and the continuity of $\tau^2(x)$, $f(x)$ and $g(x)$. $\square$

Lemma 2.5.6. Under H_0, if (e1), (e2), (f1), (f2), (k), (v3)-(v5) and (w) hold, then

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

Proof. Note that $T_n(\theta_0) - T_n(\hat{\theta}_n)$ can be written as $2Q_1 - Q_2$ with

$$Q_1 = \int \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)((Y_i - \hat{m}(X_i))^2 - v_0(X_i))\hat{f}_h^2(X_i)\right] \cdot \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)(v_0(X_i) - v(X_i, \hat{\theta}_n))\hat{f}_h^2(X_i)\right] d\hat{\varphi}_w(x)$$

$$Q_2 = \int \left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)(v_0(X_i) - v(X_i, \hat{\theta}_n))\hat{f}_h^2(X_i)\right]^2 d\hat{\varphi}_w(x).$$

It suffices to show that $nh^{d/2}Q_1 = o_p(1)$, and $nh^{d/2}Q_2 = o_p(1)$. Note that Q_1 can be written as

$$\int U_n(x)\left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)(v_0(X_i) - v(X_i, \hat{\theta}_n) - (u_n)^T\dot{v}_0(X_i) + (u_n)^T\dot{v}_0(X_i))\hat{f}_h^2(X_i)\right] d\hat{\varphi}_w(x)$$

which is the sum of the following two terms

$$Q_{11} = \int U_n(x)\left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)d_{ni}\right]\hat{f}_h^2(X_i)d\hat{\varphi}_w(x),$$
$$Q_{12} = (\hat{\theta}_n - \theta_0)^T \int U_n(x)\left[\frac{1}{n}\sum_{i=1}^{n} K_{hi}(x)\dot{v}_0(X_i)\right]\hat{f}_h^2(X_i)d\hat{\varphi}_w(x).$$

By Theorem 2.3.2, and $nh^dT_n(\theta_0) = O_p(1)$ we have $nh^d||\hat{\theta}_n - \theta_0||^2 = O_p(1)$. Therefore, by Cauchy-Schwartz inequality, assumption (v4), (k) and the fact that $\hat{f}_h(X_i)/\hat{f}_h(x) = O_p(1)$, $\int U_n(x)d\varphi(x) = O(1/nh^d)$, we have

$$nh^{d/2}|Q_{11}| \le \max_i \frac{|d_{ni}|}{||u_n||}(nh^d)^{1/2}||u_n|| \int \hat{f}_h^6(x)d\hat{\varphi}_w(x)\sqrt{n}\int U_n^2(x)d\hat{\varphi}_w(x) = o_p(1)$$

Next, observe that Q_{12} can be written as $Q_{121} - Q_{122}$, where

$$Q_{121} = u_n^T \int U_n(x) \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \dot{v}(X_i, \hat{\theta}_n) \right] \hat{f}_h^2(X_i) d\hat{\varphi}_w(x),$$
$$Q_{122} = u_n^T \int U_n(x) \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) (\dot{v}(X_i, \hat{\theta}_n) - \dot{v}_0(X_i)) \right] \hat{f}_h^2(X_i) d\hat{\varphi}_w(x).$$

Arguing as above, on the event $(nh^d)^{1/2}||u_n|| \leq k$, with Cauchy-Schwartz inequality,

$$\begin{aligned} n^2h^d|Q_{122}|^2 \leq n^2h^d||\hat{\theta}_n - \theta_0||^2 \int U_n^2(x) d\hat{\varphi}_w(x) \cdot \max_i ||\dot{v}(\hat{\theta}_n, X_i) - \dot{v}(\theta_0, X_i)||^2 d\hat{\varphi}_w(x) \\ \cdot \int \left(\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x) \hat{f}_h^2(X_i) \right)^2 d\hat{\varphi}_w(x) \\ = n^2h^d O_p\left(\frac{1}{n^2h^{2d}}\right) O_p\left(\frac{1}{nh^d}\right) O_p(h^d) O_P(1) = o_p(1). \end{aligned}$$

For Q_{121}, we can further rewrite it as the sum of D_1 and D_2, where

$$D_1 = (\hat{\theta}_n - \theta_0)^T \int Z_n(x, \hat{\theta}_n) \dot{\mu}_n(x, \theta_0) d\hat{\varphi}_w(x),$$
$$D_2 = (\hat{\theta}_n - \theta_0)^T \int Z_n(x, \hat{\theta}_n)(\dot{\mu}_n(x, \hat{\theta}_n) - \dot{\mu}_n(x, \theta_0)) d\hat{\varphi}_w(x),$$

and it is obvious to see that

$$nh^{d/2}D_1 \leq nh^{d/2}||\hat{\theta}_n - \theta_0||^2 \int Z_n^2(x, \hat{\theta}_n) d\hat{\varphi}_w(x) \int \hat{f}_h^2(x) d\hat{\varphi}_w(x) = o_p(1).$$

While using Theorem 5, and assumption (v5)

$$nh^{d/2}D_2 \leq nh^{d/2}||\hat{\theta}_n - \theta_0||^2 \int Z_n^2(x, \hat{\theta}_n) d\hat{\varphi}_w(x) \int \left[\frac{1}{n} K_{hi}(x)(\dot{v}(\hat{\theta}_n, X_i) - \dot{v}(X_i, \theta_0)) \right]^2 d\hat{\varphi}_w(x),$$

which is $o_p(1)$. The proof of Q_2 will be proceeded similar to Q_1. □

Lemma 2.5.7. Under H_0, (e1), (e2), (f1), (f2), (v3)-(v5), (w), we have

$$nh^{d/2}|T_n(\theta_0) - \tilde{T}_n(\theta_0)| = o_p(1)$$

Proof. Using Lemma 2.5.3, we have

$$nh^{d/2}|T_n(\theta_0) - \tilde{T}_n(\theta_0)| \le nh^{d/2} \int_{\mathcal{C}} U_n^2(x) d\varphi(x) \sup_{x \in \mathcal{C}} |f^6(x)/\hat{f}_w^6(x) - 1|$$
$$= nh^{d/2} O_p((nh^d)^{-1}) O_p((\log_k n)(\log n/n)^{d/(d+4)}) = o_p(1),$$

which implies the desired result. □

Lemma 2.5.8. Under the same condition as Lemma 2.5.6, $nh^{d/2}(\hat{C}_n - \tilde{C}_n) = o_p(1)$.

Proof. Since

$$\hat{C}_n = \frac{1}{n^2} \sum_{i=1}^n \int_{\mathcal{C}} K_{hi}^2(x) \hat{\xi}^2(X_i) \hat{f}_h^4(X_i) d\hat{\varphi}_w(x), \quad \tilde{C}_n = \frac{1}{n^2} \sum_{i=1}^n \int_{\mathcal{C}} K_{hi}^2(x) \xi^2(X_i) f^4(X_i) d\varphi(x).$$

As usual, for $\hat{C}_n$, by adding and subtracting the following quantities, $m(X_i)$, $v_0(X_i)$, $f(X_i)$, and $f^{-6}(x)$ in the corresponding places, i.e. $\hat{\xi}^2(X_i)$ for the first two, and $\hat{f}_h^4(X_i)$, $d\hat{\varphi}_w(x)$, respectively. Let $\Delta_{iv} := v_0(X_i) - v(X_i, \hat{\theta}_n)$, we have

$$\hat{C}_n = \frac{1}{n^2} \sum_{i=1}^n \int_{\mathcal{C}} K_{hi}^2(x)((Y_i - \hat{m}(X_i) - m(X_i) + m(X_i))^2 - v(X_i, \hat{\theta}_n) + v_0(X_i) - v_0(X_i))$$
$$(\hat{f}_h^4(X_i) - f^4(X_i) + f^4(X_i)) \left(\frac{1}{\hat{f}_w^6(x)} - \frac{1}{f^6(x)} + \frac{1}{f^6(x)} \right) dG(x)$$
$$= \frac{1}{n^2} \sum_{i=1}^n \int_{\mathcal{C}} K_{hi}^2(x) \xi^2(X_i) f^4(X_i) d\varphi(x) + \frac{1}{n^2} \sum_{i=1}^n \int_{\mathcal{C}} K_{hi}^2(x) \xi^2(X_i) \Delta_{if^4} d\varphi(x)$$
$$+ \frac{1}{n^2} \sum_{i=1}^n \int_{\mathcal{C}} K_{hi}^2(x) \xi^2(X_i) \Delta_{if^4} \left(\frac{1}{\hat{f}_w^6(x)} - \frac{1}{f^6(x)} \right) dG(x)$$

$$+\frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_{hi}^2(x)\xi^2(X_i)f^4(X_i)\left(\frac{1}{\hat{f}_w^6(x)}-\frac{1}{f^6(x)}\right)dG(x)$$
$$+\frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_{hi}^2(x)(\Delta_{im}^2)\Delta_{if^4}\left(\frac{1}{\hat{f}_w^6(x)}-\frac{1}{f^6(x)}\right)dG(x)$$
$$+\frac{2}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_{hi}^2(x)((Y_i-m(X_i))\Delta_{im}-\Delta_{iv})\Delta_{if^4}\left(\frac{1}{\hat{f}_w^6(x)}-\frac{1}{f^6(x)}\right)dG(x)$$
$$+\frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_{hi}^2(x)(\Delta_{im}^2)f^4(X_i)\left(\frac{1}{\hat{f}_w^6(x)}-\frac{1}{f^6(x)}\right)dG(x)$$
$$+\frac{2}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_{hi}^2(x)((Y_i-m(X_i))\Delta_{im}-\Delta_{iv})f^4(X_i)\left(\frac{1}{\hat{f}_w^6(x)}-\frac{1}{f^6(x)}\right)dG(x)$$
$$+\frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_{hi}^2(x)(\Delta_{im}^2)\Delta_{if^4}f^{-6}(x)dG(x)$$
$$+\frac{2}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_{hi}^2(x)((Y_i-m(X_i))\Delta_{im}-\Delta_{iv})\Delta_{if^4}f^{-6}(x)dG(x)$$
$$+\frac{1}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_{hi}^2(x)(\Delta_{im}^2)f^4(X_i)f^{-6}(x)dG(x)$$
$$+\frac{2}{n^2}\sum_{i=1}^{n}\int_{\mathcal{C}} K_{hi}^2(x)((Y_i-m(X_i))\Delta_{im}-\Delta_{iv})f^4(X_i)\left(\frac{1}{f^6(x)}\right)dG(x)$$
$$:= C_{n1}+\cdots+C_{n12}.$$

Note that C_{n1} is simply $\tilde{C}_n$. For $i = 2\ldots 12$, we will show that $nh^{d/2}C_{ni} = o_p(1)$. By Cauchy-Schwartz inequality, it is sufficient to show that $nh^{d/2}C_{nk} \to 0$ in probability for $k = 2, 4, 12$. For C_{n2}, we have

$$Enh^{d/2}C_{n2} = h^{d/2}\int_{\mathcal{C}} EK_{h1}^2(x)\xi^2(X_1)(\hat{f}_h^4(X_1)-f^4(X_1))d\varphi(x)$$
$$= h^{d/2}\int_{\mathcal{C}} EK_{h1}^2(x)\xi^2(X_1)\left[\left(\frac{1}{n}\sum_{j=1}^{n}K_{hj1}-f(X_1)+f(X_1)\right)-f^4(X_1)\right]d\varphi(x) = o(h^{4+d/2}).$$

By Lemma 2.5.3, expanding the numerator in the fraction part, also note that other functions are all continuous on the compact set $\mathcal{C}$,

$$Enh^{d/2}|C_{n4}| \leq h^{d/2} \int_{\mathcal{C}} EK_{h1}^2(x)\xi^2(X_1)f^4(x) \sup_{x\in\mathcal{C}} \left| \frac{f^6(x) - \hat{f}_w^6(x)}{\hat{f}_w^6(x)f^6(x)} \right| dG(x) = o(1).$$

Finally, for C_{n12}, it suffices to show that

$$Enh^{d/2}C_{n12} = 2h^{d/2} \int_{\mathcal{C}} EK_{h1}^2(x)((Y_1 - m(X_1))\Delta_{1m} - \Delta_{1v})f^4(X_1)d\varphi(x)$$
$$= 2h^{d/2} \int_{\mathcal{C}} EK_{h1}^2(x)((Y_1 - m(X_1))\Delta_{1m})f^4(X_1)d\varphi(x) - 2h^{d/2} \int_{\mathcal{C}} EK_{h1}^2(x)\Delta_{1v}f^4(X_1)d\varphi(x),$$

is of order $o(1)$.

The first part on the RHS is similar to T_{n2}^* with a simpler derivation, and the second part on the RHS converge to zero with assumption $(v5)$ by taking the term $\max \Delta_{1v}$ outside the integral. For $i = 3, 5, 6, 7, 8, 9, 10, 11$, we will use similar technique as the above three terms and obtaining $o_p(1)$. □

Lemma 2.5.9. Under the same conditions as Lemma 2.5.6, we have $\hat{\Gamma}_n - \Gamma = o_p(1)$.

Proof. Since we have already known that $\Gamma_n \to \Gamma$, so if suffices to show that $\hat{\Gamma}_n - \tilde{\Gamma}_n = o_p(1)$, $\tilde{\Gamma}_n - \Gamma_n = o_p(1)$, where,

$$\tilde{\Gamma}_n = \frac{h^d}{n^2} \sum_{i\neq j} \left(\int_{\mathcal{C}} K_{hi}(x)K_{hj}(x)\xi(X_i)\xi(X_j)f_h^2(X_i)f_h^2(X_j)f^{-6}(x)dG(x) \right)^2.$$

First we show that $\hat{\Gamma}_n - \tilde{\Gamma}_n = o_p(1)$. Note that

$$\hat{\Gamma}_n = \frac{h^d}{n^2} \sum_{i\neq j} \left(\int_{\mathcal{C}} K_{hi}(x)K_{hj}(x)\hat{\xi}(X_i)\hat{\xi}(X_j)\hat{f}_h^2(X_i)\hat{f}_h^2(X_j)\hat{f}_w^{-6}(x)dG(x) \right)^2$$
$$= \frac{h^d}{n^2} \sum_{i\neq j} \Bigg(\int_{\mathcal{C}} K_{hi}(x)K_{hj}(x)((Y_i - m(X_i) - \Delta_{im})^2 - v_0(X_i) + v_0(X_i) - v(X_i, \hat{\theta}_n)) \cdot$$
$$((Y_j - m(X_j) - \Delta_{jm})^2 - v_0(X_j) + v_0(X_j) - v(X_j, \hat{\theta}_n)) \cdot$$
$$(\Delta_{if^2} + f^2(X_i))(\Delta_{if^2} + f^2(X_j))(\hat{f}_w^{-6}(x) - f^{-6}(x) + f^{-6}(x))dG(x) \Bigg)^2.$$

After expanding the square terms in the above expression of $\hat{\Gamma}_n$, $\hat{\Gamma}_n$ can be written as the sum of $\tilde{\Gamma}_n$ and many other terms which include at least one term from Δ_{im}^2, Δ_{vi} and Δ_{if^2}. Hence, except for $\tilde{\Gamma}_n$, all other terms will be $o_p(1)$.

Finally, note that by Fubini Theorem,

$$\begin{aligned}
E\tilde{\Gamma}_n &= \frac{h^d}{n^2}\sum_{i\neq j} E\bigg(\int_{\mathcal{C}} K_{hi}(x)K_{hj}(x)\xi(X_i)\xi(X_j)f_h^2(X_i)f_h^2(X_j)f^{-6}(x)dG(x)\bigg)^2 \\
&= \frac{(n-1)h^d}{n}\int\int (E(K_{h1}(x)K_{h1}(y)\xi^2(X_1)f^4(X_1)))^2 d\varphi(x)d\varphi(y) = \Gamma_n
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
E(\tilde{\Gamma}_n - \Gamma_n)^2 &\leq \sum_{i\neq j} EH^4(\tilde{X}_i, \tilde{X}_j) + c\sum_{i\neq j\neq k} EH_n^2(\tilde{X}_i, \tilde{X}_j)H_n^2(\tilde{X}_j, \tilde{X}_k) \\
&\leq (n^2 + cn^3)EH^4(\tilde{X}_1, \tilde{X}_2) = O(1/nh^d).
\end{aligned}$$

This concludes the proof of Lemma 2.5.9. □

The proof of Theorem 2.3.5. Note that under the alternative hypothesis, we have $Y = m(X) + \sqrt{v_a(X)}\varepsilon$. So by adding and subtracting $v_a(X_i)$ in $T_n(\hat{\theta}_n)$, it can be written as $T_{nc1} + T_{nc2} + T_{nc3}$, where

$$\begin{aligned}
T_{nc1} &= \int_{\mathcal{C}} \left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\{(Y_i - \hat{m}(X_i))^2 - v_a(X_i)\}\hat{f}_h^2(X_i)\right]^2 d\hat{\varphi}_w(x), \\
T_{nc2} &= \int_{\mathcal{C}} \left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\{v_a(X_i) - v(X_i, \hat{\theta}_n)\}\hat{f}_h^2(X_i)\right]^2 d\hat{\varphi}_w(x),
\end{aligned}$$

and T_{nc3} is bounded above by $2\sqrt{T_{nc1}T_{nc2}}$ by Cauchy-Schwartz inequality.

By adding and subtracting $m(X_i)$, $f^2(X_i)$, and $f^{-6}(x)$ in the corresponding parts in T_{nc1}, and the dominating term would be

$$nh^{d/2}\tilde{T}_{nc1} = nh^{d/2}\int_{\mathcal{C}}\left[\frac{1}{n}\sum_{i=1}^n K_{hi}(x)\{(Y_i - m(X_i))^2 - v_a(X_i)\}f^2(X_i)\right]^2 d\varphi(x).$$

Denote $\xi_a(X_i) = (Y_i - m(X_i))^2 - v_a(X_i)$, and

$$T_{nc1} = \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)\{(Y_i - m(X_i))^2 - v_a(X_i)\} f^2(X_i) \right]^2 d\varphi(x),$$
$$\tilde{C}_{an} = \int_{\mathcal{C}} \left[\frac{1}{n^2} \sum_{i=1}^{n} K_{hi}^2(x) \xi_a^2(X_i) f^4(X_i) \right] d\varphi(x).$$

Then we can show that $nh^{d/2}(T_{nc1} - \tilde{C}_{an}) = nh^{d/2}(\tilde{T}_{nc1} - \tilde{C}_{an}) + o_p(1)$. Similar to the proof of Theorem 2.3.4, we further have $nh^{d/2}(T_{nc1} - \tilde{C}_{an}) \to N(0, \Gamma_a)$ with

$$\Gamma_a = 2 \int (\xi_a^2(x))^2 f^{-2}(x) g^2(x) dx \int \left(\int K(u)K(u+v)du \right)^2 dv.$$

It is also easy to show that $\hat{\Gamma}_n \to \Gamma_a$. The details are omitted for the sake of brevity.

For $nh^{d/2}T_{nc2} \to \infty$, as $n \to \infty$, we can show that $T_{nc2} \to \int_{\mathcal{C}} (v_a(x) - v(x, \theta_a))^2 dG(x)$ in probability. As for $\hat{C}_n$ and $\tilde{C}_{an}$, we can easily show that they are both the order of $O_p(1/(nh^d))$. Therefore, we also have $|T_{nc3}| \leq 2\sqrt{(T_{nc1} - \tilde{C}_{an} + \tilde{C}_{an})T_{nc2}} = o_p(1)$.

