Title
Three Essays on Nonparametric Hypothesis Testing

Author
Kim, Seolah

Publication Date
2020

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Three Essays on Nonparametric Hypothesis Testing

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Seolah Kim

June 2020

Dissertation Committee:

Professor Aman Ullah, Chairperson
Professor Tae-Hwy Lee
Professor Ruoyao Shi

The Dissertation of Seolah Kim is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I express my deepest gratitude to my advisor, Professor Aman Ullah. It was one of the best life decisions for me to pursue my career under his supervision. Without his insightful guidance on my research, I would have never been able to be obtain a Ph.D. at the University of California, Riverside. Not only was he a great example of a true researcher for his passion and knowledge, but he has also been my role model as a mentor with his generosity and kindness. I will carry these lessons from him for ever so long.

I thank Professor Tae-Hwy Lee and Professor Ruoyao Shi for their great support in leading me toward being more productive and thorough about research. Their intuitive comments and advice helped me improve the quality of the papers. I am grateful to Professor Michael Bates for allowing me to collaborate in research during my Ph.D. program. At the Department of Economics, I would also like to thank Gary Kuzas, Damaris Carlos, and Wendy Mello for their amazing administrative support and assistance whenever needed.

I would like to thank my grandmother, my parents, and my whole family. Thanks to their unwavering support, I was able to stand where I am now and fulfill my passion for research. I always feel assured and encouraged by their presence around me.

Throughout my Ph.D. program, I cannot express how grateful I am to have Maxwell Chomas and Sabrina Park. They have been such a good company for all my good and bad moments in life. I appreciate their considerable encouragement whenever I was in a struggle. Thanks to them, I could pull myself together and keep moving on.

Lastly, I was glad to build up long-lasting friendships with fellow graduate students including Sungjun Huh, Hanbyul Ryu, Yoon-Jae Ro, Jianghao Chu, and Jialuo Chen.

To my late grandfather, Jae-Yoon Kim

ABSTRACT OF THE DISSERTATION

Three Essays on Nonparametric Hypothesis Testing

by

Seolah Kim

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, June 2020
Professor Aman Ullah, Chairperson

Nonparametric approaches have widely been used due to their advancement in not making assumptions on the distribution of the data. Even with their extensive development, nonparametric hypothesis testing has not been developed as much as a nonparametric estimation even though it is one of the key components of the econometric analysis. This dissertation has mainly two parts. I first explore the systematic development of the current nonparametric tests and provide results on testing linearity as an illustration. Then I develop new nonparametric tests for detecting endogeneity in cross-sectional data and panel data respectively.

Elaborating each test's performance can be meaningful in that it allows us to decide on which test to use depending on the hypothesis and even construct a new test based on such a relationship. Under the hypotheses for linearity, Chapter 2 will categorize the types of nonparametric tests and discuss the analytical relationship of those tests. By imposing some conditions, I can compare the local power of each test asymptotically. While

examining the analytical relationship, I also develop a nonparametric Rao-Score test and show that it is equivalent to the Su and Ullah (2013) test.

Once analyzing the analytical relationship of the current nonparametric tests, I focus on developing a new nonparametric test for endogeneity. Since endogeneity is commonly observed in many economic contexts, detecting its presence is a preliminary step for choosing an estimation strategy. In Chapter 3, I construct a test using the control function approach under a triangular simultaneous equations model. My test can be summarized as being simple to implement as a test and being able to capture the locally nonlinear correlation with kernel weighting. Furthermore, I will apply these tests to the empirical analyses and show the contradicting results with the parametric test.

Not only in triangular simulation equations model, but also is endogeneity important model specification issue in panel data setting. The estimation strategy differs depending on the presence of endogeneity between the individual-specific components and the variable. I propose a new estimation method for the nonparametric panel random effects model and construct a new test for endogeneity using the residuals from the proposed estimation method. By obtaining the individual-specific effects in the random effects model, I construct a test over the i index instead of the i index and time. With a large T, the test performs well in terms of size and power.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Since the nonparametric methods have been introduced in statistics and econometrics, they have been powerful in econometric analyses due to their advancement in not making assumptions on the distribution of the data. Extensive development of the estimation methods range from a density estimation to a single index model (Rosenblatt (1956), Rao (1985), Silverman (1986), Ullah (1985, 1988), Horowitz (1992), Klein and Spady (1993), Ichimura (1993), Linton and Nielson (1995), Fan and Gijbels (1996), Newey et al. (1999), Li and Racine (2003), Fan and Yao (2005), among others). On the consolidated background, there is an ongoing research about a consistent nonparametric estimation.

Once running the estimation, a hypothesis testing is one of the key components of the econometric analysis to analyze the effect with the statistical significance. However, nonparametric hypothesis testing has not been developed systematically and the current nonparametric tests are ad-hoc. Elaborating each test's performance will allow us to understand the principles of the test statistic and even construct a new test based on such a relationship. For example, Ullah and Zinde-Walsh (1984) compared the robustness both in small and large sample between the likelihood function based tests. Therefore, the goal of this dissertation is to propose a systematic approach in a nonparametric hypothesis testing for linearity and then develop new nonparametric tests for the model specification and endogeneity as an illustration.

In Chapter 2, I investigate the analytical relationship between the existing nonparametric test statistics for the linearity test. Among many different nonparametric hypothesis tests, I focus on Li-Wang type conditional moment test, Su-Ullah type goodness-of-fit test, Yao-Ullah type goodness-of-fit test, and F type test. I present which nonparametric test is locally most powerful, which can be useful for empirical researchers in conducting a test for model specification. By imposing some conditions, I can compare the local power of each test asymptotically. Furthermore, I develop a nonparametric Rao-Score test for the model specification and show its equivalence to Su-Ullah type goodness-of-fit test. While a nonparametric test is superior over parametric tests due to its consistency, the Rao-Score test has not been developed yet in a fully nonparametric context. Using the local log-likelihood functions, I can develop a fully nonparametric Rao-Score test.

Chapter 3 introduces a consistent nonparametric test for endogeneity using a triangular simultaneous equations model. In such a setting, I take the control function approach to obtain the conditional moment of interest $E[U|V]$, where U is the error term of the structural equation and V is the error term from the reduced-form equation. This conversion opens a new way of constructing a test because it significantly reduces the dimension when estimating the conditional moment, which can alleviate the curse of dimensionality. In constructing a test, I use nonparametric residuals to obtain the consistency of the test. My test has strengths in that it is easy to implement as its asymptotic distribution is the standard normal and it can capture the locally nonlinear correlation with kernel weighting. Then, I apply this test using the data from Autor, Dorn, Hanson (2013) and get the contradicting results with the parametric test.

Endogeneity issue is important both in cross-sectional and panel data settings. While I develop a new nonparametric test for endogeneity with cross-sectional data, Chapter 4 proposes a new estimation method for the nonparametric panel random effects model and develops a new nonparametric test for endogeneity. In a panel data setting, testing for endogeneity in an individual fixed effects determines whether to use the fixed effects or the random effects model. There has been a difficulty in testing this hypothesis mainly because these individual fixed effects under the random effects model cannot be obtained. However, I extend the approach of Huang et al. (2019), then the individual fixed effects model can be estimated even under the random effects. With this estimation, I construct a test using the residuals from the proposed estimation strategy. Using this test, I also apply to the empirical data analyzing the productivity of the public capital in the United States.

Chapter 2

Systematic Development of Nonparametric Testing for Linearity: Small and Large Sample Properties

2.1 Introduction

Testing for linearity is an important subject in model specification, as it can prevent an estimator from being inconsistent due to misspecification of a functional form. Since a nonparametric test has superiority over any parametric tests due to its consistency, nonparametric hypothesis testing methods have been extensively developed until now. However, analytical comparisons of the local power between nonparametric tests have not been made.

The elaboration of these tests' performance is meaningful in that it can be practical for empirical researchers in specifying the model and even propose a new path for nonparametric hypothesis testing. In this regard, this paper proposes a systematic approach in current nonparametric hypothesis testing.

There are two main contributions of this paper. First, this is the first paper that categorizes the four popular nonparametric tests under the same hypotheses— Li-Wang type conditional moment test, Su-Ullah type goodness-of-fit test, Yao-Ullah type goodness-of-fit test, and F-type test. Since these tests are based on the kernel sum of squared residuals, I can compare them analytically. Since the currently developed tests are ad-hoc, this systematic development can propose a new path for the development of nonparametric hypothesis testing by providing the analytics of these tests. Second, I analyze how the relationship between these four tests is shown in their asymptotic power and different depending on the estimation method in simulations. I suggest under which conditions this inequality holds. Based on both the analytical and numerical results, I show which nonparametric test is locally most powerful for each estimation method, which can be useful for empirical researchers testing for the correct model specification.