Finally, for n large enough, we have

$$nh^{d/2}\hat{\Gamma}_n^{-1/2} |T_n(\hat{\theta}_n) - \hat{C}_n| = nh^{d/2}\hat{\Gamma}_n^{-1/2} |T_{nc1} - \tilde{C}_{an} + T_{nc2} + T_{nc3} + \tilde{C}_{an} - \hat{C}_n|,$$

which is bounded below by $nh^{d/2}\hat{\Gamma}_n^{-1/2}(T_{nc2} + o_p(1))$. This, together with the result $\hat{\Gamma}_n \to \Gamma$, implies the desired result. □

Now let us prove Theorem 2.3.6. Denote $v_{loc}(x, \theta_0) := v_0(x) + c_n \delta(x)$. Under the alternative hypothesis, we have $(Y - m(X))^2 = v_{loc}(X, \theta_0)\varepsilon^2$. The key part of the proof is to show that $nh^{d/2}(\tilde{T}_n(\theta_0) - \tilde{C}_n) \to N(\int_{\mathcal{C}} \delta^2(x) dG(x), \Gamma)$ in distribution and $nh^{d/2}(\hat{C}_n - \tilde{C}_{nL}) = o_p(1)$, where

$$\tilde{C}_{nL} := \int_{\mathcal{C}} \frac{1}{n^2} \sum_{i=1}^{n} K_{hi}^2(x) \xi_{loc}^2(X_i) f^4(X_i) d\varphi(x). \tag{2.5.6}$$

Proof. The proof largely follows the similar arguments in the proofs of previous theorem, hence only the main steps are sketched here for the sake of brevity. First we have to show that $nh^{d/2}(T_n(\hat{\theta}_n) - \tilde{T}_n(\theta_0)) = o_p(1)$, where

$$\tilde{T}_n(\theta_0) = \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)[(Y_i - m(X_i))^2 - v_0(X_i)] f^2(X_i) \right]^2 d\varphi(x).$$

Denote $(Y_i - m(X_i))^2 - v_{loc}(X_i, \theta_0) = \xi_{loc}(X_i)$, $\tilde{T}_n(\theta_0)$ can be further written as the sum of the following three terms

$$\begin{aligned}
T_{nL1}(\theta_0) &= \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)[(Y_i - m(X_i))^2 - v_{loc}(X_i, \theta_0)] f^2(X_i) \right]^2 d\varphi(x) \\
T_{nL2}(\theta_0) &= \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)[v_{loc}(X_i, \theta_0) - v_0(X_i)] f^2(X_i) \right]^2 d\varphi(x) \\
T_{nL3}(\theta_0) &= 2 \int_{\mathcal{C}} \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)[(Y_i - m(X_i))^2 - v_{loc}(X_i, \theta_0)] f^2(X_i) \right] \\
&\qquad \left[\frac{1}{n} \sum_{i=1}^{n} K_{hi}(x)[v_{loc}(X_i, \theta_0) - v_0(X_i)] f^2(X_i) \right] d\varphi(x).
\end{aligned}$$

For $T_{nL1}(\theta_0)$, it has the decomposition $nh^{d/2} T_{nL1}(\theta_0) = nh^{d/2}(\tilde{C}_{nL} + M_n)$, where

$$\begin{aligned}
\tilde{C}_{nL} &= \int_{\mathcal{C}} \frac{1}{n^2} \sum_{i=1}^{n} K_{hi}^2(x) \xi_{loc}^2(X_i) f^4(X_i) d\varphi(x) \\
M_n &= \int_{\mathcal{C}} \frac{1}{n^2} \sum_{i \neq j} K_{hi}(x) K_h(X_j - x) \xi_{loc}(X_i) f^2(X_i) \xi_{loc}(X_j) f^2(X_j) d\varphi(x).
\end{aligned}$$

We can further show that $nh^{d/2}(T_{nL1}(\theta_0) - \tilde{C}_{nL}) \xrightarrow{d} N(0, \Gamma)$, and $nh^{d/2}(\hat{C}_n - \tilde{C}_{nL}) = o_p(1), \hat{\Gamma}_n \to \Gamma$ under the local alternative hypothesis. Therefore, the desired result can be obtained after proving that $nh^{d/2} T_{nL2} \to \int_{\mathcal{C}} \delta^2(x) dG(x)$ in probability, and $nh^{d/2} T_{nL3} = o_p(1)$. As we mentioned before, the arguments are similar to those used in the proof of Theorem 2.3.4 and 2.3.5. □

Chapter 3

Checking Variance Function via Empirical Smoothing with Unknown Regression Function

3.1 Introduction

In statistical regression, we often face the problem of heteroscadesticity, and efficiency of estimation will be heavily affected if it is not properly treated. In application, there are several tests and graphical methods that can help determine the existence of the heteroscadesticity, but the methods of eliminating them often depends on experience. Thus knowing the form of the variance function may be helpful.

There are indeed numerous researches focused on such topics. Indeed, checking parametric form of variance function can be viewed as a generalization of homoscadesticity checking since constant is a special case of parametric structure. In this chapter, we study the regression model $Y = m(X) + \sqrt{v(X)}\varepsilon$ where the independent variables are random. Therefore, existing researches like Dette et al. (1999) and Dette et al. (1998) and Dette (2002) whose tests are constructed under fixed design will not be considered. Speaking of testing variance function, under standardized residual and independence of the X variables, with the derivation $(Y - m(X))^2 = v(X, \theta) + \xi$ (ξ is identically distributed with mean zero and finite variance), majority of the variance checking can be turned into a lack of fit test for the regression function. Therefore the methods commonly used in developing nonparametric lack of fit tests parameter can be easily applied in the current problem. Briefly speaking, such methodology are mainly classified as either distance-based test with empirical smoothing over the conditional residual, or empirical residual processes test with Kolmogorov-Smirnov (KS) and Cramér-von Mises (CM) distance over the distribution functions. For the latter type of tests, examples can be found in Neumeyer and Van Keilegom (2010), where they constructed tests based on estimating the error distribution in the nonparametric multiple regression setting, and both the KS, CM distance are used to propose tests to validate the model. The local power of the test is obtained at the convergence rate of $n^{-1/2}$. Other than their general model validity tests, Zhu et al. (2001) constructed CM type test to check the homoscadesticity while assuming a parametric form of the regression function. Following their footstep, Dette et al. (2007) proposed tests for checking a more general problem, the parametric form of the variance function. Based on the empirical processes of the residual, Dette et al. (2007) showed that both KS and CM type tests asymptotically follow a complicated distribution related to a standard Gaussian process under the null hypothesis, while in their simulation studies, the CM type tests is performed to show satisfactory finite sample power. Later, Dette and Hetzler (2009) constructed a test that checks the same problem as in Dette et al. (2007) utilizing the empirical process of the pseudo-residuals, where the limiting process is a Brownian Bridge under the null hypothesis. An obvious drawback on these fruitful research is that majority of the tests will not provide a one-step calculable statistic. And such methods involving bootstrap may further hinder its application on large scale data set. However most of them indeed enjoy controlled empirical levels and reasonably large power according to their finite sample simulations.

For the distance-based test, one example on checking the goodness of fit on regression function can be found in Koul and Ni (2004). Here they assumed random design, and proposed the minimum-distance test. Following their model specification, Samarakoon and Song (2011) constructed a test for variance function also using minimum-distance with assumption on previous knowledge of the regression function. Such an idea of using distance can also be seen similarly in Dette et al. (1998), where the test of homoscedasticity is constructed based on L_2 approximation of the variance function. The more recent Pardo-Fernández and Jiménez-Gamero (2019) proposed test based on the L_2 distance between two characteristic functions, one is the nonparametric standard residual approximating the real variance, and the other the semi-parametric standard residual for the null hypothesis. From the previous chapters, we see that the method of minimum distance achieved both the asymptotic normality, and shows satisfactory empirical power under different scenarios. However, the drawbacks are also obvious, a restriction on the dimensions, and quite heavy burden on the computation. The choice of the integrating weight function $g(x)$ also brings uncertainty in real application. The method of minimum distance also requires compact support for the kernel function and the integration range, therefore when performing the test, choosing domain for integration remains a problem when samples are believed to have density on infinite support. At the same time, choice of kernel function are limited as well. In this research, we will again focus on the similar hypothesis as in previous chapter, the methodology however, uses Zheng (1996), where he proposed test statistics to check the parametric form of the mean function in a regression model based on the conditional expectation of the residual with the empirical smoothing technique. In Zheng (2009), he proposed a test utilizing the same methodology to check homoscedasticity, but still need an assumption for compact integrating range. We carry similar techniques but will have less restrictions on the choice of kernel function and possessing infinite support compared to many previous research. Plus, we tested a more general hypothesis. Under this framework, Song and Du (2011) improved Zheng (1996)'s test on regression function by shrinking the variance, and the same methodology with moderate modification is used in Samarakoon and Song (2012) as well. This idea could indeed, be a future improvement on the current proposed method, but will not be discussed in this research due to its complexity.

As mentioned earlier, we begin by consider the following regression model

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

where X is a d-dimensional co-variate, Y is a scalar response variable, $m(x)$ and $v(x) > 0$ are both measurable with respect to x, and the regression error ε are subjected to the standardized version, i.e. $E(\varepsilon|X) = 0$ and $E(\varepsilon^2|X) = 1$. In this research, we will assume that the mean regression function $E(Y|X = x) = m(x)$ remains unknown, and our focus is to develop a test procedure that check whether the variance function $\text{Var}(Y|X = x) = v(x)$ has a parametric form. To be specific, we want to test

$$H_0 : P(Var(Y|X) = v(X, \theta_0)) = 1 \quad \text{for some } \theta_0 \in \Theta$$

versus

$$H_1 : P(Var(Y|X) = v(X, \theta)) < 1 \quad \text{for all } \theta \in \Theta.$$

Following ideas from Zheng (1996), we have the θ_0 under the null hypothesis to minimize the mean squared error of the variance function, $\theta_0 = \text{argmin}_{\theta\in\Theta} E[(Y - m(X))^2 - v(X, \theta)]$, therefore a $\sqrt{n}$ consistent estimate of θ, $\hat{\theta}_n$ is used when performing the test procedure. Also note that $Var(Y|X) = E((Y - E(Y|X))^2|X) = E((Y - m(X))^2|X)$. We denote $\eta_i = (Y_i - m(X_i))^2 - v(X_i, \theta_0)$. To perform the test, we may discover that under null hypothesis, $E(\eta|X) = 0$, and under alternative hypothesis, $E(\eta|X) \neq 0$. This is further strengthened to consider the conditional expectation $E(E(\eta E(\eta|X)|X))$ where it is consistently greater than 0 under alternative hypothesis, i.e. $E(E^2(\eta|X)) > 0$. Next observe the following relationship,

$$E(\eta_i|X_i) = 0 \iff E(\eta_i f^2(X_i)|X_i) = 0 \iff E(\eta_i E(\eta_i f^2(X_i)|X_i) f^3(X_i)) = 0.$$

Instead of directly using the quantity $E(\eta_i|X_i)$ to construct our test statistic, we will approximate the quantity on the left hand side (LHS) of the above relations $E(\eta_i E(\eta_i f^2(X_i)|X_i) f^3(X_i))$. First, construct an empirical kernel smoothing with Nadaraya-Watson type estimate for the conditional expectation in the center $E(\eta_i f^2(X_i)|X_i)$. With the leave one out estimation scheme we will obtain $(n-1)^{-1} \sum_{j=1, j\neq i}^{n} K_{hij} \hat{\eta}_j \hat{f}_h^2(X_j)/(\hat{f}_h(X_i))$, where $\hat{f}_h(X_i) = (n - 1)^{-1} \sum_{j=1, j\neq i} K_{hij}$. Therefore an estimate for the entire quantity $E(\eta_i E(\eta_i f^2(X_i)|X_i) f^3(X_i))$ can be written as

$$V_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \hat{\eta}_i \hat{f}_h^2(X_i) \sum_{j=1, j\neq i}^{n} K_{hij} \hat{\eta}_j \hat{f}_h^2(X_j) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \hat{\eta}_i \hat{f}_h^2(X_i) K_{hij} \hat{\eta}_j \hat{f}_h^2(X_j),$$

where

$$\begin{aligned}
\hat{\eta}_j &= (Y_j - \hat{m}(X_j))^2 - v(X_i, \hat{\theta}_n) \\
&= (Y_j - m(X_j) + m(X_j) - \hat{m}(X_j))^2 - v(X_j, \theta_0) + v(X_j, \theta_0) - v(X_j, \hat{\theta}_n) \\
&= (Y_j - m(X_j))^2 - v(X_j, \theta_0) + [v(X_j, \theta_0) - v(X_j, \hat{\theta}_n)] + [m(X_j) - \hat{m}(X_j)]^2 \\
&\quad -2[m(X_j) - \hat{m}(X_j)][Y_j - m(X_j)] \\
&:= \eta_j + \Delta v_j + \Delta m_j^2 - 2(Y_j - m(X_j))\Delta m_j \\
&= \eta_j + \Delta v_j + \Delta m_j^2 - 2\Delta m_j \varepsilon_j \sqrt{v(X_j, \theta_0)}.
\end{aligned}$$

The advantage of the above construction of test statistic is that the randomness introduced by the Nadaraya-Watson estimator where the kernel estimation of density function in the denominator is eliminated. Therefore the condition proposed by Zheng (2009) regarding the uniform consistency of density estimator to the true density will be relaxed, and the restriction over compact sets follows from this condition is no longer required in our test. In the current research we test a more general problem, the parametric form of the variance function rather than homoscedasticity alone. Also, since we are testing the same hypothesis from the previous chapter, an indication from Song and Du (2011) states that the proposed centered minimum distance test is a general form of the current constructed test if replacing the $G(x)$ with empirical cumulative distribution function of X_i.

The paper will be organized as follow. In section 2 we will show the asymptotic normality of the test while assuming the pre-obtained $\sqrt{n}$ estimation of the parameter, and in section 3 we will develop the consistency and the local power of the test. In section 4 a simulation study is carried out to show the finite sample properties of the test, while all the details of the proofs are postponed to the end.

3.2 Main Results

In the context and proof that follows, we have these assumptions.

(D1) The sample vector $\{(Y_i, X_i)\}_{i=1}^n$ where $X_i \in R^d$ are random vectors with distribution function $F(y, x)$, $x \in R^d$, here, the density of X, $f(x)$ is assumed to be continuous and first order derivative uniformly bounded.

(D2) The function $E((\eta^2(X))|X = x) := \tau^2(x)$ is continuously differentiable with respect to X and is square integrable.

(D3) The function $E(\varepsilon^8|X)$ is continuously differentiable with respect to X and is square integrable.

(H) The bandwidth satisfies $h \to 0$, $nh^d \to \infty$, and $4 - d/2 > 0$.

Note that the above condition implies a restriction on dimension where $d < 8$.

(P1) The parameter space Θ is compact and convex in R^d, and the function $E((Y - m(X))^2 - v(X, \theta))$ may attain a unique minimum at $\theta_0 \in \Theta$.

(P2) $v(X, \theta)$ is twice continuously differentiable and the following quantities are bounded with $M < \infty$,

$$
\begin{aligned}
&E(\sup_{\theta\in\Theta} v^2(\theta, X)),\ E\left(\sup_{\theta\in\Theta}\left\|\frac{\partial v(X,\theta)}{\partial\theta}\frac{\partial v(X,\theta)}{\partial\theta^T}\right\|\right),\\
&E\left(\sup_{\theta\in\Theta}\left\|[(y-m(X))^2 - v(X,\theta)]^2\frac{\partial v(X,\theta)}{\partial\theta}\frac{\partial v(X,\theta)}{\partial\theta^T}\right\|\right),\\
&E\left(\sup_{\theta\in\Theta}\left\|[(y-m(X))^2 - v(X,\theta)]^2\frac{\partial^2 v(X,\theta)}{\partial\theta\partial\theta^T}\right\|\right).
\end{aligned}
$$

(K1) K is non-negative and bounded function symmetric around zero, which is unimodal, integrates to one, and satisfies $\int K(u)du = 1$, and $\int u^2K(u)du \neq 0$.

(K2) $\int K^2(u)du < \infty$.

To show the asymptotic normality of the test statistics V_n constructed in the previous section, similar methods with Zheng (1996) will be used. By the general second order U-statistic of the form $U_n := [n(n-1)]^{-1} \sum_{i=1}^{n} \sum_{j \neq i}^{n} H_n(z_i, z_j)$, we will obtain the distribution utilizing results from Hall (1984). In the U-statistics, z_i, $i = 1, ..., n$ are independent and identically distributed sample, and H_n is any symmetric function. For V_n, we have $z_i = (\eta_i, X_i^T)$, and $H_n = K_{hij} \hat{\eta}_i \hat{\eta}_j \hat{f}_h^2(X_i) \hat{f}_h^2(X_j)$.

Define the following,

$$r_n(z_i) := E(H_n(z_i, z_j)|z_i)$$

$$\overline{r}_n := E(r_n(z_i)) = E(H_n(z_i, z_j)) < \infty$$

$$\hat{U}_n := \overline{r}_n + \frac{2}{n} \sum_{i=1}^{n} [r_n(z_i) - \overline{r}_n]$$

We will work with the degenerate U-statistics, and its asymptotic distribution are depicted through the following theorem from Hall (1984) and Zheng (1996), it is reproduced here for the sake of completeness. Here, to proceed with the theorems in Hall (1984), denote $G_n(z_1, z_2) = E(H_n(z_3, z_1) H_n(z_3, z_2)|z_1, z_2)$, and we consider a one dimensional U-Statistics to be degenerate if $E(H_n(z_i, z_j)|z_i) = 0$ when $i \neq j$.

Theorem 3.2.1. If $E[\|H_n(z_i, z_j)\|^2] = o(n)$, then $\sqrt{n}(U_n - \hat{U}_n) = o_p(1)$, and $U_n = \overline{r}_n + o_p(1)$.

Theorem 3.2.2. Assume $E(H_n(z_1, z_2)|z_1) = 0$ almost surely and $E(H_n^2(z_1, z_2)) \leq \infty$ for each n. If

$$\lim_{n \to \infty} \frac{E(G_n^2(z_1, z_2)) + n^{-1} E(H_n^4(z_1, z_2))}{E(H_n^2(z_1, z_2))^2} \to 0,$$

then

$$\frac{nU_n}{\sqrt{2E(H_n^2(z_1, z_2))}} \xrightarrow{d} N(0, 1).$$

Before we proceed to our main results and proofs, there are a number of abbreviations needed for the simplicity purposes, and to make clear use of some special notations we listed the following remark, as well as one useful expansions that will often be encountered and cited in later text.

Remark. *(i) For the term* $\tilde{V}_{nijk}$ *for some* $i, j, k = 1, ...n$*, denotes the case where all estimate of the density function* $\hat{f}_h(x)$ *in* V_{nijk} *are replaced by the true density* $f(x)$.

(ii)$K_h(X_i - X_j) := K_{hij}$, $\hat{f}_h(X_i) := \hat{f}_i$, $f(X_i) := f_i$, $m(X_i) := m_i$, $v(X_i, \theta_0) := v_{i,0}$, $\sigma(X_i) := \sigma_i$.

(iii)$\sum_{i \neq j}^{n}$ *or any* $\sum_{i \neq j \neq k}^{n}$ *only means the sum over the first index number* i*, i.e. the terms are* $n-1$ *and* $n-2$ *respectively. In this paper, it does NOT mean* $i \neq j$ *and both* i *and* j *range from* 1 *to* n.

(iv) the expansion of $(\sum_{i=1}^{n} X_i)^4$*, and its expectation with precise coefficient.*

$$\Big(\sum_{i=1}^{n} X_i\Big)^4 = \sum_{i=1}^{n} X_i^4 + \sum_{i=1}^{n}\sum_{j \neq i}^{n}(X_i^3 X_j + X_i^2 X_j^2) + \sum_{i=1}^{n}\sum_{j \neq i}^{n}\sum_{k \neq j \neq i}^{n} X_i^2 X_j X_k + \sum_{i=1}^{n}\sum_{j \neq i}^{n}\sum_{k \neq j \neq i}^{n}\sum_{l \neq k \neq j \neq i}^{n} X_i X_j X_k X_l \tag{3.2.1}$$

$$\begin{aligned} E(\sum_{i=1}^{n} X_i)^4 &= E\binom{n}{1}\binom{1}{1}X_i^4 + E\binom{n}{2}\binom{4}{3}\binom{2}{1}X_i^3 X_j + E\binom{n}{2}\binom{4}{2}\binom{2}{2}X_i^2 X_j^2 \\ &\quad + E\binom{n}{3}\binom{4}{2}\binom{3}{1}\binom{2}{1}X_i^2 X_j X_k + E\binom{n}{4}\binom{4}{1}\binom{3}{1}\binom{2}{1}X_i X_j X_k X_l \\ &= nEX_1^4 + 4n(n-1)EX_1^3 X_2 + 3n(n-1)EX_1^2 X_2^2 + 6n(n-1)(n-2)EX_1^2 X_2 X_3 \\ &\quad + n(n-1)(n-2)(n-3)EX_1 X_2 X_3 X_4, \end{aligned} \tag{3.2.2}$$

where we can check that the coefficient for each expectation adds up to n^4.

Theorem 3.2.3. Under all the assumptions listed, and the null hypothesis, $nh^{d/2}|\hat{\Sigma}|^{-1/2}V_n \to N(0,1)$ in distribution, where $\hat{\Sigma} = 2(n(n-1))^{-1}\sum_{i=1}^{n}\sum_{j \neq i}^{n} h^d K_{hij}^2 \hat{\eta}_i^2 \hat{f}_h^4(X_i)\hat{\eta}_j^2 \hat{f}_h^4(X_j)$

Proof of Theorem 3.2.3. We will show that $\hat{\Sigma}$ does converge to Σ in probability. Indeed, the statistic V_n will converge in distribution to $N(0, \Sigma)$ at the speed of $nh^{d/2}$.

To derive the asymptotic distribution of the test statistics V_n, we expand it to the following 10 terms, then we will show the asymptotic distribution with the lemmas following this expansion.

$$
\begin{aligned}
V_{n1} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i\eta_j \hat{f}_i^2\hat{f}_j^2 \\
V_{n2} &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i\Delta m_j^2 \hat{f}_i^2\hat{f}_j^2 \\
V_{n3} &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i\Delta v_j \hat{f}_i^2\hat{f}_j^2 \\
V_{n4} &= \frac{-4}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i\Delta m_j\varepsilon_j\sqrt{v(X_j,\theta_0)}\hat{f}_i^2\hat{f}_j^2 \\
V_{n5} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Delta m_i^2\Delta m_j^2 \hat{f}_i^2\hat{f}_j^2 \\
V_{n6} &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Delta v_i\Delta m_j^2 \hat{f}_i^2\hat{f}_j^2 \\
V_{n7} &= \frac{-4}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Delta m_i^2\Delta m_j\varepsilon_j\sqrt{v(X_j,\theta_0)}\hat{f}_i^2\hat{f}_j^2 \\
V_{n8} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Delta v_i\Delta v_j \hat{f}_i^2\hat{f}_j^2 \\
V_{n9} &= \frac{-4}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Delta v_i\Delta m_j\varepsilon_j\sqrt{v(X_j,\theta_0)}\hat{f}_i^2\hat{f}_j^2 \\
V_{n10} &= \frac{4}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Delta m_i\varepsilon_i\sqrt{v(X_i,\theta_0)}\Delta m_j\varepsilon_j\sqrt{v(X_j,\theta_0)}\hat{f}_i^2\hat{f}_j^2
\end{aligned}
$$

The general road map for proving the theorem is, under different conditions, we have $nh^{d/2}V_{ni} = o_p(1)$ for $i = 2, ..., 10$, and $nh^{d/2}V_{n1} \to N(0, \Sigma)$ in distribution.

Next we will provide several lemmas that will facilitate the proof of the theorem. The proof of the lemmas are put in the appendix for the sake of brevity.

Lemma 1. *Under assumption (D1), (D2), (H), (K1), (K2), (P1), (P2), $nh^{d/2}V_{n1} \to N(0, \Sigma)$, and $nh^{d/2}V_{ni} = o_p(1)$, for $i = 3, 8$.*
where Σ is the asymptotic variance that equals $2\int K^2(u)du \cdot \int [\tau^2(x)]^2 f^{10}(x)dx$

In the above lemma, we show the major part of the test statistics where the variance and the asymptotic distribution of the test is obtained. Part 3 and 8 of this decomposition contains the difference between $\sqrt{n}$ consistency estimation of the parameter and the real parameter itself. And such consistency leads to the order of $o_p(1)$ in the results.

Lemma 2. *Under assumptions (D1), (D2),(D3), (H), (K1), (K2),* $nh^{d/2}V_{ni} = o_p(1)$ *for* $i = 5, 2, 6$.

In this part of the proof, we may see restrictions where there is quantity of order $O_p(h^{4-d/2})$, to eliminate the bias in the asymptotic distribution, we would have this term converges to zero. Therefore, we assume $4 - d/2 > 0$, and this further implies $d < 8$, which is a relatively loose dimension restriction comparing to the previous chapter. This indeed, tells us that substituting regression function with NW estimator will bring restrictions to the application of the test. In some existing sufficient dimension reduction research such as Li et al. (2019), we may discover that under kernel smoothing methodology, dimensions higher than eight are considered rather inefficient in estimation, so is the test at the same time, may converge at a very slow rate. Therefore, a reasonable approach with higher dimensions would be to perform dimension reduction before the estimation and testing.

Lemma 3. *Under assumptions (D1), (D2), (D3), (H), (P1), (P2), and (K2),* $nh^{d/2}V_{ni} = o_p(1)$ *for* $i = 4, 7, 9, 10$

Lemma 4. *Under all the assumptions and with the conclusions in Lemma 2,* $\hat{\Sigma} \to \Sigma$ *in probability.*

The last two lemmas provides general results that leads to no further bias in the asymptotic distribution and shows that we can obtain a consistent variance estimation where we may further use it in performing the test. After having these results, from Lemma 1, Lemma 2, and Lemma 3, we know that $nh^{d/2}V_n \to N(0, \Sigma)$ in distribution, and by Lemma 4, and the Slutsky's theorem, we obtain a calculable statistics, and we have completed the proof. □

3.3 Consistency and Local Power

Under alternative hypothesis, we discover the fact that the quantity $E(\eta E(\eta f^2(X)|X)f^3(X)) > 0$, and its empirical estimate V_n may converge to a constant greater than zero as $n \to \infty$. That is, the test will consistently reject the alternative hypothesis. When performing the test, we will need to obtain a consistent estimate for the parameter in the variance function under null hypothesis. According to Jennrich (1969), estimation to the parameter θ_a under the form $v(X_i, \theta_a)$ is $\sqrt{n}$ consistent, i.e. $\sqrt{n}(\hat{\theta}_n - \theta_a) = O_p(1)$.

Theorem 3.3.1. Given all the assumptions listed, under alternative hypothesis $v_a(X)$, the test reject H_0 whenever

$$\left| nh^{d/2}\sqrt{\frac{n-1}{n}} V_n/\sqrt{\hat{\Sigma}} \right| > z_{\alpha/2}$$

is consistent for H_a, with $0 < \alpha < 1$.

For situations where alternatives are not consistently different from the null, we may discover a different power performance of the test. To see how sharp the test is against those alternative hypotheses, we use the local alternative of the form $v_{loc}(x) = v(X, \theta_0) + \delta_n l(X)$, where $\delta_n \to 0$ as $n \to \infty$. We can modify the order of δ_n to see if the test identifies the mis-specification or not. To be specific, we have the following result,

Theorem 3.3.2. Under the same assumption as consistency, if $\delta_n = 1/\sqrt{(nh^{d/2})}$, and $\int l^2(x)dx > 0$

$$nh^{d/2} V_n/\sqrt{\hat{\Sigma}} \to N\Big(\int l^2(x) f^6(x) dx/\sqrt{\Sigma}, 1 \Big) \tag{3.3.1}$$

in distribution.

That is, if $\delta_n = o(1/\sqrt{nh^{d/2}})$, then the test may fail to reject the alternative hypothesis as sample size goes to infinity, and if $1/\sqrt{nh^{d/2}} = o(\delta_n)$, then the test will reject the hypothesis consistently. The power of the test under the exact rate is $1 - \Phi(\int l^2(x)f^6(x)dx/\sqrt{\Sigma} - z_{\alpha/2}) + \Phi(\int l^2(x)f^6(x)dx/\sqrt{\Sigma} + z_{\alpha/2})$, where $\Phi(x)$ is the cumulative distribution function of the standard normal distribution. In the following simulation studies, we will also observe this pattern with the empirical power converges to a constant between 0.05 and 1

3.4 Simulations

We begin the simulation with the alternatives that covers different scenarios mentioned in the earlier context. Due to the construction of the test, the computation will be simpler and faster comparing to minimum distance test. Here, we choose the Epanechikov Kernel with domain $[-1, 1]$ for both the estimation and tests, and $h = n^{-1/4}$. In order to observe the effects of bandwidth under finite sample performance, we proposed a coefficient $ah = an^{-1/4}$, where $a = 0.8,\ 1,\ 1.2,\ 1.5$. Discussion on bandwidth selection with the power of the test can be seen in Gao and Gijbels (2008) for tests of regression function, where the properties on variance functions will not be a topic to be studied in this research. The model that generates the data is $Y = 1 + 2X + 3X^2 + \sqrt{1 + \theta X^2}e$, $\theta_0 = 0.5$ with X follows uniform $[0, 1]$, and e follows $N(0, 1)$. Before the testing procedure, we also need a $\sqrt{n}$ consistent estimation of the parameter, where ordinary least square estimate is used as in the simulation of Pardo-Fernández and Jiménez-Gamero (2019) for parameter θ. Treating $(Y - \hat{m}(X))^2$ as the response variable, and $v(X, \theta)$ in both the null and alternative hypothesis as linear regression function, we may obtain an estimate for θ. The simulation has run 500 times and calculate the proportion of rejection under the level 5%.

Alternative hypothesis are as follow, with $l(x) := (x^2 + 0.1)^{-1}$ for convenience with the fact that $\int (x^2 + 0.1)^{-2} dx < \infty$:

$$
\begin{aligned}
H_1 : \quad & v(x) = \exp(x+1);\ H_2 : v(x) = \frac{1}{x^2 + 0.1};\ H_3 : v(x) = v(x, \theta_0) + \frac{l(x)}{n^4}; \\
H_4 : \quad & v(x) = v(x, \theta_0) + \frac{l(x)}{\sqrt{nh^{d/2}}}; \qquad\qquad H_5 : v(x) = v(x, \theta_0) + \frac{l(x)}{\ln n}.
\end{aligned}
$$

Here we restate the test statistics used in this test, quantities include:

$$
\hat{T}_n := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K_{hij} \hat{\eta}_i \hat{f}_h^2(X_i) \hat{\eta}_j \hat{f}_h^2(X_j);\ \hat{\Sigma} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} h^d K_{hij}^2 \hat{\eta}_i^2 \hat{f}_h^4(X_i) \hat{\eta}_j^2 \hat{f}_h^4(X_j)
$$

where $\hat{\eta}_i = (Y_i - \hat{m}_i)^2 - 1 - \hat{\theta}_n X_i^2$

And in total, the test statistics is computed through the following quantities.

$$
\begin{aligned}
& nh^{d/2}(\hat{\Sigma})^{-1/2}\hat{T}_n\sqrt{\frac{n-1}{n}} \\
= \quad & \frac{[n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\hat{\eta}_i\hat{f}_h^2(X_i)\hat{\eta}_j\hat{f}_h^2(X_j)}{\{[2n(n-1)]^{-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} h^d K_{hij}^2\hat{\eta}_i^2\hat{f}_h^4(X_i)\hat{\eta}_j^2\hat{f}_h^4(X_j)\}^{1/2}} \cdot \frac{nh^{d/2}\sqrt{n-1}}{\sqrt{n}} \\
= \quad & \frac{\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\hat{\eta}_i\hat{f}_h^2(X_i)\hat{\eta}_j\hat{f}_h^2(X_j)}{\{\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}^2\hat{\eta}_i^2\hat{f}_h^4(X_i)\hat{\eta}_j^2\hat{f}_h^4(X_j)\}^{1/2}}
\end{aligned}
$$

Under the null hypothesis, the above test statistics converge asymptotically to $N(0,1)$, so is H_3 with faster local convergence rate as $1/n^4$. Both finite sample type I error are controlled under 0.05. Under H_1, H_2 and H_5, the empirical power would asymptotically goes to 1, which as expected, with H_2 having faster increase in empirical power than H_5. Under H_4, the local alternative, the empirical power will converge to a constant between 0 and 1. From the simulation results, we have observed finite sample performance that meets our expectation. The bandwidth will indeed affect the performance of the test, since this is a finite sample study, and under our choice of alternative functions, suggested bandwidth coefficient would be $a = 1$. But in real application, bandwidth still remains an open problem with the test. Therefore, cross validation may be suggested if there is a concern for bandwidth under small sample sizes.

Remark. *The choice of null and alternative hypotheses have some constant which we do not generalize as in our theoretical results. Since these are finite sample simulations, we need to take into consideration of the goodness in parameter estimation. When sample sizes are small, the estimation of the parameter may have greater probability in getting a negative value, therefore, the constant* 1 *added guarantees that even with this negative estimation, our variance function are still positive. These choices of variance function are commonly seen in many previous researches. And of course, in the bad estimation scenario, the test will generally reject the hypothesis. However, if we do not have this kind of modification, we will need to omit those simulation runs where variance function are negative, which doesn't sound reasonable and were not considered in our study.*

Table 3.1: Empirical Levels and Powers

Bandwidth	Hypothesis	100	200	300	400	500
	H_0	0.112	0.098	0.084	0.074	0.056
	H_1	0.356	0.664	0.84	0.936	0.992
	H_2	0.998	1.000	1.000	1.000	1.000
$a = 1.5$	H_3	0.116	0.114	0.108	0.090	0.060
	H_4	0.108	0.140	0.134	0.172	0.152
	H_5	0.596	0.844	0.966	0.976	1.000
	H_0	0.072	0.072	0.052	0.054	0.044
	H_1	0.358	0.662	0.828	0.918	0.978
	H_2	0.994	0.998	1.000	1.000	1.000
$a = 1.2$	H_3	0.090	0.048	0.062	0.044	0.058
	H_4	0.096	0.104	0.118	0.140	0.150
	H_5	0.608	0.868	0.968	0.990	0.996
	H_0	0.064	0.030	0.030	0.026	0.036
	H_1	0.414	0.660	0.778	0.916	0.968
	H_2	0.998	1.000	1.000	1.000	1.000
$a = 1$	H_3	0.064	0.038	0.030	0.034	0.046
	H_4	0.116	0.150	0.114	0.124	0.130
	H_5	0.598	0.842	0.996	0.988	0.998
	H_0	0.052	0.030	0.032	0.030	0.032
	H_1	0.358	0.596	0.766	0.868	0.952
	H_2	0.996	0.998	1.000	1.000	1.000
$a = 0.8$	H_3	0.046	0.020	0.026	0.028	0.032
	H_4	0.106	0.106	0.108	0.076	0.102
	H_5	0.596	0.846	0.946	0.988	0.994

3.5 Proof of the Lemmas and Theorems

Proof of Lemma 1. First, we add and subtract $f^2(X_i)$, and $f^2(X_i)$ in V_{n1}, we obtain that

$$\begin{aligned} V_{n1} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K_{hij} \eta_i \eta_j f^2(X_i) f^2(X_j) \\ &+ \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K_{hij} \eta_i \eta_j f^2(X_i) (\hat{f}_h^2(X_j) - f^2(X_j)) \\ &+ \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K_{hij} \eta_i \eta_j (\hat{f}_h^2(X_i) - f^2(X_i)) (\hat{f}_h^2(X_j) - f^2(X_j)) \\ &:= V_{n11} + 2V_{n12} + V_{n13}. \end{aligned}$$

We will show that $nh^{d/2}V_{n11}$ follows an asymptotic normal distribution with the desired variance, while $nh^{d/2}V_{n12}$ and $nh^{d/2}V_{n13}$ are $o_p(1)$.

$$\begin{aligned} nh^{d/2}V_{n12} &= \frac{h^{d/2}}{n-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i\eta_j f^2(X_i)((\hat{f}_h(X_j)-f(X_j)+f(X_j))^2 - f^2(X_j)) \\ &= \frac{h^{d/2}}{n-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i\eta_j f^2(X_i)(\hat{f}_h(X_j)-f(X_j))^2 \\ &\quad + \frac{h^{d/2}}{n-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i\eta_j f^2(X_i)(2f(X_j)(\hat{f}_h(X_j)-f(X_j))). \end{aligned}$$

By Chebyshev's inequality, $E(n^2h^dV_{n12}^2) = o(1)$ implies $nh^{d/2}V_{n12} = o_p(1)$. Next we will show the above quantity is less than the the sums of the following quantities that all converges to zero. With C-R inequality, we have,

$$\begin{aligned} E(n^2h^dV_{n12}^2) &\leq 2E\left(\frac{h^{d/2}}{n-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i\eta_j f^2(X_i)(\hat{f}_h(X_j)-f(X_j))^2\right)^2 \\ &\quad + 2E\left(\frac{h^{d/2}}{n-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i\eta_j f^2(X_i)(2f(X_j)(\hat{f}_h(X_j)-f(X_j)))\right)^2 \\ &= \frac{2}{(n-1)^2}E\left(h^d\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}^2\eta_i^2\eta_j^2 f^4(X_i)(\hat{f}_h(X_j)-f(X_j))^4\right) \\ &\quad + \frac{2}{(n-1)^2}E\left(h^d\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i^2\eta_j^2 f^4(X_i)(4f^2(X_j)(\hat{f}_h(X_j)-f(X_j))^2)\right) \\ &:= 2EV_{n121} + 2EV_{n122}. \end{aligned}$$

The last equality hold due to a fact that will be continuously used throughout the proof. Any pair of (i,j) with $\eta_i\eta_j$ that have subscripts completely unequal, say $\eta_i\eta_j\cdot\eta_i\eta_k$, or $\eta_i\eta_j\cdot\eta_k\eta_l$, will have conditional expectation equal to zero. And terms with any first order $\eta_i's$ under the null hypothesis will be zero as well after conditioning on X_i, i.e. $E(E(\eta_i|X_i)) = 0$. We can separate the conditional expectation into the product of $E(\eta_i|X_i)$ and other functions of X_i with the assumption that the samples are independent of each other.

$$\begin{aligned}
EV_{n121} &= \frac{nh^d}{n-1}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^4(X_1)(\hat{f}_h(X_2)-f(X_2))^4\Bigg) \\
&= \frac{nh^d}{n-1}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^4(X_1)\Bigg(\frac{1}{n-1}\sum_{i\neq 2}^{n}K_{hi2}-E(K_{h32}|X_2)+E(K_{h32}|X_2)-f(X_2)\Bigg)^4\Bigg) \\
&\leq \frac{8nh^d}{n-1}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^4(X_1)\Bigg(\frac{1}{n-1}\sum_{i\neq 2}^{n}K_{hi2}-E(K_{h32}|X_2)\Bigg)^4\Bigg) \\
&\quad +\frac{8nh^d}{n-1}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^4(X_1)(E(K_{h32}|X_2)-f(X_2))^4\Bigg).
\end{aligned}$$

Next, we will use the following derivations very often, with assumptions that $f(X)$ has finite second order derivative,

$$\begin{aligned}
E(K_{h32}|X_2) &= \int \frac{1}{h^d}K\Bigg(\frac{x_3-X_2}{h}\Bigg)f(x_3)dx_3 \\
&= \int K(u)f(X_2+uh)du = \int K(u)\Big(f(X_2)+uhf'(X_2)+\frac{h^2}{2}u\frac{\partial^2 f(\tilde{X}_2))}{\partial X\partial X^T}u^T\Big)du \\
&= f(X_2)+O(h^2).
\end{aligned}$$

Plug in the result, we have,

$$\begin{aligned}
& EV_{n121} \leq \frac{8nh^d}{n-1}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^4(X_1)\Bigg(\frac{1}{n-1}\sum_{i\neq 2}^{n}K_{hi2}-E(K_{h32}|X_2)\Bigg)^4\Bigg) \\
& +\frac{8nh^d}{n-1}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^4(X_1)(f(X_2)-f(X_2)+O(h^2))^4\Bigg) \\
= & \frac{8nh^d}{(n-1)^5}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^4(X_1)\Bigg(\sum_{i\neq 2}^{n}(K_{hi2}-E(K_{h32}|X_2))\Bigg)^4\Bigg) \\
& +\frac{8nh^d}{n-1}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^4(X_1)(O(h^2))^4\Bigg)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{8nh^d}{(n-1)^5} E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 f^4(X_1) \Bigg(\sum_{i\neq 2}^n (K_{hi2} - E(K_{h32}|X_2))^4 \\
&\quad + \sum_{i\neq 2}^n \sum_{j\neq i\neq 2}^n (K_{hi2} - E(K_{h32}|X_2))^2 (K_{hj2} - E(K_{h42}|X_2))^2 \Bigg)\Bigg) \\
&\quad + \frac{8nh^d}{n-1} E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 f^4(X_1)(O(h^2))^4 \Bigg) \\
&= \frac{8nh^d}{(n-1)^4} E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 f^4(X_1)((K_{h32} - E(K_{h32}|X_2))^4 \Bigg) \\
&\quad + \frac{8n(n-2)h^d}{(n-1)^4} E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 f^4(X_1)(K_{h32} - E(K_{h32}|X_2))^2 (K_{h42} - E(K_{h42}|X_2))^2) \Bigg) \\
&\quad + \frac{8nh^d}{n-1} E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 f^4(X_1) \Bigg) O(h^8).
\end{aligned}
$$

As for the above, note that the last term indeed is of order $O(h^8)$ after taking expectation. For the other tow, we may replace $E(K_{h32}|X_2)$ and $E(K_{h42}|X_2)$ with $f(X_2) + O(h^2)$, this is obtained by usual calculation. Next integrate all the variables, with slight abuse of notations that is only used in the following equation, $K_{hij} = K_h(x_i - x_j)$. We have,

$$
\begin{aligned}
EV_{n121} &\leq \frac{8nh^d}{(n-1)^4} \int K_{h12}^2 \eta_1^2 \eta_2^2 f^5(x_1)((K_{h32} - f(x_2) + O(h^2))^4 f(x_2) f(x_3) dx_1 dx_2 dx_3 \\
&\quad + \frac{8n(n-2)h^d}{(n-1)^4} \int K_{h12}^2 \eta_1^2 \eta_2^2 f^5(x_1)(K_{h32} - f(x_2) + O(h^2))^2 (K_{h42} - f(x_2) + O(h^2))^2) \\
&\qquad \cdot f(x_2) f(x_3) f(x_4) dx_1 dx_2 dx_3 dx_4 + O(h^8) \\
&= O\Bigg(\frac{1}{n^3 h^{3d}} \Bigg) + O\Bigg(\frac{1}{n^2 h^{2d}} \Bigg) + O(h^8).
\end{aligned}
$$