Since the nonparametric methods have been introduced in econometrics, there has been an extensive evolution of nonparametric tests ranging from a distributional test to testing for serial correlation (Whang and Andrews (1993), Fan and Ullah (1999), Fan et al. (2001), Hsiao and Li (2001), Horowitz and Spokoiny (2001), Linton et al. (2005), Su and Ullah (2008), Mishra et al. (2010), Su et al. (2013), among others). Among many different nonparametric hypothesis tests, I focus on four nonparametric tests, which use the kernel sum of squared residuals. These four tests are Li-Wang type conditional moment test, Ullah-type F test, Su-Ullah type goodness-of-fit test, and Yao-Ullah type goodness-of-fit test (See Ullah (1985), Fan and Li (2002), Li and Wang (1998), Su and Ullah (2013), and Yao and Ullah (2013)).

The Li-Wang type test is based on the moment condition of interest. It has an advantage for is simplicity among nonparametric tests in that it only requires the null hypothesis. For example, the parametric specification will only be needed to test for the parametric specification against the nonparametric specification. Since this requirement facilitates the procedure of constructing a test, Li-Wang type test has been used in testing a wide range of null hypotheses (See Hidalgo (1995), Zheng (1996), Lavergne and Vuong (2000), Hsiao and Li (2001), Henderson et al. (2008) among others). When constructing a test statistic, the conditional moment is converted to another one so that the test statistic can be derived for the second-order U-statistics.

The second type is the goodness-of-fit test, which corresponds to Su-Ullah type and Yao-Ullah type tests. The goodness-of-fit in the model specification is constructed using the ANOVA decomposition (See Doksum and Samarov (1995) and Huang and Chen (2008) among others), which contains the residual sum of squares. The only difference between the Su-Ullah type and Yao-Ullah type is that the former uses the local ANOVA decomposition while the latter uses the global ANOVA decomposition. An interesting approach with the goodness-of-fit is that the goodness-of-fit will be zero under the null hypothesis. The F-type test and goodness-of-fit test are similar in that these types require the estimation both under the null and the alternative hypothesis.

The third type is the F-type test. It is first proposed by Ullah (1988), and then was extended by Fan and Li (2002). The form of this F-statistic is similar to the parametric version of the F-test, but it is different in that it does not contain the degrees of freedom due to nonparametric estimation. The null hypothesis in this type of test is that the difference between the sum of parametric residuals and the sum of nonparametric residuals is equal to zero. Yatchew (1992) develops the F-type test using the sample splitting.

In addition to these four types of nonparametric tests, I construct a nonparametric Rao-Score (RS) test using two steps. Using the local log-likelihood function, I construct a local RS and then I set up the global RS by integrating the local test statistics over the support of a variable of interest. Once obtaining the global RS, I construct a fully nonparametric RS test. This type of test has not been developed in the literature, but I show the equivalence in test statistics between a nonparametric RS test and Su-Ullah type test asymptotically.

The evaluation of different test statistics has been accumulated over time in a parametric context. For example, Ullah and Zinde-Walsh (1984) showed the relationship between F, Wald, Lagrange Multiplier, and Likelihood Ratio tests, where they provided a way of developing a test statistic for any given problem based on the principles of the test. Unlike parametric hypothesis testing, any systematic development among nonparametric tests in hypothesis testing has not been done much due to its complexity. By extending the tests of Fan et al. (2001) and Fan and Yao (2008), Hong and Lee (2013) showed the approximate relationship between the F-type test and the general likelihood ratio test. Also, Su and Ullah (2013) compared the asymptotic local power between the method of moment test and their goodness-of-fit test. In this regard, I generalize the relationship of different nonparametric tests asymptotically by investigating their alternatives or modifications and present the inequality between them by their asymptotic local power.

The structure of this paper is as follows. Sections 2.2-2.3 introduce model and hypotheses as well as the examples of nonparametric test statistics. After comparing the local power of each test statistic asymptotically in Section 2.4, I present the simulation results of the different test statistics in testing for linearity with the bootstrap procedure in Section 2.5. Section 2.6 concludes the paper.

2.2 Model and Hypotheses

In this section, the general model will be discussed for testing linearity. Based on this general model, I introduce different types of kernel-based nonparametric tests using the kernel sum of squared residuals, and show how they are linked to each other asymptotically. Consider the model as follows:

$$y_i = m(x_i) + u_i \tag{2.1}$$

for $i = 1, \ldots, n$, where y_i is an observable scalar random variable and x_i' is a $p \times 1$ vector of regressors with the unknown functions $m : \mathbb{R}^p \to \mathbb{R}^1$. All the variables are *i.i.d.* over i. u_i is disturbance term such that $E[u_i \mid x_i] = 0$ is satisfied. Based on this model, The hypotheses for linearity are as follows.

$$\begin{cases} \mathbb{H}_0: & m(x_i) = E[y_i \mid x_i] = x_i\beta \\ \mathbb{H}_1: & m(x_i) = E[y_i \mid x_i] \neq x_i\beta \end{cases}$$

The model (2.1) can be re-written by accommodating the given null hypothesis, which allows constructing a goodness-of-fit test.

$$
\begin{aligned}
y_i &= m(x_i) + u_i \\
y_i - x_i\beta &= m(x_i) - x_i\beta + u_i
\end{aligned} \tag{2.2}
$$

Using the model (2.2), the ANOVA decomposition of variance following Doksum and Samarov (1995) is given as $E[(y_i - x_i\beta)^2] = E[(m(x_i) - x_i\beta)^2] + E[u_i^2]$. Under the null hypothesis, the second term becomes zero if the null is true. Following the given model, the goodness-of-fit R^2 can be constructed as $R^2 = E[(m(x_i) - x_i\beta)^2]/E[(y_i - x_i\beta)^2]$. Then another set of hypotheses for testing linearity can be set up.

$$
\begin{cases}
\mathbb{H}_0: & R^2 = 0 \\
\mathbb{H}_1: & R^2 \neq 0
\end{cases}
$$

These hypotheses will used for the Su-Ullah type and the Yao-Ullah type tests. When $R^2 = 0$, it implies that the model does not explain any variation of y, which suggests that the model is linear.

For the asymptotic properties of each test, the assumptions will be characterized as follows.

Assumptions

(A1) $\{y_i, x_i\}_{i=1}^n$ is independently and identically distributed.

(A2) $0 < V(y) < \infty$.

(A3) The marginal density $f(x)$ is differentiable, $0 < f(x) \leq B_f < \infty,$ and $|f(x) - f(x')| < m_f\,|x - x'|$ for some $0 < m_f < \infty$ is satisfied.

(A4) Let $\sigma^2(x) = E\left[u^2 \middle|\, x\right]$, and $\sigma^4(x) = E\left[u^4 \middle|\, x\right]$. $\sigma^2(x)$ is continuous at x, $E[\sigma^2(x)] < \infty$, and $E[\sigma^4(x)] < \infty$. In addition, both $E[\sigma^2(x)]$ and $E[\sigma^4(x)]$ satisfy the Lipschitz-type condition, where $|\mu(x+u) - \mu(x)| \leq \vartheta(x)\,\|u\|$ with $E[\vartheta^2(x)] < \infty$ for $\mu = \sigma^2$, and σ^4.

(A5) i) $\hat{\beta} - \beta = O(n^{-1/2})$ under $\mathbb{H}_0$. ii) $m(x)$ is twice differentiable in $\mathbb{R}$ such that $m^{(1)}(x)$ and $m^{(2)}(x)$ are continuous, and dominated by functions with finite second moments, respectively.

(A6) The kernel function $K(\cdot)$ is continuous, symmetric, and bounded with $\int K(x)dx = 1$. For $\forall x \in \mathbb{R}\,,\ |K(x)| < B_k < \infty$. I assume $\left|K^j(u) - K^j(v)\right| \leq C_1\,|u - v|\,,$ for $j = 0, 1, 2, 3$.

(A7) As $n \to \infty$, $h \to 0,\ nh^p \to \infty,\ nh^{3p/2} \to \infty$, $nh^{p+2}/(\log n)^2 \to c \in (0, \infty]$.

The model assumes the *i.i.d.* distribution of $\{y_i, x_i\}_{i=1}^n$. By (A3), the conditional variance $\sigma^2(x)$ is continuous at x and its expectation is finite. (A2) is required to construct the globally constructed goodness-of-fit test (Yao-Ullah type). In (A3), the conditional density $f(x)$ satisfies the Lipschitz continuous condition. (A5)-ii) is the standard assumptions for nonparametric estimation, which are required for all the tests except the conditional moment test (Li-Wang type). In constructing all types of test, the kernel function is used as a weighting function. Regarding properties of the kernel function, it is bounded and symmetric by (A6). As in $f(x)$, the kernel function satisfies the Lipschitz continuous function. (A7) is on the restriction of the bandwidth.

2.3 Examples of Nonparametric Test Statistics

Based on the suggested models (2.1) and (2.2), I present four types of nonparametric tests—Conditional moment test (Li-Wang type), locally constructed goodness-of-fit test (Su-Ullah type), globally constructed goodness-of-fit test (Yao-Ullah type), and F-type test.

2.3.1 Conditional Moment Test (Li-Wang Type)

The conditional moment test is first developed by Zheng (1996), and widely used in different contexts (See Li and Wang (1998), Hsiao and Li (2001), and Henderson et al. (2008) among others). As discussed earlier, the advantage in constructing this type of test lies in its simplicity that the test required only the null hypothesis.

Under $\mathbb{H}_0$, there is no misspecification, and it contains the moment condition, which becomes $E[u_i \mid x_i] = 0$, where $u_i = y_i - x_i\beta$. Instead of directly testing the moment condition, the converted condition will be used for constructing a test statistic, which is $E[u_i E[u_i \mid x_i] f(x_i)] = 0$. The equivalence of two moment conditions is simply derived.

$$\begin{aligned} E\left[f\left(x_i\right)u_i E[u_i|x_i]\right] &= \iint u_1 \left(\int u_2 f\left(x, u_2\right) du_2\right) f\left(x, u_1\right) du_1 dx \\ &= \int \left(\int u_1 f\left(u_1|x\right) du_1\right)\left(\int u_2 f\left(u_2|x\right) du_2\right) f^2\left(x\right) dx \\ &= \int \left(\int u f\left(u|x\right) du\right)^2 f^2\left(x\right) dx, \end{aligned}$$

since u_i is *i.i.d.* over i. Therefore, $E[u_i|x_i] = \int u f\left(u|x\right) du = 0$ iff $E[f\left(x_i\right) u_i E[u_i|x_i]] = 0$ since $f\left(x\right) > 0$. ∎

Define $\hat{f}(x_i) = \frac{1}{(n-1)h^p} \sum_{j \neq i}^{n} K\left(\frac{x_j - x_i}{h}\right)$ (See Rosenblatt (1956)). Then, the sample analogue of the moment condition, I_n, and its standardized test statistic J_n are proposed as follows.

$$\begin{aligned} I_n &= \hat{E}[\hat{u}_i \hat{E}[\hat{u}_i \mid x_i] \hat{f}(x_i)] \\ &= \frac{1}{n} \sum_{i=1}^{n} \hat{u}_i \left(\frac{1}{(n-1)h^p} \sum_{j \neq i}^{n} \hat{u}_j K\left(\frac{x_j - x_i}{h}\right) \right) \\ &= \frac{1}{n(n-1)h^p} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{u}_i \hat{u}_j K\left(\frac{x_j - x_i}{h}\right), \end{aligned}$$

where $\hat{u}_i = y_i - \hat{m}(x_i)$. Note that $\hat{u}_i = y_i - x_i\hat{\beta}$ under the null.

Define $\hat{\Omega}_{J_n} = \frac{2}{nh^p} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{u}_i^2 \hat{u}_j^2 K\left(\frac{x_j - x_i}{h}\right)$. Then,

$$J_n = \sqrt{n^2 h^p} I_n / \hat{\Omega}_{J_n} \sim N(0,1)$$

Following Zheng (1996) and Li and Wang (1998), J_n asymptotically follows $N(0,1)$ with the consistency of $\hat{\Omega}_{J_n}$.

2.3.2 Locally Constructed Goodness-Of-Fit Test (Su-Ullah Type)

This test is initially proposed by Su and Ullah (2013) for testing heteroskedasticity. The estimation strategy for this type is weighted least squares (WLS) with using kernel weighting. The estimation is done at the local x, which allows obtaining local estimators $\delta(x)$.

$$\begin{aligned} y_i - x_i\hat{\beta} &= m(x_i) - x_i\hat{\beta} + \hat{u}_i \\ \varepsilon_i &= X_i(x)\hat{\delta}(x) + \hat{u}_i \end{aligned} \tag{2.3}$$

Note that $\hat{\delta}(x)$ becomes a local constant estimator if $X_i(x) = 1$, and $\hat{\delta}(x)$ becomes a local linear estimator if $X_i(x) = (1, x_i - x)'$. Then, the local R^2 will be calculated in this setting. As the first step, the local ANOVA (Analysis of Variance) decomposition can be derived from estimated model.

$$TSS(x) = ESS(x) + RSS(x)$$

$$\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2 K\left(\frac{x_i - x}{h}\right) = \sum_{i=1}^{n}(X_i(x)\hat{\boldsymbol{\delta}}(x))^2 K\left(\frac{x_i - x}{h}\right) + \sum_{i=1}^{n}\hat{u}_i^2 K\left(\frac{x_i - x}{h}\right)$$

The local goodness-of-fit $\hat{R}^2(x)$ can be obtained.

$$\hat{R}^2(x) = 1 - \frac{RSS(x)}{TSS(x)} = \frac{ESS(x)}{TSS(x)} = \frac{\sum_{i=1}^{n}(X_i(x)\hat{\delta}(x))^2 K\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2 K\left(\frac{x_i - x}{h}\right)}$$

Once the local $\hat{R}^2(x)$ is constructed, the second step is to obtain the global $\hat{R}^2$ by integrating the local ANOVA decomposition over χ_n, a compact subset of the support of the probability distribution function of x. The advantage of using a local ANOVA decomposition over the global ANOVA decomposition is that the probability of the global $\hat{R}^2$ being negative is zero.

$$\int_{\chi_n} TSS(x)dx = \int_{\chi_n} ESS(x)dx + \int_{\chi_n} RSS(x)dx$$

$$TSS = ESS + RSS$$

Define $W_x = diag(K_{1x}, ..., K_{nx})$. At the global level, $H^* = \int_{\chi_n} H_x dx$, where $H_x = W_x X_i(x)(X_i'(x)W_x X_i(x))^{-1} X_i'(x)W_x$.

The Su-Ullah type test statistic is then constructed by standardizing the estimated R^2 with its bias adjustment term and the variance.

$$T_n = \frac{\sqrt{n^2 h^p}\hat{R}^2 - \hat{B}_n}{\sqrt{\hat{\Omega}_{Tn}/(n^{-1}TSS)^2}},$$

where $\hat{B}_n \equiv h^{p/2}\sum_{i=1}^n \hat{u}_i^2 H_{ii}^*/(n^{-1}TSS)$ and $\hat{\Omega}_{Tn} \equiv 2n^{-1}h^p\sum_{i=1}^n \sum_{j\neq i}^n \hat{u}_i^2\hat{u}_j^2(nH_{ij}^*)^2$. Following the similar proofs of Su and Ullah (2013), T_n asymptotically follows $N(0,1)$.