With almost the same procedure but less tedious details, adding and subtracting $E(K_{h32}|X_2)$ inside the term $(\hat{f}_h(X_2) - f(X_2))$, we may find that $n^2 h^d V_{n122}$ is much easier to prove, whereas

$$
\begin{aligned}
V_{n122} &= \frac{4}{(n-1)^2} E\Bigg(h^d \sum_{i=1}^n \sum_{j\neq i}^n K_{hij}^2 \eta_i^2 \eta_j^2 f^4(X_i)(f^2(X_j)(\hat{f}_h(X_j) - f(X_j))^2) \Bigg) \\
&= \frac{4nh^d}{n-1} E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 f^4(X_1)(f^2(X_2)(\hat{f}_h(X_2) - f(X_2))^2) \Bigg)
\end{aligned}
$$

$$
\begin{aligned}
&\le \frac{8nh^d}{(n-1)^3}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^4(X_1)f^2(X_2)(n(K_{h32}-E(K_{h32}|X_2))^2+n(n-1)O(h^4)\Bigg) \\
&= O\Bigg(\frac{1}{nh^d}\Bigg)+O(h^4).
\end{aligned}
$$

Now, we have shown that both V_{n121} and V_{n122} are $o_p(1)$. Next, with the fact that $E(\eta_i|X_i)=0$ and X_i, X_j independent of each other,

$$
\begin{aligned}
E(n^2h^dV_{n13}^2) &= \frac{h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^n\sum_{j\neq i}^n K_{hij}\eta_i\eta_j(\hat{f}_h^2(X_i)-f^2(X_i))(\hat{f}_h^2(X_j)-f^2(X_j))\Bigg)^2 \\
&= \frac{h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^n\sum_{j\neq i}^n K_{hij}^2\eta_i^2\eta_j^2(\hat{f}_h^2(X_i)-f^2(X_i))^2(\hat{f}_h^2(X_j)-f^2(X_j))^2\Bigg) \\
&= \frac{nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2(\hat{f}_h^2(X_1)-f^2(X_1))^2(\hat{f}_h^2(X_2)-f^2(X_2))^2\Bigg) \\
&= \frac{nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2((\hat{f}_h(X_1)-f(X_1))^2+2f(X_1)(\hat{f}_h(X_1)-f(X_1)))^2 \\
&\qquad\qquad \cdot((\hat{f}_h(X_2)-f(X_2))^2+2f(X_2)(\hat{f}_h(X_2)-f(X_2)))^2\Bigg) \\
&\le \frac{4nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2((\hat{f}_h(X_1)-f(X_1))^4+4f^2(X_1)(\hat{f}_h(X_1)-f(X_1))^2) \\
&\qquad\qquad \cdot((\hat{f}_h(X_2)-f(X_2))^4+4f^2(X_2)(\hat{f}_h(X_2)-f(X_2))^2)\Bigg) \\
&= \frac{4nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2((\hat{f}_h(X_1)-f(X_1))^4(\hat{f}_h(X_2)-f(X_2))^4\Bigg) \\
&\quad +\frac{32nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^2(X_1)((\hat{f}_h(X_1)-f(X_1))^2(\hat{f}_h(X_2)-f(X_2))^4\Bigg) \\
&\quad +\frac{64nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2 f^2(X_1)f^2(X_2)((\hat{f}_h(X_1)-f(X_1))^2(\hat{f}_h(X_2)-f(X_2))^2\Bigg) \\
&:= \frac{4n}{n-1}V_{131}+\frac{32n}{n-1}V_{132}+\frac{64n}{n-1}V_{133}.
\end{aligned}
$$

We may work on V_{n131} first, where the rest carries out similar expansion with much less complexities.

$$
\begin{aligned}
V_{n131} &= h^d E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 ((\hat{f}_h(X_1) - f(X_1))^4 (\hat{f}_h(X_2) - f(X_2))^4 \Bigg) \\
&= h^d E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 \Big(\frac{1}{n-1} \sum_{i \neq 1}^{n} K_{hi1} - E(K_{h31}|X_1) + E(K_{h31}|X_1) - f(X_1) \Big)^4 \\
&\qquad \cdot \Big(\frac{1}{n-1} \sum_{j \neq 2}^{n} K_{hj2} - E(K_{h31}|X_2) + E(K_{h31}|X_2) - f(X_2) \Big)^4 \Bigg).
\end{aligned}
$$

Let's focus on the term $((n-1)^{-1}\sum_{j\neq 2}^{n} K_{hj2} - E(K_{h31}|X_2) + E(K_{h31}|X_2) - f(X_2))^4$, after expansion, we may see the pattern with the other part. And here we will use the previously described condition.

$$
\begin{aligned}
& \Bigg(\frac{1}{n-1} \sum_{j \neq 2}^{n} K_{hj2} - E(K_{h31}|X_2) + E(K_{h31}|X_2) - f(X_2) \Bigg)^4 \qquad (3.5.1) \\
= & \Bigg(\frac{1}{n-1} \sum_{j=3}^{n} [K_{hj2} - E(K_{h31}|X_2)] + [E(K_{h31}|X_2) - f(X_2)] + K_{h12} \Bigg)^4 \\
\leq & \frac{8}{(n-1)^4} \Bigg(\Big(\sum_{j=3}^{n} [K_{hj2} - E(K_{h31}|X_2)] \Big) + K_{h12} \Bigg)^4 + \Bigg(O(h^2) \Bigg)^4.
\end{aligned}
$$

$$
\begin{aligned}
& \Bigg(\frac{1}{n-1} \sum_{j \neq 2}^{n} K_{hj2} - E(K_{h31}|X_2) + E(K_{h31}|X_2) - f(X_2) \Bigg)^2 \qquad (3.5.2) \\
= & \Bigg(\frac{1}{n-1} \sum_{j=3}^{n} [K_{hj2} - E(K_{h31}|X_2)] + [E(K_{h31}|X_2) - f(X_2)] + K_{h12} \Bigg)^4 \\
\leq & \frac{2}{(n-1)^4} \Bigg((\sum_{j=3}^{n} [K_{hj2} - E(K_{h31}|X_2)]) + K_{h12} \Bigg)^2 + \Bigg(O(h^2) \Bigg)^2.
\end{aligned}
$$

Now we may expect the following, with the third and first moment of the expansion equals zero, V_{n131} is less than or equal to

$$\frac{64h^d}{(n-1)^8} E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 \Bigg(\Big(\sum_{i=3}^n [K_{hi1} - E(K_{h31}|X_1)] + K_{h12} \Big)^4 + (n-1)^4 (O(h^2))^4 \Bigg) \cdot \Bigg(\Big(\sum_{j=3}^n [K_{hj2} - E(K_{h32}|X_2)] + K_{h12})^4 + (n-1)^4 (O(h^2) \Big)^4 \Bigg)$$

$$\leq \frac{256h^d}{(n-1)^8} E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 \cdot \Bigg(\Big(\sum_{i=3}^n [K_{hi1} - E(K_{h31}|X_1)]^4 + \sum_{i=3}^n \sum_{k \neq i, k=3}^n [K_{hi1} - E(K_{h31}|X_1)]^2 [K_{hk1} - E(K_{h41}|X_1)]^2 + (K_{h12})^4 + (n-1)^4 (O(h^2) \Big)^4 \Bigg)$$

$$\cdot \Bigg(\Big(\sum_{j=3}^n [K_{hj2} - E(K_{h32}|X_2)]^4 + \sum_{j=3}^n \sum_{l \neq j, l=3}^n [K_{hj2} - E(K_{h32}|X_2)]^2 [K_{hk2} - E(K_{h42}|X_2)]^2 + (K_{h12})^4 + (n-1)^4 (O(h^2) \Big)^4 \Bigg)$$

$$= \frac{256h^d}{(n-1)^8} E\Bigg(K_{h12}^2 \eta_1^2 \eta_2^2 \Bigg((\sum_{i=3}^n [K_{hi1} - E(K_{h31}|X_1)]^4)(\sum_{j=3}^n [K_{hj2} - E(K_{h32}|X_2)]^4)$$

$$+2(\sum_{i=3}^n [K_{hi1} - E(K_{h31}|X_1)]^4)(\sum_{j=3}^n \sum_{l \neq j, l=3}^n [K_{hj2} - E(K_{h32}|X_2)]^2 [K_{hk2} - E(K_{h42}|X_2)]^2)$$

$$+(\sum_{i=3}^n \sum_{k \neq i, k=3}^n [K_{hi1} - E(K_{h31}|X_1)]^2 [K_{hk1} - E(K_{h41}|X_1)]^2) \cdot (\sum_{j=3}^n \sum_{l \neq j, l=3}^n [K_{hj2} - E(K_{h32}|X_2)]^2 [K_{hk2} - E(K_{h42}|X_2)]^2)$$

$$+2(\sum_{i=3}^n [K_{hi1} - E(K_{h31}|X_1)]^4)(K_{h12} + (n-1)^4 O(h^2))^4$$

$$+2(\sum_{i=3}^n \sum_{k \neq i, k=3}^n [K_{hi1} - E(K_{h31}|X_1)]^2 [K_{hk1} - E(K_{h41}|X_1)]^2)(K_{h12}^4 + (n-1)^4 (O(h^2))^4)$$

$$+(K_{h12}^4 + (n-1)^4 (O(h^2))^4)^2 \Bigg)$$

$$= O\Big(\frac{1}{n^6 h^{6d}}\Big) + O\Big(\frac{1}{n^5 h^{5d}}\Big) + O\Big(\frac{1}{n^4 h^{4d}}\Big) + O\Big(\frac{h^8}{n^3 h^{3d}}\Big) + O\Big(\frac{h^8}{n^2 h^{2d}}\Big) + O(h^{16}) = o(1).$$

The last equality follows from the derivation that is typical to all, we show one of it and the rest are the same. By Change of Variables, Taylor Expansion, and previously mentioned assumptions, we have,

$$\frac{h^d}{(n-1)^8}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2(\sum_{i=3}^{n}\sum_{k\neq i,k=3}^{n}[K_{hi1}-E(K_{h31}|X_1)]^2[K_{hk1}-E(K_{h41}|X_1)]^2)$$

$$\cdot(\sum_{j=3}^{n}\sum_{l\neq j,l=3}^{n}[K_{hj2}-E(K_{h32}|X_2)]^2[K_{hk2}-E(K_{h42}|X_2)]^2)$$

$$= \frac{(n-2)^2h^d}{(n-1)^6}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2([K_{h31}-E(K_{h31}|X_1)]^2[K_{h41}-E(K_{h41}|X_1)]^2)$$

$$\cdot([K_{h52}-E(K_{h52}|X_2)]^2[K_{h62}-E(K_{h62}|X_2)]^2)$$

$$= \frac{(n-2)^2h^d}{(n-1)^6}\int\frac{1}{h^{10d}}K^2\left(\frac{x_2-x_1}{h}\right)\tau^2(x_1)\tau^2(x_2)\left[K\left(\frac{x_3-x_1}{h}\right)-f(x_1)+O(h^2)\right]^2$$

$$\cdot\left[K\left(\frac{x_4-x_1}{h}\right)-f(x_1)+O(h^2)\right]^2\left[K\left(\frac{x_5-x_2}{h}\right)-f(x_2)+O(h^2)\right]^2$$

$$\cdot\left[K\left(\frac{x_6-x_2}{h}\right)-f(x_2)+O(h^2)\right]^2 f(x_1)f(x_2)f(x_3)f(x_4)f(x_5)f(x_6)dx_1...dx_6 = O\Big(\frac{1}{n^4h^{4d}}\Big).$$

Therefore we may follow the similar path with terms like V_{n132}, V_{n133}, it leads to that V_{n132} equals the following:

$$h^dE\left(K_{h12}^2\eta_1^2\eta_2^2f^2(X_1)((\hat{f}_h(X_1)-f(X_1))^2(\hat{f}_h(X_2)-f(X_2))^4\right)$$

$$\leq \frac{16h^d}{(n-1)^6}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2f^2(X_1)\left((\sum_{i=3}^{n}[K_{hi1}-E(K_{h31}|X_1)]+K_{h12})^2+(n-1)^2(O(h^2))^2\right)$$

$$\cdot\left((\sum_{j=3}^{n}[K_{hj2}-E(K_{h32}|X_2)]+K_{h12})^4+(n-1)^4(O(h^2))^4\right)$$

$$\leq \frac{128h^d}{(n-1)^6}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2f^2(X_1)\left((\sum_{i=3}^{n}[K_{hi1}-E(K_{h31}|X_1)]^2+(K_{h12})^2+(n-1)^2(O(h^2))^2\right)$$

$$\cdot\Bigg((\sum_{j=3}^{n}[K_{hj2}-E(K_{h32}|X_2)]^4+\sum_{j=3}^{n}\sum_{l\neq j,l=3}^{n}[K_{hj2}-E(K_{h32}|X_2)]^2[K_{hk2}-E(K_{h42}|X_2)]^2$$

$$+(K_{h12})^4+(n-1)^4\cdot(O(h^2))^4\Bigg)$$

$$= O\left(\frac{1}{n^4h^{4d}}\right)+O\left(\frac{1}{n^3h^{3d}}\right)+O\left(\frac{h^4}{n^2h^{2d}}\right)+O\left(\frac{h^8}{nh^d}\right)+O(h^{12})=o(1).$$

Now, by observation, through similar route of proof, we obtain the last equality. And,

$$\begin{aligned}
V_{133} &= h^dE\left(K_{h12}^2\eta_1^2\eta_2^2f^2(X_1)f^2(X_2)((\hat{f}_h(X_1)-f(X_1))^2(\hat{f}_h(X_2)-f(X_2))^2\right)\\
&\le \frac{4h^d}{(n-1)^4}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2f^2(X_1)\left((\sum_{i=3}^{n}[K_{hi1}-E(K_{h31}|X_1)]+K_{h12})^2+(n-1)^2(O(h^2))^2\right)\\
&\qquad\cdot\left((\sum_{j=3}^{n}[K_{hj2}-E(K_{h32}|X_2)]+K_{h12})^2+(n-1)^2(O(h^2))^2\right)\\
&\le \frac{16h^d}{(n-1)^4}E\Bigg(K_{h12}^2\eta_1^2\eta_2^2f^2(X_1)\\
&\qquad\cdot\left((\sum_{i=3}^{n}[K_{hi1}-E(K_{h31}|X_1)]^2+(K_{h12})^2+(n-1)^2(O(h^2))^2\right)\\
&\qquad\cdot\left((\sum_{j=3}^{n}[K_{hj2}-E(K_{h32}|X_2)]^2+(K_{h12})^2+(n-1)^2\cdot(O(h^2))^2\right)\\
&= O\left(\frac{1}{n^2h^{2d}}\right)+O\left(\frac{h^2}{nh^d}\right)+O(h^8)=o(1).
\end{aligned}$$

Next we will show that $nh^{d/2}V_{n11}$ converges weakly to normal distribution with mean zero and variance $\Sigma = 2\int K^2(u)du\int[\tau^2(x)]^2f^{10}(x)dx$. Following Theorem 3.2.2, calculate the following quantities, with $Z_i=(\eta_i,X_i^T)$,

$$\begin{aligned}
E[H_n(Z_1,Z_2)|Z_1] &= E(K_{h12}\eta_1\eta_2f^2(X_1)f^2(X_2)|(\eta_1,X_1^T))\\
&= (\eta_1f^2(X_1))E(K_{h12}\eta_2f^2(X_2)|(\eta_1,X_1^T))\\
&= (\eta_1f^2(X_1))E(K_{h12}E(\eta_2f^2(X_2)|X_2^T))\\
&= (\eta_1f^2(X_1))E(K_{h12}f^2(X_2)E(\eta_2|X_2^T))=0
\end{aligned}$$

And this shows that $H_n(z_1, z_2)$ is a degenerate U-statistic.

$$\begin{aligned}
& E[G_n^2(z_1, z_2)] = E(E^2(H_n(z_3, z_1)H_n(z_3, z_2)|z_1, z_2)) \\
= \; & E(E^2(K_{h31}\eta_3\eta_1 f^2(X_3)f^2(X_1)K_{h32}\eta_3\eta_2 f^2(X_3)f^2(X_2)|(\eta_1, X_1^T), (\eta_2, X_2^T))) \\
= \; & E(\eta_1^2\eta_2^2 f^4(X_1)f^4(X_2)E^2(K_{h31}K_{h32}\eta_3^2 f^4(X_3)|(X_1^T, X_2^T))) \\
= \; & E(\eta_1^2\eta_2^2 f^4(X_1)f^4(X_2)\left[\int \frac{1}{h^{2d}}K\left(\frac{x_3 - X_1}{h}\right)K\left(\frac{x_3 - X_2}{h}\right)\tau^2(x_3)f^5(x_3)dx_3\right]^2 |(X_1^T, X_2^T)) \\
= \; & E(\eta_1^2\eta_2^2 f^4(X_1)f^4(X_2)\frac{1}{h^{2d}}\left[\int K(u)K\left(u + \frac{X_1 - X_2}{h}\right)\tau^2(X_1 + uh)f^5(X_1 + uh)du\right]^2 |(X_1^T, X_2^T)) \\
= \; & \frac{1}{h^{2d}}\int \tau^2(x_1)\tau^2(x_2)f^5(x_1)f^5(x_2) \\
& \qquad \cdot\left[\int K(u)K\left(u + \frac{x_1 - x_2}{h}\right)\tau^2(x_1 + uh)f^5(x_1 + uh)du\right]^2 dx_1dx_2 \\
= \; & \frac{1}{h^d}\int \tau^2(x_1)\tau^2(x_1 - vh)f^5(x_1)f^5(x_1 - vh) \\
& \qquad \cdot\left[\int K(u)K(u + v)\tau^2(x_1 + uh)f^5(x_1 + uh)du\right]^2 dx_1dv = O\left(\frac{1}{h^d}\right).
\end{aligned}$$

Next let's find the order of $H_n^2(z_1, z_2)$,

$$\begin{aligned}
E(H_n^2(z_1, z_2)) &= E[EH_n^2(z_1, z_2)|X_1, X_2] \\
&= \int K_{h12}^2\tau^2(x_1)\tau^2(x_2)f^5(x_1)f^5(x_2)dx_1dx_2 \\
&= \int \frac{1}{h^d}K^2(u)\tau^2(x_2 + uh)\tau^2(x_2)f^5(x_2 + uh)f^5(x_2)dudx_2 \\
&= \int \frac{1}{h^d}K^2(u)[\tau^2(x_2) + uh(\tau^2(x_2))' + 1/2u^2h^2(\tau^2(x_2))'' + o(h^2)]\tau^2(x_2) \\
&\qquad \cdot[f^5(x_2) + 5uhf^4(x_2) + 10u^2h^2f^3(x_2) + o(h^2)]f^5(x_2)dudx_2 \\
&= O\left(\frac{1}{h^d}\right) + O\left(\frac{1}{h^{d-2}}\right) = O\left(\frac{1}{h^d}\right).
\end{aligned}$$

The last inequality is obtained by Taylor expansion, Dominate Convergence Theorem and the continuity of both the $f(x)$ and $\tau(x)$. Similar techniques and expansions will be used throughout the rest of the research, and also notice that since $\int uK(u)du = 0$, so is the term

$\int uK^2(u)du = 0$. Next, we have,

$$\begin{aligned} EH_n^4(z_1, z_2) &= \int K_h^4(x_2 - x_1)\tau^4(x_1)\tau^4(x_2)f^9(X_1)f^9(x_2)dx_1dx_2 \\ &= \frac{1}{h^{3d}} \int K^4(u)\tau^4(x_1)\tau^4(x_1 + uh)f^9(x_1)f^9(x_1 + uh)dx_1du = O\left(\frac{1}{h^{3d}}\right). \end{aligned}$$

Thus, from the condition in Theorem 3.2.2, we see that

$$\frac{O(h^{-d}) + O(n^{-1}h^{-3d})}{O(h^{-2d})} = O(h^d) + O\left(\frac{1}{nh^d}\right),$$

converges to zero is satisfied. Therefore, the variance

$$\Sigma = 2h^d EH_n^2(z_1, z_2) = 2\int K^2(u)du \int (\tau^2(x_2))^2 f^{10}(x_2)dx_2.$$

Next, we will show that $nh^{d/2}V_{n3}$ and $nh^{d/2}V_{n8}$ are $o_p(1)$. With V_{n3}, we may again have the similar decomposition,

$$\begin{aligned} \frac{1}{2}V_{n3} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{hij}\eta_i\Delta v_j \hat{f}_h^2(X_i)\hat{f}_h^2(X_j) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{hij}\eta_i\Delta v_j (\hat{f}_h^2(X_i) - f^2(X_i) + f^2(X_i))(\hat{f}_h^2(X_j) - f^2(X_j) + f^2(X_j)) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{hij}\eta_i\Delta v_j f^2(X_i)f^2(X_j) \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{hij}\eta_i\Delta v_j (\hat{f}_h^2(X_i) - f^2(X_i))f^2(X_j) \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{hij}\eta_i\Delta v_i (\hat{f}_h^2(X_j) - f^2(X_j))f^2(X_i) \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{hij}\eta_i\Delta v_j (\hat{f}_h^2(X_i) - f^2(X_i))(\hat{f}_h^2(X_j) - f^2(X_j)) \\ &:= V_{n31} + 2V_{n32} + V_{n33} + V_{n34} \end{aligned}$$

We modify Lemma 3.3b from Zheng (1996), the following result that will help in proving that $nh^{d/2}V_{n31} = o_p(1)$

$$\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i G(X_j) = O_p\left(\frac{1}{\sqrt{n}}\right) \tag{3.5.3}$$

The proof of the above result can be easily obtained by changing ε_i with η_i and $\sigma^2(X_i)$ with $\tau^2(X_i)$.Therefore, we may directly go to the following derivation.