Rao-Score Test

Before introducing the next type of the test, I will propose a nonparametric Rao-Score test and show its equivalence to Su-Ullah type test. Then, the new hypotheses under the new model will be given as below.

$$\begin{cases} \mathbb{H}_0 : \delta(x) &= 0 \\ \mathbb{H}_1 : \delta(x) \neq 0 \end{cases}$$

Under the null hypothesis, the model is correctly specified. Under the alternative, the model is incorrect. Consider a local log-likelihood function given in Fan et al. (2001).

$$l(\delta(x), \sigma^2(x); h, x) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\sigma^2(x) - \frac{1}{2\sigma^2(x)h^p}(\hat{\varepsilon} - X(x)\delta(x))'W_x(\hat{\varepsilon} - X(x)\delta(x))$$

The score function is given as

$$\begin{aligned} S(\delta(x)) &= \frac{\partial(\delta(x), \sigma^2(x); h, x)}{\partial\delta(x)} = \frac{2}{2\sigma^2(x)h^p}X_(x)'W_x(\hat{\varepsilon} - X(x)\delta(x)) \\ S(\sigma^2(x)) &= \frac{\partial(\delta(x), \sigma^2(x); h, x)}{\partial\sigma^2(x)} = -\frac{n}{2}\frac{1}{\sigma^2(x)} + \frac{1}{2\sigma^4(x)h^p}(\hat{\varepsilon} - X(x)\delta(x))'W_x(\hat{\varepsilon} - X(x)\delta(x)). \end{aligned}$$

Based on the score functions, the second-order derivative of the local log-likelihood function with respect to $\delta(x)$ is

$$\frac{\partial S(\delta(x))}{\partial \delta(x)} = \frac{\partial^2(\delta(x), \sigma^2(x); h, x)}{\partial \delta(x) \partial \delta(x)'} = -\frac{1}{\sigma^2(x) h^p} X(x)' W_x X(x)$$

Note that the conditional variance under the null $\tilde{\sigma}^2(x)$ is obtained by the first-order condition as follows:

$$\tilde{\sigma}^2(x) = \frac{(\hat{\varepsilon} - X(x)\delta(x))' W_x (\hat{\varepsilon} - X(x)\delta(x))}{nh^p} \Bigg|_{H_0} = \frac{1}{nh^p} \hat{\varepsilon}' W_x \hat{\varepsilon}.$$

Based on the obtained score function and $\tilde{\sigma}^2(x)$, I can construct the local Rao-Score:

$$\begin{aligned}
\widehat{RS}(x) &= (S(\delta(x)) \mid_{\mathbb{H}_0})' \left[I^{-1}(\delta(x)) \mid_{\mathbb{H}_0} \right] (S(\delta(x)) \mid_{\mathbb{H}_0})' \\
&= (S(\delta(x)) \mid_{\mathbb{H}_0})' \left(\frac{\partial S(\delta(x))}{\partial \delta(x)} \Bigg|_{\mathbb{H}_0} \right)^{-1} (S(\delta(x)) \mid_{\mathbb{H}_0}) \\
&= \left(\frac{1}{\tilde{\sigma}^2(x) h^p} X(x)' W_x \hat{\varepsilon} \right)' \left(\frac{1}{\tilde{\sigma}^2(x) h^p} X(x)' W_x X(x) \right)^{-1} \left(\frac{1}{\tilde{\sigma}^2(x) h^p} X(x)' W_x \hat{\varepsilon} \right) \\
&= \frac{1}{h^p} \frac{\hat{\varepsilon}' W_x X(x) (X(x)' W_x X(x))^{-1} X(x)' W_x \hat{\varepsilon}}{\tilde{\sigma}^2(x)} \\
&= \frac{1}{h^p} \frac{\hat{\varepsilon}' H_x \hat{\varepsilon}}{\frac{1}{nh^p} \hat{\varepsilon}' W_x \hat{\varepsilon}} \\
&= n \frac{\hat{\varepsilon}' H_x \hat{\varepsilon}}{\hat{\varepsilon}' W_x \hat{\varepsilon}}
\end{aligned}$$

As $\hat{R}^2(x) = \sum_{i=1}^n (X_i(x)\hat{\delta}(x))^2 K\left(\frac{x_i - x}{h}\right) / \sum_{i=1}^n (y_i - x_i\hat{\beta})^2 K\left(\frac{x_i - x}{h}\right)$, it is easy to show that $\widehat{RS}(x) = n\hat{R}^2(x)$ at the local level.

Once constructing a local measure, the global Rao-Score can be constructed by integrating the local measures over χ_n. Using $\tilde{\sigma}^2 = \frac{1}{nh^p}\hat{\varepsilon}'W\hat{\varepsilon}$,

$$\begin{aligned}
\widehat{RS} &= \int_{\chi_n} \frac{1}{h^p}\frac{\hat{\varepsilon}'H_x\hat{\varepsilon}}{\tilde{\sigma}^2(x)}\,dx \\
&= \int_{\chi_n} \frac{1}{h^p}\frac{\hat{\varepsilon}'H_x\hat{\varepsilon}}{\frac{1}{nh^p}\hat{\varepsilon}'W\hat{\varepsilon}}\,dx \\
&= n\int_{\chi_n} \frac{\hat{\varepsilon}'H_x\hat{\varepsilon}}{\hat{\varepsilon}'W\hat{\varepsilon}}\,dx \\
&= n\frac{\hat{\varepsilon}'H\hat{\varepsilon}}{\hat{\varepsilon}'W\hat{\varepsilon}}
\end{aligned}$$

The difference between the global goodness-of-fit and the global Rao-Score is that the global goodness-of-fit is established by integrating the ESS_x and TSS_x respectively. On the other hand, the global Rao-Score is constructed by integrating the $n(ESS_x/TSS_x)$. I can show in Theorem 1 that both global measures are equivalent. The proof of Theorem 1 is given in Appendix A.

Theorem 1 *As* $n \to \infty$, $\int_{\chi_n}\left(\hat{\varepsilon}'H_x\hat{\varepsilon}/(n^{-1}\hat{\varepsilon}W_x\hat{\varepsilon})\right)dx = \int_{\chi_n}\hat{\varepsilon}'W_x\hat{\varepsilon}\ dx/(n^{-1}\int_{\chi_n}\hat{\varepsilon}W_x\hat{\varepsilon}dx)$.

Using Theorem 1, asymptotically at the global level,

$$\widehat{RS} = n\hat{R}^2$$

To derive the asymptotic distribution of Rao-score Type test, I modify the global Rao-Score following the similar proofs from Theorem 5 in Fan et al. (2001).

Then, the standardized Rao-Score test statistic S_n can be derived. Define $\hat{\mu}_n = \sum_{i=1}^{n} \hat{u}_i^2 H_{ii}^* / (n^{-1}TSS) = h^{-p/2}\hat{B}_n$ and $\hat{\Gamma}_n = 2n^{-2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \hat{u}_i^2\hat{u}_j^2 (nH_{ij}^*)^2 = h^{-p}\hat{\Omega}_n$.

$$
\begin{aligned}
S_n &= \frac{nR^2 - \hat{\mu}_n}{\sqrt{\hat{\Gamma}_n/(n^{-1}TSS)^2}} \\
&= \frac{nR^2 - \hat{\mu}_n}{\sqrt{\hat{\Gamma}_n/(n^{-1}TSS)^2}} \sim N(0,1)
\end{aligned}
$$

The modification of the test statistic S_n gives the equivalence to the test statistic T_n.

$$
\begin{aligned}
S_n &= \frac{h^{-p/2}(nh^{p/2}R^2 - h^{p/2}\hat{\mu}_n)}{\sqrt{h^{-p}h^{p}\hat{\Gamma}_n/(n^{-1}TSS)^2}} \\
&= \frac{h^{-p/2}}{\sqrt{h^{-p}}} T_n \\
&= T_n
\end{aligned}
$$

Therefore, even though both test statistics are constructed from the weighted least squares and the local log-likelihood function respectively, they are equivalent and follow the standard normal distribution asymptotically.

2.3.3 Globally Constructed Goodness-Of-Fit Test (Yao-Ullah Type)

This type of test is constructed by the global ANOVA decomposition, which was proposed by Yao and Ullah (2013) for testing relevant variables in the model. Different from the locally constructed goodness-of-fit test, this test is directly constructed from the global R_G^2. For testing linearity, the global R_G^2 and $\hat{R}_G^2$ are derived as below. Note that R_G^2 is always between 0 and 1, but $\hat{R}_G^2$ may not lie between 0 and 1. Given the ANOVA decomposition,

$E[(y_i - x_i\hat{\beta})^2] = E[(\hat{m}(x_i) - x_i\hat{\beta})^2] + E[(y_i - \hat{m}(x_i))^2] + 2E[(y_i - \hat{m}(x_i))] \cdot E[\hat{m}(x_i) - x_i\hat{\beta}]$. But, $2\sum_{i=1}^{n}[(y_i - \hat{m}(x_i))] \cdot E[\hat{m}(x_i)]$ might not be equal to zero. Due to this potential issue, it is necessary to impose a condition in constructing the test statistic.

$$
\begin{aligned}
R_G^2 &= 1 - \frac{E\left[u_i^2\right]}{E\left[(y_i - x_i\hat{\beta})^2\right]} = 1 - \frac{E\left[(y_i - m(x_i))^2\right]}{E\left[(y_i - x_i\hat{\beta})^2\right]} \\
\hat{R}_G^2 &= 1 - \frac{E\left[\hat{u}_i^2\right]}{E\left[(y_i - x_i\hat{\beta})^2\right]} = 1 - \frac{E\left[(y_i - \hat{m}(x_i))^2\right]}{E\left[(y_i - x_i\hat{\beta})^2\right]}.
\end{aligned}
$$

Define $\hat{A}_n = \frac{1}{n^3 h^{2p}} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \frac{\hat{u}_i^2}{\hat{f}(x_i)^2} K^2\left(\frac{x_j - x_i}{h}\right)$ [1]. The estimated goodness-of-fit and test statistic are given as follows.

$$
\begin{aligned}
\hat{R}_G^2 &= \left[1 - \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{m}(x_i))^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2}\right] \cdot I\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2 \geq \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{m}(x_i))^2\right) \\
T_{Gn} &= \frac{nh^{p/2}}{\sqrt{\hat{V}_G}}\left\{\hat{R}_G^2 + I\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2 \geq \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{m}(x_i))^2\right)\frac{\hat{A}_n}{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2}\right\} \\
&\sim N(0,1)
\end{aligned}
$$

Note that second term of T_{Gn} is bias adjustment term for T_{Gn}, and it is defined differently depending on the estimation method. Also, $\hat{V}_G$ is shown as

$$
\hat{V}_G = \frac{\hat{\sigma}_{\phi G}^2}{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2},
$$

where $\hat{\sigma}_{\phi G}^2 = \left[\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}^{n} K\left(\frac{x_j - x_i}{h}\right)\frac{\hat{u}_i^2\hat{u}_j^2}{h^p \hat{f}(x_i)^2}\right]\left(\int 2(2K(\psi) - \kappa(\psi))^2 d\psi\right)$, and $\kappa(\psi)$ is two-fold convolution kernel of $K(\cdot)$.

[1] In Yao and Ullah (2013), there is an additional bias adjustment term, but it is cancelled out because the estimation in this setting is leave-one-out estimation.

2.3.4 F-Type Test

Ullah (1985) initially introduced a nonparametric F-type test by using the kernel squared sum of residuals. In Ullah (1985), the following statistic U_n was proposed, which has a similar form to the parametric F-test.

$$U_n = \frac{RSS_1 - RSS_0}{RSS_0} = \frac{\sum_{i=1}^{n}(y_i - \hat{m}(x_i))^2 - \sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2}{\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2}$$

Based on this test statistic, Fan and Li (2002) derived asymptotic normality of U_n and developed the new test statistic F_n.

Define $\hat{\Omega}_{Fn} = \frac{1}{n^3h^{2p}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{\hat{f}_i^2(x_i)}\hat{u}_i^2\hat{u}_j^2(\kappa\left(\psi\right) - 2K(\frac{x_j-x_i}{h}))^2$, B_{Fn} is a bias adjustment term, where $B_{Fn} = \frac{1}{n^3h^{2p}}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{\hat{u}_i^2}{\hat{f}(x_i)^2}K^2(\frac{x_j-x_i}{h})$, and $\hat{u}_i = y_i - x_i\hat{\beta}$ under the null.

$$F_n = nh^{p/2}\frac{U_n + B_{Fn}}{\sqrt{\hat{\Omega}_{Fn}}} \sim N(0,1)$$

By Fan and Li (2002), F_n follows the standard normal distribution under the null.

2.4 Asymptotic Local Power Properties

With the test statistics from the Section 2.3, I focus on deriving the relationship of the given four test statistics—Li-Wang type test, Su-Ullah type goodness-of-fit test, Yao-Ullah goodness-of-fit test, and F-type test. Even though it is difficult to compare these test statistics, it is feasible to directly compare the test statistics asymptotically to analyze the asymptotic local power of each test statistic. When the conditions are imposed as below, then the inequality among the test statistics can be obtained.

For the asymptotic properties under the alternative, I introduce the Pitman local alternatives as follows:

$$H_1(\delta_n) : m_1(x_i) = x_i\beta + \delta_n l(x_i)\,,$$

where $l(\cdot)$ is continuously differentiable and bounded, and $\delta_n = n^{-1/2}h^{-p/4}$. Due to their complicated form, two conditions will be imposed for comparing the asymptotic local power of the test statistics.

(B1) $E\left[l(x_i)\right] = 0$.

(B2) X is uniformly distributed.

If (B1) holds, this means that $V\left(m_1(x_i)\right) = E[(m_1(x_i))^2]$. Also, the uniformly distribution of x_i implies that $f(x_i) = 1/vol(\chi)$. In the following sections 2.4.1-2.4.4, the asymptotic mean and the variance will be shown. In all calculations, the Gaussian kernel is used and $p = 1$ for the simplicity.[2]

2.4.1 Method of Moment Test Statistic

With the two conditions above, the test statistic can be simplified as below. I will present the mean, the variance, and the test statistic in order.

$$\begin{aligned} E\left[nh^{1/2}I_n\right] &= \lim_{n\to\infty} E[l^2(x_i)f(x_i)] \\ &= \frac{1}{vol(X)} \lim_{n\to\infty} E[l^2(x_i)] \end{aligned}$$

[2]The detailed derivation for each type of the test in this section will be given in Appendix A.

$$
\begin{aligned}
\Omega_{J_n} &= 2 \lim_{n\to\infty} E[\sigma^4(x_i) f^2(x_i)] \int K^2(u)du \\
&= \frac{2}{(vol(X))^2} \lim_{n\to\infty} E[\sigma^4(x_i)] \int K^2(u)du \\
&= \frac{1}{2\sqrt{\pi}} \cdot \frac{2}{(vol(X))^2} \lim_{n\to\infty} E[\sigma^4(x_i)] \\
&\simeq 0.2821 \frac{2}{(vol(X))^2} \lim_{n\to\infty} E[\sigma^4(x_i)]
\end{aligned}
$$

Define $\Gamma = \lim_{n\to\infty} E[l^2(x_i)]/\sqrt{2}\lim_{n\to\infty} E[\sigma^4(x_i)]$. Based on the above information,

$$
\begin{aligned}
J_n &= \frac{nh^{1/2} I_n}{\sqrt{\Omega_{J_n}}} \\
&\simeq \frac{1}{\sqrt{0.2821}} \lim_{n\to\infty} E[l^2(x_i)]/\sqrt{2} \lim_{n\to\infty} E[\sigma^4(x_i)] \\
&= 1.882\Gamma
\end{aligned}
$$

2.4.2 Locally Constructed R^2 Test Statistic

Likewise, the asymptotic mean and the asymptotic variance will be calculated. Define $\sigma_\varepsilon^2 = E[(y_i - x_i\beta)^2]$,and $\mu_q(x)$ refers to the stack of $x^j, 0 \leq |j| \leq q$ in the lexicographical order. The local constant least squares estimation is when $q = 0$, and the local linear least squares estimation is when $q = 1$.

$$
E[nh^{1/2}\hat{R}^2 - \hat{B}_n] = \lim_{n\to\infty} E\left[l^2(x_i)\right]/\sigma_\varepsilon^2
$$

$$
\Omega_{T_n,q} = 2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^2 \int (\int K(z)\mu_q(z)'\mu_q(z+x)K(z+x)dz)^2 dx
$$

Note that the value of Ω_{T_n} [3] changes depending on the estimation method, and I present the variance when the local constant least squares estimation ($T_{n,0}$) and the local linear least squares estimation ($T_{n,1}$) are applied respectively.

$$
\begin{aligned}
\Omega_{T_n,0} &= 2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^2 \int (\int K(z)\mu_0(z)'\mu_0(z+x)K(z+x)dz)^2 dx \\
&= 2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^2 \int \kappa^2(u)du \\
&= \frac{1}{2\sqrt{2\pi}} \cdot 2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^2 \\
&\simeq 0.1995 \cdot 2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^2
\end{aligned}
$$

$$
\begin{aligned}
\Omega_{T_n,1} &= 2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^2 \int \left(\int K(z)\mu_1(z)'\mu_1(z+x)K(z+x)dz\right)^2 dx \\
&= 2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^2 \int \left(\int \kappa(x)(1+z(z+x))dz\right)^2 dx \\
&= \frac{27}{32\sqrt{2\pi}} \cdot 2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^2 \\
&\simeq 0.3366 \cdot 2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^2
\end{aligned}
$$