First note that,

$$\begin{aligned}\Delta v_i &= v(X_i,\theta_0) - v(X_i,\hat{\theta}_n) \\ &= v(X_i,\theta_0) - \Big(v(X_i,\theta_0) + \frac{\partial v(X_i,\theta_0)}{\partial\theta}(\theta_0-\hat{\theta}_n) + (\theta_0-\hat{\theta}_n)^T\frac{\partial^2 v(X_i,\tilde{\theta})}{\partial\theta\partial\theta^T}(\theta_0-\hat{\theta}_n)\Big) \\ &= \frac{\partial v(X_i,\theta_0)}{\partial\theta}(\hat{\theta}_n-\theta_0) + (\hat{\theta}_n-\theta_0)^T\frac{\partial^2 v(X_i,\tilde{\theta})}{\partial\theta\partial\theta^T}(\hat{\theta}_n-\theta_0).\end{aligned} \tag{3.5.4}$$

Therefore, we plug back into all the $V'_{n3}s$, we get

$$\begin{aligned}V_{n31} &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i f^2(X_i)f^2(X_j)\frac{\partial v(X_i,\theta_0)}{\partial\theta}(\hat{\theta}_n-\theta_0) \\ &\qquad + \frac{2(\hat{\theta}_n-\theta_0)^T}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i f^2(X_i)f^2(X_j)\frac{\partial^2 v(X_i,\tilde{\theta})}{\partial\theta\partial\theta^T}(\hat{\theta}_n-\theta_0) \\ &:= V_{n311}(\hat{\theta}_n-\theta_0) + (\hat{\theta}_n-\theta_0)^T V_{n312}(\hat{\theta}_n-\theta_0).\end{aligned}$$

Following Zheng (1996), we know that with the equation (3.5.3), $V_{n311} = O_p(n^{-1/2})$, and for V_{n312}

$$\begin{aligned}E(\|V_{n312}\|) &\leq E\left[K_{h12}\tau(X_1)f^2(X_1)f^2(X_2)\left\|\frac{\partial^2 v(x_3,\tilde{\theta})}{\partial\theta\partial\theta^T}\right\|\right] \\ &= \int K_h(x_1-x_2)\tau(x_1)\left\|\frac{\partial^2 v(x_3,\tilde{\theta})}{\partial\theta\partial\theta^T}\right\| f^3(x_1)f^3(x_2)dx_1dx_2 = O(1)\end{aligned}$$

Therefore, by the $\sqrt{n}$ consistency of the $\hat{\theta}_n$, we can conclude that $V_{n31} = O(n^{-1})$, so is $nh^{d/2}V_{n31} = O(h^{d/2})$. Here, borrow the similar idea, we may directly derive the following from V_{n8}.

$$\begin{aligned}
V_{n8} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Delta v_i \Delta v_j (\hat{f}_h^2(X_i) - f^2(X_i) + f^2(X_i))(\hat{f}_h^2(X_j) - f^2(X_j) + f^2(X_j)) \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Delta v_i \Delta v_j f^2(X_i) f^2(X_j) \\
&\quad + \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Delta v_i \Delta v_j f^2(X_i)(\hat{f}_h^2(X_j) - f^2(X_j)) \\
&\quad + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Delta v_i \Delta v_j (\hat{f}_h^2(X_i) - f^2(X_i))(\hat{f}_h^2(X_j) - f^2(X_j)) \\
&:= V_{n81} + V_{n82} + V_{n83}.
\end{aligned}$$

And with similar ideas, we will try to utilize the $\sqrt{n}$ consistency of the parameter, we may obtain the following derivation for the product of $\Delta v_i's$,

$$\begin{aligned}
V_{n81} &= \frac{(\hat{\theta}_n - \theta_0)^T}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij} f^2(X_i) f^2(X_j) \frac{\partial v(X_i,\tilde{\theta})}{\partial \theta}\frac{\partial v(X_i,\tilde{\theta})}{\partial \theta^T}(\hat{\theta}_n - \theta_0) \\
&:= (\hat{\theta}_n - \theta_0)^T V_{n811}(\hat{\theta}_n - \theta_0).
\end{aligned}$$

Integrating V_{n811} with the previous assumption, it implies that the first order derivatives with respect to variance function are bounded, we know that this quantity is of order $O_p(1)$. and therefore $nh^{d/2}V_{n81} = O_p(h^{d/2})$.

For the rest of the proof, after observing similar patterns occurs, we will show that $nh^{d/2}V_{n32}$, $nh^{d/2}V_{n33}$, $nh^{d/2}V_{n34}$, $nh^{d/2}V_{n82}$, $nh^{d/2}V_{n83}$ are all $o_p(1)$. And for simplicity purposes, we will only show $nh^{d/2}V_{n33}$, $nh^{d/2}V_{n34}$ is of order $o_p(1)$. It is achieved by using Cauchy-Schwartz inequality, and $\eta_i = O_p(1), \forall i$. Following these we will also show that $nh^{d/2}V_{n82}$, $nh^{d/2}V_{n83}$ are of $o_p(1)$ as well.

Consider one half of the term V_{n33},

$$\frac{1}{2}V_{n33} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\eta_i \Delta v_j (\hat{f}_h^2(X_j) - f^2(X_j)) f^2(X_i).$$

$$
\begin{aligned}
\frac{1}{2}E(n^2h^dV_{n32}^2) &= E\Bigg(\frac{h^d}{(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_{hij}^2\eta_i^2\Delta v_j^2(\hat{f}_h^2(X_i)-f^2(X_i))^2f^4(X_j) \\
&\quad+\frac{h^d}{(n-1)^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{k\neq j\neq i}^{n}K_{hij}K_{hik}\eta_i^2\Delta v_j\Delta v_k(\hat{f}_h^2(X_i)-f^2(X_i))^2f^2(X_j)f^2(X_k)\Bigg) \\
&= \frac{nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\Delta v_2^2(\hat{f}_h^2(X_1)-f^2(X_1))^2f^4(X_2)\Bigg) \\
&\quad+\frac{n(n-2)h^d}{(n-1)}E\Bigg(K_{h12}K_{h13}\eta_1^2\Delta v_2\Delta v_3(\hat{f}_h^2(X_1)-f^2(X_1))^2f^2(X_2)f^2(X_3)\Bigg).
\end{aligned}
$$

Again, utilize the following relationship

$$
(\hat{f}_1^2-f_1^2)^2=((\hat{f}_1-f_1)^2+2f_1(\hat{f}_1-f_1))^2\leq 2(\hat{f}_1-f_1)^4+4f_1^2(\hat{f}_1-f_1)^2. \tag{3.5.5}
$$

$$
\begin{aligned}
&\frac{nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\Delta v_2^2(\hat{f}_h^2(X_1)-f^2(X_1))^2f^4(X_2)\Bigg) \\
\leq\ &\frac{2nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\Delta v_2^2(\hat{f}_h(X_1)-f(X_1))^4f^4(X_2)\Bigg) \\
&+\frac{4nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\Delta v_2^2(\hat{f}_h(X_1)-f(X_1))^2f^2(X_1)f^4(X_2)\Bigg) \\
=\ &\frac{2nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\Delta v_2^2\Big(\frac{1}{n-1}\sum_{i\neq 1}K_{hi1}-E(K_{h21}|X_1)+E(K_{h21}|X_1)-f(X_1)\Big)^4f^4(X_2)\Bigg) \\
&+\frac{4nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\Delta v_2^2\Big(\frac{1}{n-1}\sum_{i\neq 1}K_{hi1}-E(K_{h21}|X_1)+E(K_{h21}|X_1)-f(X_1)\Big)^2f^2(X_1)f^4(X_2)\Bigg) \\
\leq\ &\frac{16nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\Delta v_2^2\Big[\Big(\frac{1}{n-1}\sum_{i\neq 1}K_{hi1}-E(K_{h21}|X_1)\Big)^4+(E(K_{h21}|X_1)-f(X_1))^4\Big]f^4(X_2)\Bigg) \\
&+\frac{8nh^d}{(n-1)}E\Bigg(K_{h12}^2\eta_1^2\Delta v_2^2\Bigg[\Bigg(\frac{1}{n-1}\sum_{i\neq 1}K_{hi1}-E(K_{h21}|X_1)\Bigg)^2 \\
&\qquad+(E(K_{h21}|X_1)-f(X_1))^2\Bigg]f^2(X_1)f^4(X_2)\Bigg) \\
=\ &O_p\Bigg(\frac{1}{n^4h^{3d}}\Bigg)+O_p\Bigg(\frac{1}{n^3h^{2d}}\Bigg)+O_p\Bigg(\frac{1}{n^2h^d}\Bigg)+O\Bigg(\frac{h^{d+8}}{n}\Bigg)+O\Bigg(\frac{h^{d+4}}{n}\Bigg).
\end{aligned}
$$

The last equality is obtained by using the expansion (3.2.2), and the fact that $\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})$.

And we have similar result,

$$\begin{aligned} & \frac{n(n-2)h^d}{(n-1)} E\Bigg(K_{h12}K_{h13}\eta_1^2 \Delta v_2 \Delta v_3 (\hat{f}_h^2(X_1) - f^2(X_1))^2 f^2(X_2) f^2(X_3) \Bigg) \\ = \;& O_p\Bigg(\frac{1}{n^3h^{3d}}\Bigg) + O_p\Bigg(\frac{1}{n^2h^{2d}}\Bigg) + O_p\Bigg(\frac{1}{nh^d}\Bigg) + O\Bigg(h^{d+8}\Bigg) + O\Bigg(h^{d+4}\Bigg). \end{aligned}$$

Again, applying similar reasoning,

$$\begin{aligned} & E(n^2h^dV_{n34}^2) \\ = \;& \frac{h^d}{(n-1)^2} E\Bigg(\sum_{i=1}^n \sum_{j\neq i}^n K_{hij}\eta_i \Delta v_j (\hat{f}_h^2(X_i) - f^2(X_i))(\hat{f}_h^2(X_j) - f^2(X_j)) \Bigg)^2 \\ = \;& \frac{h^d}{(n-1)^2} E\Bigg(\sum_{i=1}^n \sum_{j\neq i}^n K_{hij}^2 \eta_i^2 \Delta v_j^2 (\hat{f}_h^2(X_i) - f^2(X_i))^2 (\hat{f}_h^2(X_j) - f^2(X_j))^2 \Bigg) \\ & + \frac{h^d}{(n-1)^2} E\Bigg(\sum_{i=1}^n \sum_{j\neq i}^n \sum_{k\neq j\neq i}^n K_{hij}K_{hik}\eta_i^2 \Delta v_j \Delta v_k (\hat{f}_h^2(X_i) - f^2(X_i))^2 \\ & \qquad \cdot (\hat{f}_h^2(X_j) - f^2(X_j))(\hat{f}_h^2(X_k) - f^2(X_k)) \Bigg) \\ \leq \;& \frac{4nh^d}{(n-1)} E\Bigg(K_{h12}^2 \eta_1^2 \Delta v_2^2 [(\hat{f}_h(X_1) - f(X_1))^4 + 4f^2(X_1)(\hat{f}_h(X_1) - f(X_1))^2] \\ & \qquad \cdot [(\hat{f}_h(X_2) - f(X_2))^4 + 4f^2(X_2)(\hat{f}_h(X_2) - f(X_2))^2] \Bigg) \\ & + \frac{4n(n-2)h^d}{(n-1)} E\Bigg(K_{h12}K_{h13}\eta_1^2 \Delta v_2 \Delta v_3 [(\hat{f}_h(X_1) - f(X_1))^4 + 4f^2(X_1)(\hat{f}_h(X_1) - f(X_1))^2] \\ & \cdot [(\hat{f}_h(X_2) - f(X_2))^2 + 2f(X_2)(\hat{f}_h(X_2) - f(X_2))] \\ & \cdot [(\hat{f}_h(X_3) - f(X_3))^2 + 2f(X_3)(\hat{f}_h(X_3) - f(X_3))] \Bigg). \end{aligned}$$

Continue with the derivation, it is not hard to observe that $E(n^2h^dV_{n34}^2)$ is less than the following quantity,

$$
\begin{aligned}
&\frac{4nh^d}{(n-1)}E\Big(K_{h12}^2\eta_1^2\Delta v_2^2[(\hat{f}_h(X_1)-f(X_1))^4(\hat{f}_h(X_2)-f(X_2))^4]\Big)\\
&+\frac{32nh^d}{(n-1)}E\Big(K_{h12}^2\eta_1^2\Delta v_2^2 f^2(X_2)[(\hat{f}_h(X_1)-f(X_1))^4(\hat{f}_h(X_2)-f(X_2))^2]\Big)\\
&+\frac{64nh^d}{(n-1)}E\Big(K_{h12}^2\eta_1^2\Delta v_2^2 f^2(X_1)f^2(X_2)[(\hat{f}_h(X_1)-f(X_1))^2(\hat{f}_h(X_2)-f(X_2))^2]\Big)\\
&+\frac{4n(n-2)h^d}{(n-1)}E\Big(K_{h12}K_{h13}\eta_1^2\Delta v_2\Delta v_3[(\hat{f}_h(X_1)-f(X_1))^4\\
&\qquad\qquad\cdot(\hat{f}_h(X_2)-f(X_2))^2(\hat{f}_h(X_3)-f(X_3))^2]\Big)\\
&+\frac{8n(n-2)h^d}{(n-1)}E\Big(K_{h12}K_{h13}\eta_1^2\Delta v_2\Delta v_3[(\hat{f}_h(X_1)-f(X_1))^4\\
&\qquad\qquad\cdot(\hat{f}_h(X_2)-f(X_2))^2 f(X_3)(\hat{f}_h(X_3)-f(X_3))]\Big)\\
&+\frac{16n(n-2)h^d}{(n-1)}E\Big(K_{h12}K_{h13}\eta_1^2\Delta v_2\Delta v_3[(\hat{f}_h(X_1)-f(X_1))^4\\
&\qquad\qquad\cdot f(X_2)(\hat{f}_h(X_2)-f(X_2))(\hat{f}_h(X_3)-f(X_3))^2]\Big)\\
&+\frac{16n(n-2)h^d}{(n-1)}E\Big(K_{h12}K_{h13}\eta_1^2\Delta v_2\Delta v_3[(\hat{f}_h(X_1)-f(X_1))^4\\
&\qquad\qquad\cdot f(X_2)(\hat{f}_h(X_2)-f(X_2))f(X_3)(\hat{f}_h(X_3)-f(X_3))]\Big)\\
&+\frac{16n(n-2)h^d}{(n-1)}E\Big(K_{h12}K_{h13}\eta_1^2\Delta v_2\Delta v_3[f^2(X_1)(\hat{f}_h(X_1)-f(X_1))^2\\
&\qquad\qquad\cdot(\hat{f}_h(X_2)-f(X_2))^2(\hat{f}_h(X_3)-f(X_3))^2]\Big)\\
&+\frac{32n(n-2)h^d}{(n-1)}E\Big(K_{h12}K_{h13}\eta_1^2\Delta v_2\Delta v_3[f^2(X_1)(\hat{f}_h(X_1)-f(X_1))^2\\
&\qquad\qquad\cdot(\hat{f}_h(X_2)-f(X_2))^2 f(X_3)(\hat{f}_h(X_3)-f(X_3))]\Big)\\
&+\frac{32n(n-2)h^d}{(n-1)}E\Big(K_{h12}K_{h13}\eta_1^2\Delta v_2\Delta v_3[f^2(X_1)(\hat{f}_h(X_1)-f(X_1))^2\\
&\qquad\qquad\cdot f(X_2)(\hat{f}_h(X_2)-f(X_2))(\hat{f}_h(X_3)-f(X_3))^2]\Big)\\
&+\frac{64n(n-2)h^d}{(n-1)}E\Big(K_{h12}K_{h13}\eta_1^2\Delta v_2\Delta v_3[f^2(X_1)(\hat{f}_h(X_1)-f(X_1))^2\\
&\qquad\qquad\cdot f(X_2)(\hat{f}_h(X_2)-f(X_2))f(X_3)(\hat{f}_h(X_3)-f(X_3))]\Big).
\end{aligned}
$$

From previous analysis, with derivation such as (3.5.1) and (3.5.2) we may know that

$\hat{f}_i \to f_i$, and the term with lowest moments of $(\hat{f}_i - f_i)$ will converge to zero with the slowest speed. Therefore, we focus on the last term. By (3.5.4), we have both $\Delta v_2 \Delta v_3 = O_p(n^{-1})$. Using some abbreviations,

$$
\begin{aligned}
& \frac{n(n-2)h^d}{(n-1)} E\left(\| K_{h12} K_{h13} \eta_1^2 \Delta v_2 \Delta v_3 [f_1^2 (\hat{f}_1 - f_1)^2 f_2 (\hat{f}_2 - f_2) f_3 (\hat{f}_3 - f_3)] \| \right) \\
\leq \; & \frac{(n-2)h^d}{(n-1)} E\left(\| K_{h12} K_{h13} \eta_1^2 f_1^2 f_2 f_3 [(\hat{f}_1 - f_1)^2 (\hat{f}_2 - f_2)(\hat{f}_3 - f_3)] \| \right) O_p(1) \\
= \; & \frac{(n-2)h^d}{(n-1)} E\left(\Big\| K_{h12} K_{h13} \eta_1^2 f_1^2 f_2 f_3 \Big[\Big(\frac{1}{n-1} \sum_{i=2}^{n} K_{hi1} - E(K_{hi1}|X_1) + E(K_{hi1}|X_1) - f_1 \Big)^2 \right. \\
& \cdot \Big(\frac{1}{n-1} \sum_{j \neq 2}^{n} K_{hj2} - E(K_{hj2}|X_2) + E(K_{hj2}|X_2) - f_2 \Big) \\
& \left. \qquad \cdot \Big(\frac{1}{n-1} \sum_{k \neq 3}^{n} K_{hk3} - E(K_{hk3}|X_3) + E(K_{hk1}|X_3) - f_3 \Big) \Big] \Big\| \right) O_p(1).
\end{aligned}
$$

By observation, the speed of convergence is much clear since all the term that contains centered version $(n-1)^{-1} \sum_{j \neq 2}^{n} K_{hj2} - E(K_{hj2}|X_2)$ or $(n-1)^{-1} \sum_{k \neq 3}^{n} K_{hk3} - E(K_{hk3}|X_3)$, which will be zero when taking expectation, so the only term that is not zero would be

$$
\frac{(n-2)h^d}{(n-1)} E\left(\Big\| K_{h12} K_{h13} \eta_1^2 f_1^2 f_2 f_3 \Big[\Big(\frac{1}{n-1} \sum_{i=2}^{n} K_{hi1} - E(K_{hi1}|X_1) + E(K_{hi1}|X_1) - f_1 \Big)^2 \Big] \Big\| \right) O_p(h^4)
$$

And it is not hard to see that the above term is of order $O_p((nh^{d-4})^{-1}) + O_p(h^{d+8})$. Other terms in V_{n341} can be proved in the similar method to be of order $o_p(1)$.