Then, there are two versions of the test based on the estimation methods.

$$
\begin{aligned}
T_{n,0} &= \frac{nh^{1/2}\hat{R}^2 - \hat{B}_n}{\sqrt{\Omega_{T_n,0}}} \\
&\simeq \frac{1}{\sqrt{0.1995}} \lim_{n\to\infty} E[l^2(x_i)]/\sqrt{2} \lim_{n\to\infty} E[\sigma^4(x_i)] \\
&= 2.238\Gamma
\end{aligned}
$$

[3] As the estimation methods used are the local constant and local linear estimations, the given variance formula is a simplified version of that in Su and Ullah (2013).

$$
\begin{aligned}
T_{n,1} &= \frac{nh^{1/2}\hat{R}^2 - \hat{B}_n}{\sqrt{\Omega_{T_n,1}}} \\
&\simeq \frac{1}{\sqrt{0.3366}} \lim_{n\to\infty} E[l^2(x_i)]/\sqrt{2} \lim_{n\to\infty} E[\sigma^4(x_i)] \\
&= 1.723\Gamma
\end{aligned}
$$

From this calculation, the power of the locally constructed R^2 test increases when using the local constant estimator due to its relatively smaller variance than that of the local linear estimator. This result has been also shown in Su and Ullah (2013).

2.4.3 Globally Constructed R^2 Test Statistic

Under the alternative, the probability of $I\left(\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2 \geq \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{m}(x_i))^2\right)$ becomes 1 as $n \to \infty$[4]. Then, the second term of $T_G n$ converges to zero.

$$
\begin{aligned}
E\left[nh^{1/2}\hat{R}_G^2\right] &= \lim_{n\to\infty} \frac{1}{\sigma_\varepsilon^2} E\left[\frac{l^2(x_i)}{f(x_i)}\right] \\
&= vol(X)\cdot \lim_{n\to\infty} E\left[l^2(x_i)\right]/\sigma_\varepsilon^2
\end{aligned}
$$

$$
\begin{aligned}
\Omega_{T_{Gn}} &= 2\lim_{n\to\infty}(vol(X))^2 E[\sigma^4(x_i)] \int (\kappa(u) - 2K(u))^2 du/\sigma_\varepsilon^4 \\
&= \frac{\sqrt{3}+4\sqrt{6}-8}{2\sqrt{6\pi}} \cdot 2(vol(X))^2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^4 \\
&\simeq 0.4065 \cdot 2(vol(X))^2 \lim_{n\to\infty} E[\sigma^4(x_i)]/\sigma_\varepsilon^4
\end{aligned}
$$

[4]Please refer to the proof of Theorem 3 in Yao and Ullah (2013).

Different from the locally constructed goodness-of-fit test (Su-Ullah type test), the variance formula stays the same when R^2 is globally constructed.

$$
\begin{aligned}
T_{Gn} &= \frac{nh^{1/2}\hat{R}_G^2}{\sqrt{\Omega_{T_{Gn}}}} \\
&\simeq \frac{1}{\sqrt{0.4065}} \lim_{n\to\infty} E[l^2(x_i)]/\sqrt{2} \lim_{n\to\infty} E[\sigma^4(x_i)] \\
&= 1.568\Gamma
\end{aligned}
$$

2.4.4 F-Type Test Statistic

The asymptotic mean and the variance of of F-Type test can be written as follows.

$$
\begin{aligned}
\mathrm{E}[nh^{1/2}(U_n + B_{Fn})] &= \mathrm{E}\left[\frac{l^2(x_i)}{f(x_i)}\right] \\
&= vol(X) \cdot \mathrm{E}[l^2(x_i)] \\
\Omega_{F_n} &= 2 \lim_{n\to\infty} \mathrm{E}\left[\frac{\sigma^4(x_i)}{f^2(x_i)}\right] \int (\kappa(u) - 2K(u))^2 du \\
&= 2(vol(X))^2 \lim_{n\to\infty} \mathrm{E}[\sigma^4(x_i)] \int (\kappa(u) - 2K(u))^2 du \\
&\simeq 0.4065 \cdot 2(vol(X))^2 \lim_{n\to\infty} \mathrm{E}[\sigma^4(x_i)]
\end{aligned}
$$

As $P\left(I\left(\frac{1}{n}\sum_{i=1}^n \left(y_i - x_i\hat{\beta}\right)^2 \geq \frac{1}{n}\sum_{i=1}^n (y_i - \hat{m}(x_i))^2\right)\right) \xrightarrow{p} 1$ asymptotically under the alternative, the F-type test becomes identical to the globally constructed R^2 test.

$$
\begin{aligned}
F_n = \frac{nh^{1/2}(U_n + B_{Fn})}{\sqrt{\Omega_{F_n}}} &\simeq \frac{1}{\sqrt{0.4065}} \lim_{n\to\infty} E[l^2(x_i)]/\sqrt{2} \lim_{n\to\infty} E[\sigma^4(x_i)] \\
= 1.568\Gamma
\end{aligned}
$$

2.4.5 Asymptotic Local Power

Depending on the estimation method, each test statistic is ranked based on its asymptotic local power given (B1) and (B2). For the local constant least squares estimation,

$$T_{Gn} \leq F_n < J_n < T_{n,0}$$

$$\iff 1.568\Gamma \leq 1.568\Gamma < 1.882\Gamma < 2.238\Gamma$$

The inequality between T_{Gn} and F_n is due to the indicator function term in T_{Gn}, which can increase the variance and make the test statistic decrease. For the local linear least squares estimation, the rank of the asymptotic local power between J_n and T_n is switched.

$$T_{Gn} \leq F_n < T_{n,1} < J_n$$

$$\iff 1.568\Gamma \leq 1.568\Gamma < 1.723\Gamma < 1.882\Gamma$$

Based on the asymptotic relationship, I can conclude that the locally constructed R^2 test is most powerful for the local constant estimation while the conditional method of moment test is most powerful for the local linear estimation.

2.5 Simulation

This section will show how the analytical results from the previous section match in the simulations for a small sample and large sample. In addition, I present the simulation results using a wild bootstrap procedure to improve the test performance in a finite sample.

2.5.1 Data Generating Processes

For calculating the size, I follow the data generating process of Hardle and Mammen (1993) and Li and Wang (1998), where $x_i \sim U[0,1]$ and $u_i \sim N(0, .1)$. DGP_1 is to estimate the size of the test under the null hypothesis. I also construct DGP_2 to estimate the power of each test by applying the same data generating process for x_i and u_i as DGP_1, but x_i is not linear in y_i anymore. Note that the uniform distribution of x_i is the key condition in comparing the test statistics.

$$DGP_1: \; y_i = 1 + x_i + u_i$$

$$DGP_2: \; y_i = x_i - x_i^2 + u_i$$

I consider two samples sizes, 100 with 1000 repetitions and 400 with 500 repetitions. Wild bootstrap is applied and its procedure will be introduced in the following session. The number of bootstrapping B is fixed as 399. I implement both local constant (LC) and local linear (LL) least-squares estimation. The Gaussian kernel is chosen for the estimation and test statistics. The rule-of-thumb bandwidths are used for x_i, $h = c \cdot std(x_i) \cdot n^{-1/5}$ by taking different values of $c = 0.5, \; 1, \;$ and 1.5. For the conditional moment test, as it is constructed under the null, using either LC or LL method does not affect the test statistic.

2.5.2 Bootstrap Procedure

To improve the test performance in a small sample, I propose a wild bootstrap test as an alternative. I apply a wild bootstrap method using Mammen's distribution. Steps to get a bootstrap test statistic are given below.

1. Estimate $\hat{m}(x_i) = x_i\hat{\beta}$, where $\hat{\beta}$ is the estimates from the linear regression under the null hypothesis.

2. Generate u_i^* as the wild bootstrap error. I construct $u_i^* = \frac{1-\sqrt{5}}{2}\hat{u}_i$ with the probability of $\frac{1+\sqrt{5}}{2}$ and $u_i^* = \frac{1+\sqrt{5}}{2}\hat{u}_i$ with the probability of $1 - \frac{1+\sqrt{5}}{2}$. It is easy to show $E\left[u_i^*\right] = 0$, $E\left[u_i^{*2}\right] = \hat{u}_i^2$, and $E\left[u_i^{*3}\right] = \hat{u}_i^3$.