Comparing V_{n82} and V_{n83}, it is easy to see that $nh^{d/2} V_{n83} = o_p(1)$ once we have $nh^{d/2} V_{n82} = o_p(1)$. By C-S inequality and (3.5.4), since,

$$
\begin{aligned}
\frac{1}{2} V_{n82} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} K_{hij} \Delta v_i \Delta v_j f^2(X_i) (\hat{f}_h^2(X_j) - f^2(X_j)) \\
\frac{1}{4} \| E(n^2 h^d V_{n82}^2) \| &\leq \frac{h^d}{(n-1)^2} E\left(\sum_{i=1}^{n} \sum_{j \neq i}^{n} K_{hij} \| \Delta v_i \Delta v_j \| f^2(X_i) (\hat{f}_h^2(X_j) - f^2(X_j)) \right)^2
\end{aligned}
$$

$$
\begin{aligned}
&\leq \quad \frac{h^d}{n(n-1)^2} E\left(\sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{hij} f^2(X_i)(\hat{f}_h^2(X_j) - f^2(X_j)) \right)^2 O_p(1) \\
&= \quad \frac{h^d}{n(n-1)^2} E\left(\sum_{i=1}^{n} \sum_{j\neq i}^{n} K_h^2(X_i - X_j) f^4(X_i)(\hat{f}_h^2(X_j) - f^2(X_j))^2 \right. \\
&\quad + \sum_{i=1}^{n} \sum_{j\neq i}^{n} \sum_{k\neq j\neq i}^{n} K_{hij} K_h(X_i - X_k) f^4(X_i)(\hat{f}_h^2(X_j) - f^2(X_j))(\hat{f}_h^2(X_k) - f^2(X_k)) \\
&\quad + \sum_{i=1}^{n} \sum_{j\neq i}^{n} \sum_{k\neq j\neq i}^{n} K_{hij} K_h(X_j - X_k) f^2(X_i) f^2(X_k)(\hat{f}_h^2(X_j) - f^2(X_j))^2 \\
&\quad \left. + \sum_{i=1}^{n} \sum_{j\neq i}^{n} \sum_{k\neq j\neq i}^{n} \sum_{l\neq k\neq j\neq i}^{n} K_{hij} K_{hlk} f^2(X_i) f^2(X_k)(\hat{f}_h^2(X_j) - f^2(X_j))(\hat{f}_h^2(X_l) - f^2(X_l)) \right) O_p(1) \\
&:= \quad V_{n821} + V_{n822} + V_{n823} + V_{n824}.
\end{aligned}
$$

It is easy to see that the V_{n824} with most complexity has the biggest order while other terms would be $o_p(1)$ with the following derivation,

$$
\begin{aligned}
V_{824} &= \frac{h^d}{n^2(n-1)^2} E\left(\sum_{i=1}^{n} \sum_{j\neq i}^{n} \sum_{k\neq j\neq i}^{n} \sum_{l\neq k\neq j\neq i}^{n} K_{hij} K_{hlk} f^2(X_i) f^2(X_k)(\hat{f}_h^2(X_j) - f^2(X_j)) \right. \\
&\qquad \left. \cdot(\hat{f}_h^2(X_l) - f^2(X_l)) \right) O_p(1) \\
&= \frac{h^d(n^2-5n+6)}{n(n-1)} E\left(K_{h12} K_{h34} f_1^2 f_3^2 ((\hat{f}_2 - f_2)^2 + 2f_2(\hat{f}_2 - f_2)) \right. \\
&\qquad \left. \cdot((\hat{f}_4 - f_4)^2 + 2f_4(\hat{f}_4 - f_4)) \right) O_p(1) \\
&= \frac{h^d(n^2-5n+6)}{n(n-1)} E\left(K_{h12} K_{h34} f_1^2 f_3^2 (\hat{f}_2 - f_2)^2 (\hat{f}_4 - f_4)^2 \right) O_p(1) \\
&\quad + \frac{2h^d(n^2-5n+6)}{n(n-1)} E\left(K_{h12} K_{h34} f_1^2 f_3^2 f_4 ((\hat{f}_2 - f_2)^2 (\hat{f}_4 - f_4)) \right) O_p(1) \\
&\quad + \frac{2h^d(n^2-5n+6)}{n(n-1)} E\left(K_{h12} K_{h34} f_1^2 f_3^2 f_2 ((\hat{f}_4 - f_4)^2 (\hat{f}_2 - f_2)) \right) O_p(1) \\
&\quad + \frac{4h^d(n^2-5n+6)}{n(n-1)} E\left(K_{h12} K_{h34} f_1^2 f_3^2 f_2 f_4 (\hat{f}_2 - f_2)(\hat{f}_4 - f_4) \right) O_p(1)
\end{aligned}
$$

$$
\frac{4h^d(n^2-5n+6)}{n(n-1)} E\Bigg(K_{h12}K_{h34} f_1^2 f_3^2 f_2 f_4 (\hat{f}_2 - f_2)(\hat{f}_4 - f_4)\Bigg) O_p(1)
$$

$$
= \frac{4h^d(n^2-5n+6)}{n(n-1)} E\Bigg(K_{h12}K_{h34} f_1^2 f_3^2 f_2 f_4 \Bigg(\frac{1}{n-1}\sum_{i\neq 2}^{n} K_{hi2} - E(K_{h32}|X_2) + E(K_{h32}|X_2) - f_2\Bigg)
$$

$$
\cdot \Bigg(\frac{1}{n-1}\sum_{j\neq 2}^{n} K_{hj4} - E(K_{h54}|X_4) + E(K_{h54}|X_4) - f_4\Bigg)\Bigg) O_p(1).
$$

Without showing the tedious expansion and integration, change of variable, the order is $O(1/n) + O(h^{d+4})$. For other terms in V_{n824}, it is not hard to observe the difference lies in the order of $\hat{f}_i - f_i$, $i = 2, 4$, so we may conclude that $V_{n824} = o_p(1)$, and so is V_{n821}, V_{n822} and V_{n823}. It is also easy to see that $V_{n83} = o_p(V_{n82})$, and we have finished the proof of the lemma. □

Proof of Lemma 2. For the proof of the following two lemmas, we will see many terms involving terms like $(\hat{f}_i - f_i)^r$, $r = 1, 2, 4$. And the detailed proof of these terms are omitted due to repetitiveness. In fact, all of these terms are $o_p(1)$ since $\hat{f}_i \to f_i$. Now, Let's focus on V_{n5}, this is one of the terms that introduces the bias while substituting the unknown regression function with NW estimator, so are the other terms we will prove in this lemma.

$$
V_{n5} = \frac{1}{n(n-1)} \sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij} \Delta m_i^2 \Delta m_j^2 \hat{f}_h^2(X_i)\hat{f}_h^2(X_j).
$$

$$
E(nh^{d/2}V_{n5}) = E\Bigg(\frac{h^{d/2}}{n-1}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K_{hij}\Big(\frac{1}{n-1}\sum_{k\neq i}^{n} K_{hki}(Y_k - m_i)\Big)^2
$$

$$
\cdot\Big(\frac{1}{n-1}\sum_{l\neq j}^{n} K_{hlj}(Y_l - m_j)\Big)^2 \Bigg)
$$

$$
= \frac{nh^{d/2}}{(n-1)^4} E\Bigg(K_{h12}\Big(\sum_{k\neq 1}^{n} K_{hk1}(Y_k - m_1)\Big)^2\Big(\sum_{l\neq 2}^{n} K_{hl2}(Y_l - m_2)\Big)^2\Bigg)
$$

$$
= \frac{nh^{d/2}}{(n-1)^4} E\Bigg(K_{h12}\Big(\sum_{k\neq 1}^{n} K_{hk1}(m_k - m_1 + \varepsilon_k\sqrt{v(X_k,\theta_0)})\Big)^2
$$

$$
\cdot\Big(\sum_{l\neq 2}^{n} K_{hl2}(m_l - m_2 + \varepsilon_l\sqrt{v(X_l,\theta_0)})\Big)^2\Bigg)
$$

$$
\begin{aligned}
&\leq \frac{4nh^{d/2}}{(n-1)^4} EK_{h12}\left(\sum_{k\neq 1}^{n} K_{hk1}(m_k - m_1)\right)^2 \left(\sum_{l\neq 2}^{n} K_{hl2}(m_l - m_2)\right)^2 \\
&\quad + \frac{4nh^{d/2}}{(n-1)^4} EK_{h12}\left(\sum_{k\neq 1}^{n} K_{hk1}(m_k - m_1)\right)^2 \left(\sum_{l\neq 2}^{n} K_{hl2}\varepsilon_l \sqrt{v(X_l,\theta_0)}\right)^2 \\
&\quad + \frac{4nh^{d/2}}{(n-1)^4} EK_{h12}\left(\sum_{l\neq 2}^{n} K_{hl2}(m_l - m_2)\right)^2 \left(\sum_{k\neq 1}^{n} K_{hk1}\varepsilon_k \sqrt{v(X_k,\theta_0)}\right)^2 \\
&\quad + \frac{4nh^{d/2}}{(n-1)^4} \left(\sum_{k\neq 1}^{n} K_{hk1}\varepsilon_k \sqrt{v(X_k,\theta_0)}\right)^2 \left(\sum_{l\neq 2}^{n} K_{hl2}\varepsilon_l \sqrt{v(X_l,\theta_0)}\right)^2 \\
&:= 4(V_{n51} + V_{n52} + V_{n53} + V_{n54}).
\end{aligned}
$$

$$
\begin{aligned}
EV_{n51} &= \frac{nh^{d/2}}{(n-1)^4} EK_{h12}\left(\sum_{k=3}^{n} K_{hk1}(m_k - m_1) + K_{h21}(m_2 - m_1)\right)^2 \\
&\qquad \left((\sum_{l=3}^{n} K_{hl2}(m_l - m_2) + K_{h21}(m_1 - m_2)\right)^2 \\
&= \frac{nh^{d/2}}{(n-1)^4} EK_{h12}\Bigg[\left(\sum_{k=3}^{n}\sum_{l=3}^{n} K_{hk1}(m_k - m_1)K_{hl2}(m_l - m_2)\right)^2 \\
&\quad + \left(K_{h21}(m_2 - m_1)\sum_{l=3}^{n}(K_{hl2}(m_l - m_2)\right)^2 \\
&\quad + \left(K_{h21}(m_2 - m_1)\sum_{k=3}^{n}(K_{hk1}(m_k - m_1)\right)^2 + K_{h12}^4(m_1 - m_2)^4\Bigg] \\
&:= V_{n511} + V_{n512} + V_{n513} + V_{n514}.
\end{aligned}
$$

Before entering the following derivation, note that we think of two pairs of index (k, l) and (s, t), where may equal to each other within pair, pairwise equal, pairwise unequal with one index, and pairwise unequal with all terms, V_{n511} can be expanded as the following:

$$
\begin{aligned}
&\frac{nh^{d/2}}{(n-1)^4} EK_{h12}\left(\sum_{k=3}^{n}\sum_{l=3}^{n} K_{hk1}(m_k - m_1)K_{hl2}(m_l - m_2)\right)^2 \\
&= \frac{nh^{d/2}}{(n-1)^4} EK_{h12}\left(\sum_{k=3}^{n} K_{hk1}^2 K_{hk2}^2 (m_k - m_1)^2 (m_k - m_2)^2\right.
\end{aligned}
$$

$$+\sum_{k=3}^{n}\sum_{k\neq l=3}^{n} K_{hk1}^2(m_k-m_1)^2K_{hl2}^2(m_l-m_2)^2$$

$$+\sum_{k=l,s=t,k\neq s}^{n} K_{hk1}K_{hk2}K_{hs1}K_{hs2}(m_k-m_1)(m_k-m_2)(m_s-m_1)(m_s-m_2)$$

$$+\sum_{k=l,s\neq t,l=s}^{n} K_{hk1}^2K_{hk2}K_{ht2}(m_k-m_1)^2(m_k-m_2)(m_t-m_2)$$

$$+\sum_{k\neq l,s\neq t,l=s}^{n} K_{hk1}^2K_{hk2}K_{ht2}(m_k-m_1)^2(m_k-m_2)(m_t-m_2)$$

$$\left.+\sum_{k\neq l,s\neq t,l\neq s,t\neq l}^{n} K_{hk1}K_{hl2}K_{hs1}K_{ht2}(m_k-m_1)(m_k-m_2)(m_s-m_1)(m_t-m_2)\right).$$

The above looks a little bit mess, and each summation are of order $O(n)$, $O(n(n-1))$, $O(n(n-1))$, $O(2n(n-1)(n-2))$, $O(2n(n-1)(n-2))$,$O(n(n-1)(n-2)(n-3))$ respectively, the next step would be to show that the term follows will be of order $o_(1)$.

Observe that each one of the K_{hij} will have at least $\frac{1}{h^d}$, therefore $\frac{1}{h^{5d}}$. And since we are taking expectations, we also count how many variables we need to integrate, this will result in only h^d term left after integration. As for the bias part with m_1-m_2 say, we can enjoy the benefit of Taylor expansion to the second order, therefore bias is of order $O(h^2)$. Next we expand the last one with most complexity to show the calculation. We use abbreviated expression, without loss of generality, let $k=3$, $l=4$, $s=5$, $t=6$

$$\begin{aligned}
&\frac{nh^{d/2}}{(n-1)^4}EK_{h12}\left(\sum_{k\neq l,s\neq t,l\neq s,t\neq l}^{n} K_{hk1}K_{hl2}K_{hs1}K_{ht2}(m_k-m_1)(m_k-m_2)(m_s-m_1)(m_t-m_2)\right)\\
=\;&\frac{n^2(n-2)(n-3)h^{d/2}}{(n-1)^3}\int\frac{1}{h^{5d}}K\left(\frac{x_2-x_1}{h}\right)K\left(\frac{x_3-x_1}{h}\right)K\left(\frac{x_4-x_2}{h}\right)K\left(\frac{x_5-x_1}{h}\right)\\
&\cdot K\left(\frac{x_6-x_2}{h}\right)(m(x_3)-m_1)(m(x_4)-m_2)(m(x_5)-m_1)(m(x_6)-m_2)\\
&\qquad\cdot f(x_1)f(x_2)f(x_3)f(x_4)f(x_5)f(x_6)dx_1dx_2dx_3dx_4dx_5dx_6.\\
=\;&\frac{n^2(n-2)(n-3)h^{d/2}}{(n-1)^3}\int K(a)K(u)K(v)K(s)K(t)(m(x_1+uh)-m_1)\\
&\cdot(m(x_1+(a+v)h)-m(x_1+ah))(m(x_1+sh)-m_1)(m(x_1+(a+t)h)-m(x_1+ah))
\end{aligned}$$

$$
\begin{aligned}
&\cdot f(x_1)f(x_1+ah)f(x_1+uh)f(x_1+(a+v)h)f(x_1+sh)f(x_1+(a+t)h)dadudvdsdtdx_1\\
= \quad & \frac{n^2(n-2)(n-3)h^{d/2}}{(n-1)^3}\int K(a)K(u)K(v)K(s)K(t)(uhm'(x_1)+\frac{h^2}{2}u\frac{\partial^2 m(\tilde{x}_1)}{\partial x\partial x^T}u^T)\\
&\cdot(vhm'(x_1)+\frac{h^2}{2}(v^2+2av)\frac{\partial^2 m(\tilde{x}_1)}{\partial x\partial x^T}(v^2+2av)^T)(shm'(x_1)+\frac{h^2}{2}s\frac{\partial^2 m(\tilde{x}_1)}{\partial x\partial x^T}s^T)\\
&\cdot(thm'(x_1)+\frac{h^2}{2}(t^2+2tv)\frac{\partial^2 m(\tilde{x}_1)}{\partial x\partial x^T}(t^2+2tv)^T)f(x_1)(f(x_1)+ahf'(x_1)+\frac{h^2}{2}a\frac{\partial^2 f(\tilde{x}_1))}{\partial x\partial x^T}a^T)\\
&\cdot(f(x_1)+uhf'(x_1)+\frac{h^2}{h}u\frac{\partial^2 f(\tilde{x}_1)}{\partial x\partial x^T}u^T)(f(x_1)+shf'(x_1)+\frac{h^2}{2}s\frac{\partial^2 f(\tilde{x}_1)}{\partial x\partial x^T}s^T)\\
&\cdot(f(x_1)+(a+v)hf'(x_1)+\frac{h^2}{2}(a+v)\frac{\partial^2 f(\tilde{x}_1)}{\partial x\partial x^T}(a+v)^T))\\
&\cdot(f(x_1)+(a+t)hf'(x_1)+\frac{h^2}{2}(a+t)\frac{\partial^2 f(\tilde{x}_1)}{\partial x\partial x^T}(a+t)^T))dadudvdsdtdx_1=O(nh^{d/2+8}).
\end{aligned}
$$

The change of variable are $\frac{x_2-x_1}{h}=a, \frac{x_3-x_1}{h}=u$, $\frac{x_4-x_2}{h}=v$, $\frac{x_5-x_1}{h}=s$, $\frac{x_6-x_2}{h}=t$, and fix x_1, $x_2=x_1+ah$, $x_3=x_1+uh$, $x_4=x_2+vh=x_1+(a+v)h$, $x_5=x_1+sh$, $x_6=x_2+th=x_1+(a+t)h$. $\tilde{x}_1$ is inside the neighbor of x_1 with radius h. And last equality hold by the conditions K(1) and K(2).

The other terms all share the same order of bias, with orders $O(h^{8-d/2})$, $O(h^8)$ and $O((nh^{3d/2-8})^{-1})$. So, next we will focus on V_{n512}, Note that it has exactly the same order as V_{n513}, and V_{n514} is of order $O(h^{d/2+8}/n^3h^{3d})$ which converges to zero the fastest. Next show that V_{n512} is of order $o(1)$ which is simpler than V_{n511}.

$$
\begin{aligned}
& \frac{nh^{d/2}}{(n-1)^4}EK_{h12}\left(K_{h21}(m_2-m_1)\sum_{k=3}^{n}(K_{hk1}(m_k-m_1)\right)^2\\
= \quad & \frac{nh^{d/2}}{(n-1)^4}EK_{h12}^3(m_2-m_1)^2\left(\sum_{k=3}^{n}K_{hk1}^2(m_k-m_1)^2+\sum_{k=3}^{n}\sum_{l\neq k}^{n}K_{hk1}K_{hl1}(m_k-m_1)(m_l-m_1)\right)\\
= \quad & \frac{n(n-2)h^{d/2}}{(n-1)^4}EK_{h12}^3(m_2-m_1)^2K_{h31}^2(m_3-m_1)^2\\
&+\frac{n(n-3)(n-4)h^{d/2}}{(n-1)^4}EK_{h12}^3(m_2-m_1)^2K_{h31}K_{h41}(m_3-m_1)(m_4-m_1)\\
= \quad & O\left(\frac{1}{n^2h^{5/2d-8}}\right)+O\left(\frac{1}{nh^{3/2d-8}}\right).
\end{aligned}
$$

Here, since $nh^{d/2+8}\to 0$, $nh^{3/2d-8}\to\infty$,$h^{8-d/2}$, and $nh^d\to 0$, implies that $d/2+8>d$, and

$3/2d-8<d$, implies $d<16$, which satisfies our previous restriction. For V_{n514}, by the usual change of variables, we will obtain the order of $O_p((n^3h^{7d/2-8})^{-1})$.

Now let's focus on V_{n52}, so similar will be carried out with V_{n53},

$$
\begin{aligned}
V_{n52} &= \frac{4nh^{d/2}}{(n-1)^4} EK_{h12}\left(\sum_{k\neq 1}^{n} K_{hk1}(m_k-m_1)\right)^2 \left(\sum_{l\neq 2}^{n} K_{hl2}\varepsilon_l\sqrt{v(X_l,\theta_0)}\right)^2 \\
&= \frac{4nh^{d/2}}{(n-1)^4} EK_{h12}\left(\sum_{k\neq 1}^{n} K_{hk1}^2(m_k-m_1)^2 + \sum_{k\neq 1}^{n}\sum_{i\neq k\neq 1}^{n} K_{hk1}K_{hi1}(m_k-m_1)\right. \\
&\left.\cdot(m_i-m_1)\right)\left(\sum_{l\neq 2}^{n} K_{hl2}^2\varepsilon_l^2 v(X_l,\theta_0) + \sum_{l\neq 1}^{n}\sum_{j\neq l\neq 1}^{n} K_{hl2}K_{hj2}\varepsilon_l\varepsilon_j\sqrt{v(X_l,\theta_0)v(X_j,\theta_0)}\right) \\
&= \frac{4nh^{d/2}}{(n-1)^4} EK_{h12}\left(\sum_{k\neq 1}^{n}\sum_{l\neq 2}^{n} K_{hl2}^2K_{hk1}^2(m_k-m_1)^2\varepsilon_l^2 v(X_l,\theta_0)\right. \\
&\left. + \sum_{k\neq 1}^{n}\sum_{i\neq k\neq 1}^{n}\sum_{l\neq 2}^{n} K_{hl2}^2K_{hk1}K_{hi1}\varepsilon_l^2 v(X_l,\theta_0)(m_k-m_1)(m_i-m_1)\right) \\
&= O\left(\frac{1}{nh^{3d/2-4}}\right) + O(h^{4-d/2})
\end{aligned}
$$

With restriction of $4-d/2>0$, implies $d<8$.

Similarly, we know that,

$$
\begin{aligned}
V_{n51} &= \frac{4nh^{d/2}}{(n-1)^4} EK_{h12}\left(\sum_{k\neq 1}^{n} K_{hk1}\varepsilon_k\sqrt{v(X_k,\theta_0)}\right)^2 \left(\sum_{l\neq 2}^{n} K_{hl2}\varepsilon_l\sqrt{v(X_l,\theta_0)}\right)^2 \\
&= O(1/(nh^{3d/2})),
\end{aligned}
$$

with the previous restriction $d/2<4$, this is equivalent to the order that nh^{d+4}, and this is allowed to be infinity. We have proven that $nh^{d/2}V_{n5}=o_p(1)$.