3. Generate y_i^*, where $y_i^* = x_i\hat{\beta} + \hat{u}_i^*$.

4. Using the bootstrap sample $\{y_i^*, x_i\}_{i=1}^n$, regress y_i^* on x_i to obtain $\hat{m}^*(x_i) = x_i\hat{\beta}^*$, and get $\hat{u}_i^* = y_i^* - \hat{m}^*(x_i)$.

5. Compute the bootstrap test statistic and repeat above procedure for B times. In this simulation, the number of bootstrapping is 399.

6. Based on the empirical distribution of each test, calculate the critical value c^* and obtain the p-value. If p-value is less than 0.05 at 5% significance level, the null is rejected.

Following these bootstrap procedures, I can obtain the asymptotic distribution of each test. The results will be given in Table 2.1-2.4. Following the asymptotic properties of each test statistic, the bootstrap tests are consistent.

2.5.3 Simulation Results

The simulation results are given in Table 2.1-2.4 to compare the test performance by differing the estimation methods. The size performance is shown in Table 2.1 and Table 2.2. For Table 2.1, the size of each test under the local constant least squares estimation is presented for each sample size, bandwidth, and the significance level. While each bootstrap test seems oversized in a finite sample, it gets closer to the correct size in a large sample. For the Yao-Ullah type test, the size is undersized even in a bootstrap test when the bandwidth is large. This can be explained by the fact that the Yao-Ullah type test includes an indicator function to prevent the goodness-of-fit from being negative, which can happen as the local constant estimator gets closer to the parametric linear model with the increase in a bandwidth. However, each test's size performs well overall.

The size performance using a local linear estimator in Table 2.2 is similar to the results in Table 2.1. At each significance level, the estimated size is better in a large sample for all tests. In particular, the bootstrap size of the Yao-Ullah type test is recovered in a larger bandwidth. In addition, the noticeable result is that the asymptotic size is better when implementing a local linear estimation than implementing a local constant estimation even though the bootstrap size is still closer to the correct size.

More importantly, Table 2.3-2.4 presents the power of each test, which can verify the analytical comparison between tests in Section 2.4.5. Table 2.3 presents the power of each test using the local constant estimator. At all significance levels and all sample sizes, the Su-Ullah Type test is the most powerful, followed by the Li-Wang type test, the Yao-Ullah type test, and F-Type test. As the sample size increases, the difference in the power performance between the Su-Ullah type test and the Li-Wang type test becomes narrower. Also, asymptotically the F-Type and the Yao-Ullah test are equivalent. On the other hand, the bootstrap power of Yao-Ullah test is smaller because of the indicator function element in their test statistic. Both tests become almost identical with a larger sample size.

In Table 2.4, the power of each test when using the local linear estimator is shown. The interesting aspect of the test comparison is that the Li-Wang type test outperforms the Su-Ullah type test in both sample sizes with a smaller bandwidth. When the bandwidth increases, the power of the Su-Ullah type test is higher than that of Li-Wang type test. This can be because the estimation gets closer to the local constant estimator with the increase in bandwidth. This tendency is consistent for each sample size and the significance level.

In summary, I analyze the numerical relationship among four types of nonparametric tests. For all tests, the bootstrap size performs better than the asymptotic size as it is closer to the correct size. Each test performs almost equally well, and its performance varies with the bandwidth. However, there is a clear difference between the tests in estimating the power, and the power comparison between these test statistics matches with the analytical comparison in Section 2.4.5. One interesting result is that the Su-Ullah type test outperforms the Li-Wang type test in power as the bandwidth increases with the local linear least squares estimation. When using the local constant least squares estimation, the numerical comparison of the four tests is identical to the analytical comparison for any level of bandwidth.

2.6 Conclusion

Since the extensive development in nonparametric hypothesis tests, there has not been any systematic development in nonparametric hypothesis testing due to its complexity. Such

Table 2.1: Size of Each Test Using Local Constant Estimation

			Bootstrap Test			Asymptotic Test		
	α	Tests	$c = 0.5$	$c = 1.06$	$c = 1.5$	$c = 0.5$	$c = 1.06$	$c = 1.5$
$n = 100$	1%	J_n	0.008	0.010	0.012	0.001	0.000	0.000
		T_n	0.007	0.012	0.012	0.000	0.000	0.000
		T_{Gn}	0.008	0.014	0.013	0.001	0.002	0.000
		F_n	0.008	0.014	0.013	0.001	0.002	0.000
	5%	J_n	0.051	0.051	0.060	0.012	0.003	0.000
		T_n	0.052	0.050	0.061	0.010	0.000	0.000
		T_{Gn}	0.048	0.059	0.052	0.002	0.009	0.004
		F_n	0.048	0.059	0.058	0.002	0.009	0.004
	10%	J_n	0.101	0.107	0.113	0.024	0.008	0.002
		T_n	0.106	0.111	0.111	0.020	0.002	0.000
		T_{Gn}	0.103	0.112	0.067	0.007	0.018	0.013
		F_n	0.103	0.112	0.108	0.007	0.018	0.013
$n = 400$	1%	J_n	0.016	0.014	0.012	0.010	0.002	0.000
		T_n	0.014	0.010	0.010	0.004	0.000	0.000
		T_{Gn}	0.016	0.008	0.008	0.002	0.000	0.000
		F_n	0.016	0.008	0.008	0.002	0.000	0.000
	5%	J_n	0.050	0.034	0.042	0.020	0.008	0.002
		T_n	0.038	0.040	0.048	0.018	0.002	0.000
		T_{Gn}	0.044	0.048	0.022	0.004	0.006	0.000
		F_n	0.044	0.048	0.044	0.004	0.006	0.000
	10%	J_n	0.088	0.084	0.072	0.028	0.016	0.004
		T_n	0.080	0.078	0.084	0.022	0.004	0.000
		T_{Gn}	0.100	0.062	0.022	0.008	0.010	0.002
		F_n	0.100	0.084	0.078	0.008	0.010	0.002

Note: α denotes the significance level. The different level of bandwidth depends on the value of c, where $h = c \cdot std(x_i)n^{-1/5}$.

Table 2.2: Size of Each Test Using Local Linear Estimator

			Bootstrap			Asymptotic		
	α	Tests	$c = 0.5$	$c = 1.06$	$c = 1.5$	$c = 0.5$	$c = 1.06$	$c = 1.5$
$n = 100$	1%	J_n	0.008	0.010	0.012	0.001	0.000	0.000
		T_n	0.008	0.015	0.013	0.003	0.001	0.001
		T_{Gn}	0.008	0.011	0.012	0.003	0.012	0.020
		F_n	0.008	0.011	0.012	0.003	0.012	0.020
	5%	J_n	0.051	0.051	0.060	0.012	0.003	0.000
		T_n	0.047	0.054	0.062	0.014	0.013	0.005
		T_{Gn}	0.047	0.056	0.060	0.010	0.039	0.050
		F_n	0.047	0.056	0.060	0.010	0.039	0.050
	10%	J_n	0.101	0.107	0.113	0.024	0.008	0.002
		T_n	0.105	0.106	0.108	0.031	0.024	0.011
		T_{Gn}	0.107	0.110	0.108	0.022	0.062	0.084
		F_n	0.107	0.110	0.108	0.022	0.062	0.084
$n = 400$	1%	J_n	0.016	0.014	0.012	0.010	0.002	0.000
		T_n	0.014	0.006	0.010	0.010	0.004	0.000
		T_{Gn}	0.014	0.006	0.008	0.002	0.006	0.008
		F_n	0.014	0.006	0.008	0.002	0.006	0.008
	5%	J_n	0.050	0.034	0.042	0.020	0.008	0.002
		T_n	0.042	0.038	0.042	0.018	0.014	0.008
		T_{Gn}	0.052	0.044	0.044	0.010	0.024	0.030
		F_n	0.052	0.044	0.044	0.010	0.024	0.030
	10%	J_n	0.088	0.084	0.072	0.028	0.016	0.004
		T_n	0.096	0.080	0.080	0.036	0.022	0.012
		T_{Gn}	0.102	0.088	0.074	0.016	0.036	0.044
		F_n	0.102	0.088	0.074	0.016	0.036	0.044

Note: α denotes the significance level. The different level of bandwidth depends on the value of c, where $h = c \cdot std(x_i)n^{-1/5}$.