Remark. *There are indeed numerous methods existing that cope with the "curse of dimensionality" in nonparametric regression function tests. And one of the method considered in sufficient dimension reduction is the projection pursuit that project the regression space onto a lower dimension space. In variance checking however, we may notice that this is no longer a simple lack of power problem, but a restriction in applicability of the test. If we observe* *carefully in V_{n52}, the restriction mainly comes from the term $h^{-d/2}$, and this comes from the Nadaraya-Watson estimator we proposed for the regression function $m(x)$, $x \in R^d$. Therefore, in checking variance function, we may need to consider the projection of the regression space, this will be verified in later studies.*

Next show $nh^{d/2}V_{n2} = o_p(1)$. Similar to the previous methods that deals with $\hat{f}_i \to f_i$, we will omit the proof. Using Chebyshev inequality, C-R inequality, and the fact that $E(\eta_i|X_i) = 0$,

$$
\begin{aligned}
& E(n^2h^dV_{n2}^2) = \frac{h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_{hij}\eta_i\Delta m_j^2\hat{f}_i^2\hat{f}_j^2\Bigg)^2 \\
= \quad & \frac{h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_{hij}^2\eta_i^2\Big(\frac{1}{n-1}\sum_{k\neq j}^{n}K_{hkj}(Y_k-m_j)\Big)^4 f_i^4\Bigg) + o(1) \\
= \quad & \frac{h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_{hij}^2\eta_i^2\Big(\frac{1}{n-1}\sum_{k\neq j}^{n}K_{hkj}(m_k-m_j+\varepsilon_k\sqrt{v(X_k,\theta_0)})\Big)^4 f_i^4\Bigg) + o(1) \\
\leq \quad & \frac{8h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_{hij}^2\eta_i^2\Big(\frac{1}{n-1}\sum_{k\neq j}^{n}K_{hkj}(m_k-m_j)\Big)^4 f_i^4\Bigg) \\
& \qquad + \frac{8h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_{hij}^2\eta_i^2\Big(\frac{1}{n-1}\sum_{k\neq j}^{n}K_{hkj}(\varepsilon_k\sqrt{v(X_k,\theta_0)})\Big)^4 f_i^4\Bigg) + o(1) \\
:= \quad & V_{n21} + V_{n22}.
\end{aligned}
$$

By direct observation, it is not difficult to see that similar technique would be carried out as in the proof of V_{n5}, for the sake of brevity, using expansion (3.2.2), we have that $V_{n21} = O(h^8)$, and $V_{n22} = O((n^2h^{2d})^{-1})$. $nh^{d/2}V_{n6}$ would follow with much easier expansion. From the proof of V_{n3} and the above, it suffices to show that $E(\|0.25n^2h^dV_{n6}^2\|)$ equals the following,,

$$
\begin{aligned}
& \frac{h^d}{(n-1)^2}E\Bigg(\|\sum_{i=1}^{n}\sum_{j\neq i}^{n}K_{hij}\Delta v_i\Delta m_j^2\hat{f}_i^2\hat{f}_j^2\|\Bigg)^2 \\
= & \frac{h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^{n}\sum_{i\neq j}^{n}K_{hij}\Big(\frac{1}{n-1}\sum_{k\neq j}^{n}K_{hk2}(Y_k-m_j)\Big)^2 f_i^2\Bigg)^2 O\Big(\frac{1}{n}\Big) + o(1)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{h^d}{n} E\Bigg(K_{h12}^2 \Big(\frac{1}{n-1} \sum_{k\neq 2}^{n} K_{hk2}(Y_k - m_2) \Big)^4 f_1^4 \Bigg)^2 O(1) + o(1) \\
&\quad + h^d E\Bigg(K_{h12} K_{h13} \Big(\frac{1}{n-1} \sum_{k\neq 2}^{n} K_{hk2}(Y_k - m_2) \Big)^2 \Big(\frac{1}{n-1} \sum_{k\neq 3}^{n} K_{hk3}(Y_k - m_3) \Big)^2 f_1^4 \Bigg)^2 O(1) \\
&\quad + nh^d E\Bigg(K_{h12} K_{h34} \Big(\frac{1}{n-1} \sum_{k\neq 2}^{n} K_{hk2}(Y_k - m_2) \Big)^2 \Big(\frac{1}{n-1} \sum_{k\neq 4}^{n} K_{hk4}(Y_k - m_4) \Big)^2 f_1^2 f_3^2 \Bigg)^2 O(1) \\
&= O\Big(\frac{1}{nh^d} \Big) + O\Big(h^4 \Big).
\end{aligned}
$$

Thus we have completed the proof of the lemma.

□

Proof of Lemma 3. First we prove the one that is the most complicated, then the rest would follow with similar technique. The technique for all the term that shared in this lemma involves the use of the assumption $E(\varepsilon_i|X_i) = 0$. To show $nh^{d/2}V_{n10} = o_p(1)$, we show instead $E(n^2 h^d V_{n10}^2) = o(1)$. The above quantity equals,

$$
\begin{aligned}
& \frac{h^d}{(n-1)^2} E\Bigg(\sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{hij} \Delta m_i \varepsilon_i \sqrt{v(X_i,\theta_0)} \Delta m_j \varepsilon_j \sqrt{v(X_j,\theta_0)} \hat{f}_i^2 \hat{f}_j^2 \Bigg)^2 \\
= \; & \frac{h^d}{(n-1)^2} E\Bigg(\sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{hij} \Bigg(\frac{1}{n-1} \sum_{k\neq i}^{n} K_{hki} (m_k - m_i + \varepsilon_k \sqrt{v(X_k,\theta_0)}) \Bigg) \varepsilon_i \hat{f}_i^2 \\
& \qquad \cdot \Bigg(\frac{1}{n-1} \sum_{l\neq j}^{n} K_{hlj} (m_l - m_j + \varepsilon_l \sqrt{v(X_l,\theta_0)}) \Bigg) \varepsilon_j \hat{f}_j^2 \Bigg)^2 \\
= \; & \frac{h^d}{(n-1)^2} E\Bigg(\sum_{i=1}^{n} \sum_{j\neq i}^{n} K_{hij} \Bigg(\frac{1}{n-1} \sum_{k\neq i}^{n} K_{hki} (m_k - m_i) + \frac{1}{n-1} \sum_{k\neq i}^{n} K_{hki} \varepsilon_k \sqrt{v(X_k,\theta_0)} \Bigg) \\
& \qquad \cdot \Bigg(\frac{1}{n-1} \sum_{l\neq j}^{n} K_{hlj} (m_l - m_j) + \frac{1}{n-1} \sum_{l\neq j}^{n} K_{hlj} \varepsilon_l \sqrt{v(X_l,\theta_0)} \Bigg) \varepsilon_i \hat{f}_i^2 \varepsilon_j \hat{f}_j^2 \Bigg)^2
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{nh^d}{n-1}E\Bigg(K_{h12}\Bigg[\frac{1}{(n-1)^2}\Bigg(\sum_{k\neq 1}^{n}K_{hk1}(m_k-m_1)\Bigg)\Bigg(\sum_{l\neq 2}^{n}K_{hl2}(m_l-m_2)\Bigg)\\
&\qquad +\frac{2}{(n-1)^2}\Bigg(\sum_{k\neq 1}^{n}K_{hk1}\varepsilon_k\sqrt{v(X_k,\theta_0)}\Bigg)\Bigg(\sum_{l\neq 2}^{n}K_{hl2}(m_l-m_2)\Bigg)\\
&\qquad \frac{1}{(n-1)^2}\Bigg(\sum_{k\neq 1}^{n}K_{hk1}\varepsilon_k\sqrt{v(X_k,\theta_0)}\Bigg)\Bigg(\sum_{l\neq 2}^{n}K_{hl2}\varepsilon_l\sqrt{v(X_l,\theta_0)}\Bigg)\Bigg]\varepsilon_i\hat{f}_i^2\varepsilon_j\hat{f}_j^2\Bigg)^2\\
&:\leq 3V_{n10a}+6V_{n10b}+3V_{n10c}.
\end{aligned}
$$

Where the last line is obtained by using the Cauchy Schwartz inequality, we will later separately show that all of them converge to zero in probability.

Here, we begin with V_{n10a} and the proof of V_{n10c}, where V_{n10b} follows with similar technique. Also note that, we begin with $\tilde{V}_{n10a}$, indeed it would be easy to show that the replacement of $\hat{f}_{i,j}$ with $f_{i,j}$ will have terms converges to zero at an even faster rate due to the consistency.

$$
\begin{aligned}
&\tilde{V}_{n10a} = \frac{nh^d}{n-1}E\Bigg(K_{h12}\Bigg[\frac{1}{(n-1)^2}\Bigg(\sum_{k\neq 1}^{n}K_{hk1}(m_k-m_1)\Bigg)\Bigg(\sum_{l\neq 2}^{n}K_{hl2}(m_l-m_2)\Bigg)\Bigg]\varepsilon_i\varepsilon_j f_i^2 f_j^2\Bigg)^2\\
&= \frac{nh^d}{(n-1)^5}E\Bigg(K_{h12}^2\Bigg(\sum_{k\neq 1}^{n}K_{hk1}(m_k-m_1)\Bigg)^2\Bigg(\sum_{l\neq 2}^{n}K_{hl2}(m_l-m_2)\Bigg)^2\varepsilon_i^2\varepsilon_j^2 f_i^4 f_j^4\Bigg)\\
&= \frac{nh^d}{(n-1)^5}E\Bigg(K_{h12}^2\Bigg(\sum_{k\neq 1}^{n}K_{hk1}^2(m_k-m_1)^2+\sum_{k\neq 1}^{n}\sum_{i\neq k\neq 1}^{n}K_{hk1}K_{hi1}(m_k-m_1)(m_i-m_1)\Bigg)\\
&\quad\cdot\Bigg(\sum_{l\neq 2}^{n}K_{hl2}^2(m_l-m_2)^2+\sum_{l\neq 2}^{n}\sum_{j\neq l\neq 2}^{n}K_{hl2}K_{hj2}(m_l-m_2)(m_j-m_2)\Bigg)E(\varepsilon_i^2|X_i)E(\varepsilon_j^2|X_j)f_i^4 f_j^4\Bigg)\\
&= \frac{nh^d}{(n-1)^5}E\Bigg(K_{h12}^2\Bigg(\sum_{k\neq 1}^{n}K_{hk1}^2(m_k-m_1)^2+\sum_{k\neq 1}^{n}\sum_{i\neq k\neq 1}^{n}K_{hk1}K_{hi1}(m_k-m_1)(m_i-m_1)\Bigg)\\
&\quad\cdot\Bigg(\sum_{l\neq 2}^{n}K_{hl2}^2(m_l-m_2)^2+\sum_{l\neq 2}^{n}\sum_{j\neq l\neq 2}^{n}K_{hl2}K_{hj2}(m_l-m_2)(m_j-m_2)\Bigg)\sigma_i^2\sigma_j^2 f_i^4 f_j^4\Bigg)\\
&= \frac{nh^d}{(n-1)^5}E\Bigg(K_{h12}^2\Bigg(\sum_{k\neq 1}^{n}\sum_{l\neq 2}^{n}K_{hk1}^2K_{hl2}^2(m_k-m_1)^2(m_l-m_2)^2\Bigg)\sigma_i^2\sigma_j^2 f_i^4 f_j^4\Bigg)\\
&\quad+\frac{2nh^d}{(n-1)^5}E\Bigg(K_{h12}^2\Bigg(\sum_{k\neq 1}^{n}\sum_{l\neq 2}^{n}\sum_{j\neq l\neq 2}^{n}K_{hk1}^2K_{hl2}K_{hj2}(m_k-m_1)^2(m_l-m_2)(m_j-m_2)\Bigg)\sigma_i^2\sigma_j^2 f_i^4 f_j^4\Bigg)
\end{aligned}
$$

$$+\frac{nh^d}{(n-1)^5}E\Bigg(K_{h12}^2\Bigg(\sum_{k\neq 1}^n\sum_{i\neq k\neq 1}^n\sum_{l\neq 2}^n\sum_{j\neq l\neq 2}^n K_{hk1}K_{hi1}K_{hl2}K_{hj2}(m_k-m_1)(m_i-m_1)$$

$$\cdot(m_l-m_2)(m_j-m_2)\Bigg)\sigma_i^2\sigma_j^2f_i^4f_j^4\Bigg)$$

$$= O_p\Bigg(\frac{1}{n^2h^{2d-8}}\Bigg)+O_p\Bigg(\frac{1}{nh^{d-8}}\Bigg)+O_p(h^8).$$

The last equation follows from the assumptions and the Taylor expansion after change of variables. Next, we have,

$$\begin{aligned}
\tilde{V}_{n10c} &= \frac{nh^d}{n-1}E\Bigg(K_{h12}\Bigg[\frac{1}{(n-1)^2}\Bigg(\sum_{k\neq 1}^n K_{hk1}\varepsilon_k\sqrt{v(X_k,\theta_0)}\Bigg)\Bigg(\sum_{l\neq 2}^n K_{hl2}\varepsilon_l\sqrt{v(X_l,\theta_0)}\Bigg)\Bigg]\varepsilon_i\varepsilon_j f_i^2f_j^2\Bigg)^2 \\
&= \frac{nh^d}{(n-1)^5}E\Bigg(K_{h12}\Bigg[\Bigg(\sum_{k\neq 1}^n K_{hk1}\varepsilon_k\sqrt{v(X_k,\theta_0)}\Bigg)^2\Bigg(\sum_{l\neq 2}^n K_{hl2}\varepsilon_l\sqrt{v(X_l,\theta_0)}\Bigg)^2\Bigg]\sigma_i^2\sigma_j^2f_i^4f_j^4\Bigg) \\
&= \frac{nh^d}{(n-1)^5}E\Bigg(K_{h12}^2\Bigg(\sum_{k\neq 1}^n K_{hk1}^2\varepsilon_k^2v(X_k,\theta_0)+\sum_{k\neq 1}^n\sum_{i\neq k\neq 1}^n K_{hk1}K_{hi1}\varepsilon_k\varepsilon_i\sqrt{v(X_k,\theta_0)v(X_i,\theta_0)}\Bigg) \\
&\quad\cdot\Bigg(\sum_{l\neq 2}^n K_{hl2}^2\varepsilon_l^2v(X_l,\theta_0)+\sum_{l\neq 2}^n\sum_{j\neq l\neq 2}^n K_{hl2}K_{hj2}\varepsilon_l\varepsilon_j\sqrt{v(X_l,\theta_0)v(X_j,\theta_0)}\Bigg)\sigma_i^2\sigma_j^2f_i^4f_j^4\Bigg).
\end{aligned}$$

From now on, things may get complicated yet with much simplification following the routine of V_{n10a}. First, notice that we need to begin our sums from 3 instead of 1 or 2, and we will repeatedly use the fact that $E(\varepsilon_i|X_i)=0$ and $X_i's$ are independent of each other. Therefore, terms that involve terms with ε_i to the first order will be zero, and the dominant term would be

$$\frac{2nh^d}{(n-1)^5}E\Bigg(K_{h12}^2\Bigg(\sum_{k=3}^n\sum_{l=3}^n K_{hk1}^2K_{hl2}^2\varepsilon_k^2\varepsilon_l^2v(X_k,\theta_0)v(X_l,\theta_0)\Bigg)\sigma_i^2\sigma_j^2f_i^4f_j^4\Bigg)$$

$$= O\Bigg(\frac{1}{n^2h^{2d}}\Bigg).$$

The coefficient 2 comes from the fact that when the pair (i,k) and (j,l) equals, there will be again $(n-2)^2$ many of them, and they won't be zero. The other terms are either of smaller order or equals zero after taking expectation. And this concludes the proof of V_{n10}.

Next for V_{n4}, we would show that $En^2h^dV_{n4}^2 = o(1)$ and apply Chebyshev inequality. Note the similarity between the role of η_i and ε_i, both have conditional mean zero. Here we ignore the coefficient -4.

$$\begin{aligned}
E(n^2h^dV_{n4}^2) &= \frac{h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^n\sum_{i\neq i}^n K_{hij}\eta_i\Delta m_j\varepsilon_j\sqrt{v(X_j,\theta_0)}\hat{f}_i^2\hat{f}_j^2\Bigg)^2 \\
&= \frac{h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^n\sum_{i\neq i}^n K_{hij}^2\eta_i^2\Delta m_j^2\varepsilon_j^2 v(X_j,\theta_0)\hat{f}_i^2\hat{f}_j^2\Bigg) \\
&= \frac{h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^n\sum_{i\neq i}^n K_{hij}^2\eta_i^2\Big(\frac{1}{n-1}\sum_{k\neq j}^n K_{nkj}(Y_k-m_j)\Big)^2\varepsilon_j^2 v(X_j,\theta_0)f_i^2\Bigg) + o(1) \\
&\leq \frac{2h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^n\sum_{i\neq i}^n K_{hij}^2\eta_i^2\Big(\frac{1}{n-1}\sum_{k\neq j}^n K_{nkj}(m_k-m_j)\Big)^2\varepsilon_j^2 v(X_j,\theta_0)f_i^2\Bigg) \\
&\quad + \frac{2h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^n\sum_{i\neq i}^n K_{hij}^2\eta_i^2\Big(\frac{1}{n-1}\sum_{k\neq j}^n K_{nkj}(\varepsilon_k\sqrt{V(X_k,\theta_0)})\Big)^2\varepsilon_j^2 v(X_j,\theta_0)f_i^2\Bigg) + o(1) \\
&= O(h^4) + O(1/nh^d) + o(1).
\end{aligned}$$

The last equality is obtained with the usual technique used in the proof of V_{n10}. Next we will work with V_{n9} where we may utilize many existing results in V_{n3}. Here it suffices to show that $E(n^2h^dV_{n9}^2) = o(1)$.

$$\begin{aligned}
E(\|n^2h^dV_{n9}^2\|) &= \frac{h^d}{(n-1)^2}E\Bigg\|\Bigg(\sum_{i=1}^n\sum_{i\neq j}^n K_{hij}\Delta v_i\Delta m_j\varepsilon_j\sqrt{v(X_j,\theta_0)}\hat{f}_i^2\hat{f}_j^2\Bigg)\Bigg\|^2 \\
&= \frac{h^d}{(n-1)^2}E\Bigg\|\Bigg(\sum_{i=1}^n\sum_{i\neq j}^n K_{hij}^2\Delta v_i^2\Delta m_j^2\varepsilon_j^2 v(X_j,\theta_0)\hat{f}_i^4\hat{f}_j^4 \\
&\quad + \sum_{i=1}^n\sum_{i\neq j}^n\sum_{k\neq i\neq j}^n K_{hij}K_{hkj}\Delta v_i\Delta v_k\Delta m_j^2\varepsilon_j^2 v(X_j,\theta_0)\hat{f}_i^2\hat{f}_k^2\hat{f}_j^4\Bigg)\Bigg\|^2 \\
&\leq \frac{h^d}{(n-1)^2}E\Bigg(\sum_{i=1}^n\sum_{i\neq j}^n K_{hij}^2\|\Delta v_i^2\|\Delta m_j^2\varepsilon_j^2 v(X_j,\theta_0)\hat{f}_i^4\hat{f}_j^4 \\
&\quad + \sum_{i=1}^n\sum_{i\neq j}^n\sum_{k\neq i\neq j}^n K_{hij}K_{hkj}\|\Delta v_i\|\|\Delta v_k\|\Delta m_j^2\varepsilon_j^2 v(X_j,\theta_0)\hat{f}_i^2\hat{f}_k^2\hat{f}_j^4\Bigg)^2
\end{aligned}$$

$$= \frac{h^d}{(n-1)^2} E\Bigg(\sum_{i=1}^{n} \sum_{i \neq j}^{n} K_{hij}^2 \Big(\frac{1}{n-1} \sum_{l \neq j}^{n} K_{hlj}(Y_l - m_j) \Big)^2 \varepsilon_j^2 v(X_j, \theta_0) f_i^4 f_j^2$$

$$+ \sum_{i=1}^{n} \sum_{i \neq j}^{n} \sum_{k \neq i \neq j}^{n} K_{hij} K_{hkj} \Big(\frac{1}{n-1} \sum_{l \neq j}^{n} K_{hlj}(Y_l - m_j) \Big)^2 \varepsilon_j^2 v(X_j, \theta_0) f_i^2 f_k^2 f_j^2 \Bigg)^2 O\Big(\frac{1}{n}\Big) + o(1)$$

Writing $Y_l = m_l + \varepsilon_l \sqrt{v(X_l, \theta_0)}$ gives us the usual expansion, and it is not difficult to see that the above quantity is dominated by the second part of the summation, which is of order $O(h^4) + O(1/nh^d) + o(1)$.

Finally, we will show that $E(n^2 h^d V_{n7}^2) = o(1)$, note that we can utilize the fact with $E(\varepsilon_i | X_i) = 0$ again.

$$E(n^2 h^d V_{n7}^2) = \frac{h^d}{(n-1)^2} E\Bigg(\sum_{i=1}^{n} \sum_{j \neq i}^{n} K_{hij} \Delta m_i^2 \Delta m_j \varepsilon_j \sqrt{v(X_j, \theta_0)} \hat{f}_i^2 \hat{f}_j^2 \Bigg)^2$$

$$= \frac{h^d}{(n-1)^8} E\Bigg(\sum_{i=1}^{n} \sum_{j \neq i}^{n} K_{hij} \Big(\sum_{k \neq i}^{n} K_{hki}(Y_k - m_i) \Big)^2 \Big(\sum_{l \neq j}^{n} K_{hlj}(Y_l - m_j) \Big) \varepsilon_j \sqrt{v(X_j, \theta_0)} \hat{f}_j \Bigg)^2$$

$$= \frac{h^d}{(n-1)^8} E\Bigg(\sum_{i=1}^{n} \sum_{j \neq i}^{n} K_{hij}^2 \Big(\sum_{k \neq i}^{n} K_{hki}(Y_k - m_i) \Big)^4 \Big(\sum_{l \neq j}^{n} K_{hlj}(Y_l - m_j) \Big)^2 \varepsilon_j^2 v(X_j, \theta_0) f_j^2$$

$$+ \sum_{i=1}^{n} \sum_{j \neq i}^{n} \sum_{s \neq j \neq i} K_{hij} K_{hsj} \Big(\sum_{k \neq i}^{n} K_{hki}(Y_k - m_i) \Big)^2 \Big(\sum_{k \neq s}^{n} K_{hks}(Y_k - m_s) \Big)^2$$

$$\cdot \Big(\sum_{l \neq j}^{n} K_{hlj}(Y_l - m_j) \Big)^2 \varepsilon_j^2 v(X_j, \theta_0) f_j^2 \Bigg) + o(1)$$

$$= O\Big(\frac{1}{n^2 h^{2d}} \Big) + O(h^{12}) + o(1)$$

The last equality comes from usual calculation except that the combination in the latter half of the expansion may appear with two equality pair, therefore instead of $O(1/n^3 h^{3d})$, the dominate term is of order $O(1/n^2 h^{2d})$. □

Proof of Lemma 4. First note that from the previous lemmas, we may eliminate the bias term of the $\hat{\Sigma}$, that is,

$$\begin{aligned}
\hat{\Sigma} &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} h^d K_{hij}^2 \hat{\eta}_i^2 \hat{f}_h^4(X_i)\hat{\eta}_j^2 \hat{f}_h^4(X_j) \\
&= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} h^d K_{hij}^2 \eta_i^2 f^4(X_i)\eta_j^2 f^4(X_j) + o_p(1) \\
&:= 2V_{ns} + o_p(1).
\end{aligned}$$

Here, observing the structure of V_{ns}, we may notice that it is a U-statistic with the kernel function $H_n(z_1, z_2) = h^d K_h^2(X_i - X_j)\eta_i^2 f^4(X_i)\eta_j^2 f^4(X_j)$. Therefore, similar to the proof of Lemma 3.3b and Lemma 3.3e in Zheng (1996), adding the condition that function $\|M(x)\| \leq b(x)$ for $x \in R^d$ with $b(X_i)$ square integrable with respect to the measure of X_i,

$$\begin{aligned}
& E(\|H_n(z_i, z_j)\|^2) \\
\leq\ & 2E\left(\frac{h^d}{2}K_h^2(X_i - X_j)\eta_i^2 f^4(X_i) m_j\right)^2 + 2E\left(\frac{h^d}{2}K_h^2(X_i - X_j)\eta_j^2 f^4(X_j) m_i\right)^2 \\
=\ & \int \frac{1}{h^{2d}} K^4\left(\frac{x_i - x_j}{h}\right)(\tau^2(x_i))^2 M^2(x_j) f^9(x_i) f(x_j) dx_i dx_j \\
=\ & \frac{1}{h^d}\int K^4(u)(\tau^2(x_i))^2 M^2(x_i + uh) f^9(x_i) f(x_i + uh) dx_i du \\
=\ & O_p\left(\frac{1}{h^d}\right) = o_p(n).
\end{aligned}$$

So, we have,obtained a similar result as Lemma 3.3b in Zheng (1996), following the previous routine,

$$\begin{aligned}
\bar{r}_n &= E(h^d K_h^2(X_i - X_j)\eta_i^2 f^4(X_i)\eta_j^2 f^4(X_j)) \\
&= \int K^2(u)\tau^2(x_i)\tau^2(x_i + uh) f^5(x_i) f^5(x_i + uh) dx_i du \\
&= \int K^2(u) du \int (\tau^2(x))^2 f^{10}(x) dx + o(1) \\
&= \Sigma/2 + o(1).
\end{aligned}$$

By Theorem 3.2.1, we have $\hat{\Sigma} = \Sigma + o_p(1)$.

□

Proof of Theorem 3.3.1. First we show that $V_n \to \int (v_a(x) - v(x, \theta_0))^2 f^{10}(x)dx$. Similar to the proof of Lemma 4, plus we also need the results from all previous lemmas, under the alternative hypothesis, let $\eta_{ai} = (Y_i - \hat{m}_i)^2 - v_a(X_i)$ the expansion of the statistic V_n can be written as,

$$\begin{aligned} V_{an}(x) &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij} \hat{\eta}_{ai} \hat{f}_h^2(X_i) \hat{\eta}_{aj} \hat{f}_h^2(X_j) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij} [(Y_i - \hat{m}_i)^2 - v(\hat{\theta}_n, X_i)][(Y_j - \hat{m}_j)^2 - v(\hat{\theta}_n, X_j)] \hat{f}_h^2(X_i) \hat{f}_h^2(X_j) \\ &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij} [(Y_i - m_i + m_i - \hat{m}_i)^2 - v_a(X_i) + v_a(X_i) - v(X_i, \theta_a) \\ &\quad + v(X_i, \theta_a) - v(\hat{\theta}_n, X_i)][(Y_j - m_j + m_j - \hat{m}_j)^2 - v_a(X_j) + v_a(X_j) - v(X_j, \theta_a) \\ &\quad + v(X_j, \theta_a) - v(\hat{\theta}_n, X_j)] \hat{f}_h^2(X_i) \hat{f}_h^2(X_j). \end{aligned}$$

After expansion we obtain 15 terms this time, we will notice one term converges to a constant, and the rest are of order $o_p(1)$, or after multiplying $nh^{d/2}$ becomes $O_p(1)$ instead of infinity. We will show the following representative terms with detail, while the rest follows the same procedure while proving the case under Null hypothesis.

$$V_{an1} := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij} [(Y_i - m_i)^2 - v_a(X_i)][(Y_i - m_j)^2 - v_a(X_j)] \hat{f}_h^2(X_i) \hat{f}_h^2(X_j).$$

$$V_{an2} := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij} [v_a(X_i) - v(X_i, \theta_a)][v_a(X_j) - v(X_j, \theta_a)] \hat{f}_h^2(X_i) \hat{f}_h^2(X_j).$$

$$V_{an3} := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij} [(Y_i - m_i)^2 - v_a(X_i)][v_a(X_j) - v(X_j, \theta_a)] \hat{f}_h^2(X_i) \hat{f}_h^2(X_j).$$

$$V_{an4} := \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij} [v(X_i, \theta_a) - v(X_i, \hat{\theta}_n)][v_a(X_j) - v(X_j, \theta_a)] \hat{f}_h^2(X_i) \hat{f}_h^2(X_j).$$

Following the case under null hypothesis, it would be easy to see that $nh^{d/2}V_{an1}$ converges to a normal distribution. Next, note that $V_{an1} + V_{an2}$ is also a U-statistic with $H_n(z_i, z_j) = K_{hij}(\eta_{ai} + \nu_{ai})(\eta_{aj} + \nu_{aj})f^2(X_i)f^2(X_j)$, and similar to the previous proofs, also, plus the fact that $E((\eta_{ai} + \nu_{ai})f^2(X_i)|X_i) = (v_a(X_i) - v(X_i, \theta_a))f^2(X_i)$, we have,

$$\begin{aligned}
\bar{r}_n &= E(E[H_n(z_i, z_j)|X_i, X_j]) \\
&= \frac{1}{h^d}\int K\left(\frac{x_i - x_j}{h}\right)(v_a(x_i) - v(x_i, \theta_a))(v_a(x_j) - v(x_j, \theta_a))f^5(x_i)f^5(x_j)dx_i dx_j \\
&= \int K(u)[v_a(x_i) - v(x_i, \theta_a)][v_a(x_i - hu) - v(x_i - hu, \theta_a)]f^5(x_i)f^5(x_i - hu)dx_i du \\
&= \int [v_a(x) - v(x, \theta_a)]^2 f^{10}(x)dx + o(1).
\end{aligned}$$

And following Theorem 3.2.1, we have proven equation $V_n \to \int (v_a(x) - v(x, \theta_0))^2 f^{10}(x)dx$, and the proof of equation $\hat{\Sigma} \to 2\int K^2(u)du \int [\tau^2(x) + (v_a(x) - v(x, \theta_a))]^2 f^{10}(x)dx > 0$ follows from Lemma 4.

Under the alternative hypothesis, again with the $\hat{\eta}_{ai}$ and η_{ai}, the difference may be represented as $o_p(1)$ from the proof under null hypothesis,

$$\begin{aligned}
\hat{\Sigma} &= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} h^d K_{hij}^2 \hat{\eta}_{ai}^2 \hat{\eta}_{aj}^2 \hat{f}_h^4(X_i)\hat{f}_h^4(X_j) \\
&= \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n} h^d K_{hij}^2 \eta_{ai}^2 \eta_{aj}^2 f^4(X_i) f^4(X_j) + o_p(1).
\end{aligned}$$

It is easy to observe that the above term is a U-statistic with $H_n(z_i, z_j) = h^d K_{hij}^2 \eta_{ai}^2 \eta_{aj}^2 f^4(X_i) f^4(X_j)$, and $E(\eta_{ai}^2 f^4(X_i)|X_i) = \tau^2(X_i) + (v_a(X_i) - v(X_i, \theta_a))^2$, therefore

$$\begin{aligned}
\bar{r}_n &= E(E[H_n(z_i, z_j)|X_i, X_j]) \\
&= \frac{1}{h^d}\int K^2\left(\frac{x_i - x_j}{h}\right)(\tau^2(x_i) + (v_a(x_i) - v(x_i, \theta_a))^2) \\
&\qquad \cdot(\tau^2(x_j) + (v_a(x_j) - v(x_j, \theta_a))^2)f^5(x_i)f^5(x_j)dx_i dx_j
\end{aligned}$$

$$
\begin{aligned}
&= \int K^2(u)[\tau^2(x_i) + v_a(x_i) - v(x_i, \theta_a)][v_a(x_i - hu) + v_a(x_i - hu) - v(x_i - hu, \theta_a)] \\
&\qquad\qquad\qquad\qquad \cdot f^5(x_i) f^5(x_i - hu) dx_i du \\
&= \int K^2(u) du \int [\tau^2(x) + (v_a(x) - v(x, \theta_a))^2]^2 f^{10}(x) dx + o(1).
\end{aligned}
$$

as desired. Therefore, we may conclude that the statistics goes to infinity as $n \to \infty$.

Next, we may show that V_{an3} and V_{an4} is of order $o_p(1)$. Use the fact that $E(\eta_a|X) = 0$, and with the fact that $\int (v_a(x) - v(x, \theta_a))^2 dx < \infty$, by Cauchy-Schwartz inequality,

$$
\begin{aligned}
& E(V_{an3}^2) \\
&= E\left(\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij} [(Y_i - m_i)^2 - v_a(X_i)][v_a(X_j) - v(X_j, \theta_a)] \hat{f}_h^2(X_i) \hat{f}_h^2(X_j) \right)^2 \\
&= E\left(\frac{1}{n^2(n-1)^2} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij}^2 \eta_i^2 [v_a(X_j) - v(X_j, \theta_a)]^2 \hat{f}_h^4(X_i) \hat{f}_h^4(X_j) \right. \\
&\quad + \frac{1}{n^2(n-1)^2} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \sum_{k=1, k\neq j\neq i}^{n} K_{hij} K_{hik} \eta_i^2 [v_a(X_j) - v(X_j, \theta_a)] \\
&\qquad\qquad \left. \cdot [v_a(X_k) - v(X_k, \theta_a)] \hat{f}_h^4(X_i) \hat{f}_h^2(X_j) \hat{f}_h^2(X_k) \right)^2 \\
&= O\left(\frac{1}{n^2 h^d} \right) + O\left(\frac{1}{n} \right) + o(1),
\end{aligned}
$$

where the last $o(1)$ stands for the bias from $\hat{f}_h(X_i)$ and $f(X_i)$. And similarly for V_{an4}, again we use the fact that $\|v(x, \theta_a) - v_a(x)\| = O_p(n^{-1/2})$, $E(\|V_{an4}^2\|)$ equals the following,

$$
\begin{aligned}
& E\left(\| \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij} [v(X_i, \theta_a) - v(X_i, \hat{\theta}_n)][v_a(X_j) - v(X_j, \theta_a)] \hat{f}_h^2(X_i) \hat{f}_h^2(X_j) \| \right)^2 \\
&\le O_p(n^{-1}) E\left(\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} K_{hij} [v_a(X_j) - v(X_j, \theta_a)] \hat{f}_h^2(X_i) \hat{f}_h^2(X_j) \right)^2 = O_p\left(\frac{1}{n} \right).
\end{aligned}
$$

And we have completed the proof of consistency.

□

Proof of Theorem 3.3.2. The proof of the local power begins similarly with the quantity V_n, except under the local alternative $v(X,\theta_0)+\delta_n l(X)$, and we denote it as $v_{loc}(X)$ say, further $\eta_{loc,i}$ as the variance function replaced with the local alternative,

$$\begin{aligned}
V_n &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{hij}\hat{\eta}_i\hat{\eta}_j\hat{f}_h^2(X_i)\hat{f}_h^2(X_j) \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{hij}\hat{\eta}_i\hat{\eta}_j\hat{f}_h^2(X_i)\hat{f}_h^2(X_j) \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{hij}[(Y_i-\hat{m}_i)^2-v(X_i,\hat{\theta}_n)] \\
&\qquad \cdot[(Y_j-\hat{m}_j)^2-v(X_j,\hat{\theta}_n)]\hat{f}_h^2(X_i)\hat{f}_h^2(X_j) \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{hij}[(Y_i-\hat{m}_i)^2-v_{loc}(X_i)+v_{loc}(X_i)-v(X_i,\hat{\theta}_n)] \\
&\qquad \cdot[(Y_j-\hat{m}_j)^2-v_{loc}(X_j)+v_{loc}(X_j)-v(X_j,\hat{\theta}_n)]\hat{f}_h^2(X_i)\hat{f}_h^2(X_j) \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{hij}[\eta_{loc,i}+v_{loc}(X_i)-v(X_i,\theta_0)+v(X_i,\theta_0)-v(X_i,\hat{\theta}_n)] \\
&\qquad \cdot[\eta_{loc,j}+v_{loc}(X_j)-v(X_j,\theta_0)+v(X_j,\theta_0)-v(X_j,\hat{\theta}_n)]\hat{f}_h^2(X_i)\hat{f}_h^2(X_j) \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{hij}[\eta_{loc,i}+\delta_n l(X_i)][\eta_{loc,j}+\delta_n l(X_j)]f^2(X_i)f^2(X_j)+o_p(1) \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{hij}\eta_{loc,i}\eta_{loc,j}f^2(X_i)f^2(X_j) \\
&\quad +\frac{2\delta_n}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{hij}\eta_{loc,i}l(X_j)f^2(X_i)f^2(X_j) \\
&\quad +\frac{\delta_n^2}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n} K_{hij}l(X_i)l(X_j)f^2(X_i)f^2(X_j)+o_p(1) \\
&:= V_{nl1}+2V_{nl2}+V_{nl3}+o_p(1).
\end{aligned}$$

By putting the δ_n in the front, our proof maybe more straight forward when multiplying $nh^{d/2}$ in the front, and this is how we find the local alternative rate $1/\sqrt{nh^{d/2}}$. Since $nh^{d/2}V_{nl1}\to N(0,\Sigma)$ in distribution with Σ having the same form as null hypothesis while changing variance function to the $v_{loc}(X)$. Then, for V_{nl3}, with the local alternative rate $1/\sqrt{nh^{d/2}}$,

$$E(nh^{d/2}V_{nl3}) = E(K_{hij}l(X_i)l(X_j)f^2(X_i)f^2(X_j)) = \int l^2(x)f^6(x)dx$$

And we also have

$$\begin{aligned} E(n^2h^dV_{nl2}^2) &= E\Bigg[\frac{2h^{3d/4}}{n^{1/2}(n-1)^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}K_{hij}^2\eta_{loc,i}^2 l^2(X_j)f^4(X_i)f^4(X_j) \\ &\quad + \sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{k=1,k\neq i\neq j}^{n}K_{hij}K_{hik}\eta_{loc,i}^2 l(X_j)l(X_k)f^4(X_i)f^2(X_j)f^2(X_k)\Bigg] \\ &= O\Big(\frac{1}{n^{1/2}h^{d/4}}\Big) + O(n^{1/2}h^{3d/4}) = o_p(1). \end{aligned}$$

Therefore, following the previous proof, with Slutsky's theorem, the above convergence in probability to zero plus the convergence in probability to $\int l^2(x)f^6(x)dx$, and the convergence in distribution to normal of V_{nl1}, we obtain the desired result in the theorem. □

Bibliography

Lixing Zhu, Yasunori Fujikoshi, and Kanta Naito. Heteroscedasticity checks for regression models. *Science in China Series A: Mathematics*, 44(10):1236–1252, Oct 2001.

Holger Dette, Axel Munk, and Thorsten Wagner. Estimating the variance in nonparametric regression—what is a reasonable choice? *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(4):751–764, 1998.

Holger Dette and Benjamin Hetzler. A simple test for the parametric form of the variance function in nonparametric regression. *Annals of the Institute of Statistical Mathematics*, 61(4):861, 2009.

Holger Dette, Natalie Neumeyer, and Ingrid Van Keilegom. A new test for the parametric form of the variance function in non-parametric regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):903–917, 2007.

John Xu Zheng. A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75(2):263–289, 1996.

Holger Dette. A consistent test for heteroscedasticity in nonparametric regression based on the kernel method. *Journal of statistical planning and inference*, 103(1-2):311–329, 2002.

Xu Zheng. Testing heteroscedasticity in nonlinear and nonparametric regressions. *Canadian Journal of Statistics*, 37(2):282–300, 2009.

J.C. Pardo-Fernández and M.D. Jiménez-Gamero. A model specification test for the variance function in nonparametric regression. *AStA Adv Stat Anal*, 103(3):387–410, 2019. doi: 10.1007/s10182-018-00336-y.

Hira L Koul and Pingping Ni. Minimum distance regression model checking. *Journal of Statistical Planning and Inference*, 119(1):109–141, 2004.

Nishantha Samarakoon and Weixing Song. Minimum distance conditional variance function checking in heteroscedastic regression models. *Journal of Multivariate Analysis*, 102(3): 579 – 600, 2011. ISSN 0047-259X.

Nithansha Samarakoon and Weixing Song. Empirical smoothing lack-of-fit tests for variance function. *Journal of Statistical Planning and Inference*, 142(5):1128–1140, 2012.

Weixing Song and Juan Du. A note on testing the regression functions via nonparametric smoothing. *The Canadian Journal of Statistics*, 39(1):108–125, 2011.

Yue-pok Mack and Bernard W Silverman. Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 61 (3):405–415, 1982.

Jiti Gao and Irene Gijbels. Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association*, 103(484):1584–1594, 2008.

Jeffrey Hart. *Nonparametric smoothing and lack-of-fit tests*. Springer Science & Business Media, 2013.

I.G. Frankel J. Liang E.J. Fitzsimmons G.L. Newmark. Evaluating police-caused crashes. *Accepted for presentation and publication in Transportation Research Record*, 99th Annual Meeting of the Transportation Research Board, Washington, D.C., 2020.

Denis Bosq. *Nonparametric statistics for stochastic processes: estimation and prediction*, volume 110. Springer Science & Business Media, 2012.

Peter Hall. Central limit theorem for integrated square error of multivariate nonparametric density estimators. *Journal of multivariate analysis*, 14(1):1–16, 1984.

Holger Dette et al. A consistent test for the functional form of a regression based on a difference of variance estimators. *The Annals of Statistics*, 27(3):1012–1040, 1999.

Natalie Neumeyer and Ingrid Van Keilegom. Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis*, 101(5):1067–1078, 2010.

Lingzhu Li, Sung Nok Chiu, and Lixing Zhu. Model checking for regressions: An approach bridging between local smoothing and global smoothing methods. *Computational Statistics & Data Analysis*, 138:64–82, 2019.

Robert I Jennrich. Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, 40(2):633–643, 1969.