Table 2.3: Power of Each Test Using Local Constant Estimator

			Bootstrap			Asymptotic		
	α	Tests	$c = 0.5$	$c = 1.06$	$c = 1.5$	$c = 0.5$	$c = 1.06$	$c = 1.5$
$n = 100$	1%	J_n	0.319	0.414	0.470	0.239	0.222	0.155
		T_n	0.385	0.456	0.501	0.267	0.137	0.014
		T_{Gn}	0.309	0.421	0.381	0.137	0.277	0.192
		F_n	0.309	0.421	0.382	0.137	0.277	0.192
	5%	J_n	0.547	0.659	0.704	0.383	0.344	0.251
		T_n	0.617	0.688	0.729	0.400	0.235	0.067
		T_{Gn}	0.517	0.636	0.588	0.230	0.399	0.293
		F_n	0.517	0.636	0.596	0.230	0.399	0.293
	10%	J_n	0.662	0.757	0.799	0.464	0.423	0.334
		T_n	0.722	0.793	0.822	0.474	0.321	0.116
		T_{Gn}	0.653	0.735	0.642	0.307	0.479	0.364
		F_n	0.653	0.738	0.702	0.307	0.479	0.364
$n = 400$	1%	J_n	0.974	0.988	0.992	0.962	0.976	0.968
		T_n	0.986	0.992	0.992	0.974	0.968	0.932
		T_{Gn}	0.970	0.988	0.982	0.910	0.974	0.946
		F_n	0.970	0.988	0.982	0.910	0.974	0.946
	5%	J_n	0.990	0.998	0.998	0.984	0.990	0.984
		T_n	0.996	0.998	0.998	0.990	0.986	0.968
		T_{Gn}	0.986	0.996	0.990	0.942	0.986	0.974
		F_n	0.986	0.996	0.994	0.942	0.986	0.974
	10%	J_n	1.000	0.998	0.998	0.984	0.992	0.990
		T_n	1.000	0.998	0.998	0.990	0.990	0.980
		T_{Gn}	0.994	0.996	0.990	0.962	0.988	0.982
		F_n	0.994	0.998	0.996	0.962	0.988	0.982

Note: α denotes the significance level. The different level of bandwidth depends on the value of c, where $h = c \cdot std(x_i)n^{-1/5}$.

Table 2.4: Power of Each Test Using Local Linear Estimator

			Bootstrap			Asymptotic		
	α	Tests	$c=0.5$	$c=1.06$	$c=1.5$	$c=0.5$	$c=1.06$	$c=1.5$
$n=100$	1%	J_n	0.319	0.414	0.470	0.239	0.222	0.155
		T_n	0.304	0.440	0.508	0.235	0.326	0.299
		T_{Gn}	0.267	0.411	0.489	0.181	0.455	0.573
		F_n	0.267	0.411	0.489	0.181	0.455	0.573
	5%	J_n	0.547	0.659	0.704	0.383	0.344	0.251
		T_n	0.521	0.678	0.722	0.395	0.442	0.428
		T_{Gn}	0.486	0.655	0.712	0.306	0.617	0.708
		F_n	0.486	0.655	0.712	0.306	0.617	0.708
	10%	J_n	0.662	0.757	0.799	0.464	0.423	0.334
		T_n	0.645	0.777	0.826	0.474	0.530	0.510
		T_{Gn}	0.614	0.764	0.795	0.391	0.685	0.770
		F_n	0.614	0.764	0.795	0.391	0.685	0.770
$n=400$	1%	J_n	0.974	0.988	0.992	0.962	0.976	0.968
		T_n	0.962	0.990	0.992	0.960	0.984	0.986
		T_{Gn}	0.966	0.986	0.994	0.922	0.986	0.994
		F_n	0.966	0.986	0.994	0.922	0.986	0.994
	5%	J_n	0.990	0.998	0.998	0.984	0.990	0.984
		T_n	0.988	0.998	0.998	0.984	0.992	0.996
		T_{Gn}	0.986	0.998	0.998	0.960	0.994	0.996
		F_n	0.986	0.998	0.998	0.960	0.994	0.996
	10%	J_n	1.000	0.998	0.998	0.984	0.992	0.990
		T_n	0.998	0.998	0.998	0.988	0.996	0.996
		T_{Gn}	0.992	0.998	0.998	0.976	0.996	0.998
		F_n	0.992	0.998	0.998	0.976	0.996	0.998

Note: α denotes the significance level. The different level of bandwidth depends on the value of c, where $h = c \cdot std(x_i)n^{-1/5}$.

development is important because the most powerful test can be identified and even a new test can be proposed based on the obtained analytical results.

Among many different nonparametric hypothesis tests, I focus on four nonparametric tests—Li-Wang type conditional moment test, Su-Ullah type goodness-of-fit test, Yao-Ullah type goodness-of-fit test, and the F-type test—because they have in common in that they are based on residual sums of residuals using kernel weighting methods. Under some conditions, I found the inequality between these four tests in their asymptotic power and it becomes different depends on the estimation method.

In simulations, both the size and power are estimated for a small and large sample by differing the estimation methods. Overall, the bootstrap size is better than the asymptotic size for all tests. The size performance between the tests is almost equal to each other. For the power analysis, the numerical results match with the analytical results. At all significance levels and all sample sizes, the Su-Ullah Type test is the most powerful when the local constant estimator is used. When implementing the local linear estimation, the Li-Wang type test outperforms the Su-Ullah type test in both sample sizes with a smaller bandwidth.

To conclude, I propose a systematic development of a nonparametric hypothesis testing focusing on the tests by providing information about the asymptotic local power analysis. This systematic approach can propose a new path for nonparametric hypothesis testing in the future. Furthermore, it can be practically used for empirical researchers in conducting a model specification test for linearity.

Chapter 3

A Consistent Nonparametric Test for Endogeneity

3.1 Introduction

Endogeneity is commonly observed in many economic contexts. While assuming endogeneity by economic theory, econometricians have focused on developing consistent estimation methods to tackle endogeneity (See Card (2001), Miguel et al. (2004), and Coglianese et al. (2017) among others). However, variables can be exogenous in one setting, but endogenous in another setting even in the same data context (See Kocherlakota and Yi (1996), Semykina (2018) among others). Therefore, detecting the presence of endogeneity is important as a preliminary step for determining the estimation strategy in any empirical analysis. Due to a challenging testing procedure, there are only a few tests for endogeneity. This paper develops a consistent nonparametric test for endogeneity to aid in more accurate estimation strategy of the model.

My nonparametric test is based on a nonparametric triangular simultaneous equations model from Newey et al. (1999) and Su and Ullah (2008). This model is essential to incorporate endogeneity by introducing instrumental variables. Triangular simultaneous equations consist of a structural equation (or second-stage equation) and a reduced-form equation (or first-stage equation). In addition, nonparametric estimation in each equation is run to overcome the weaknesses of a parametric estimation because the misspecification of a model undermines the consistency of a test.

Under the given setting, I take the control function approach (CFA), which allows an endogenous factor to enter the structural equation. This endogenous factor in the structural equation is presented as the conditional moment $E[U|V]$, where U is the error terms from a structural equation and V is the error terms of a reduced-form equation. The CFA is practical in that it is equivalent to a two-stage least squares estimation but simpler in that the estimation can be done only through the structural equation (See Blundell and Powell (2003), Das et al. (2003), Horowitz (2011), Kasy (2011), Murtazashvilli and Wooldridge (2016) among others).

Another advantage of implementing the CFA also relates to constructing a test for endogeneity, and it has not been used for any nonparametric test for endogeneity. In a conventional triangular simultaneous equations setting, the moment condition of interest for testing endogeneity is $E[U|X, Z] = 0$, where X is a set of potentially endogenous variables, and Z is a set of potential instruments. With the CFA, I can convert the moment condition of interest to a simpler form to construct a test with the reduced dimension. This contributes to resolving the curse of dimensionality in a nonparametric setting as well as the computational burden in the estimation of the conditional moment.

Using the converted moment condition, I set up the null hypothesis as no endogeneity against a presence of endogeneity. Then I construct a conditional moment test using kernel weighting (Li and Wang (1998), Hsiao and Li (2001), Henderson et al. (2008), Wang et al. (2018) among others). The conditional moment test is simple to construct as it only requires the null hypothesis compared to other nonparametric tests using the estimation under the alternatives (See Gonzalo (1993), Fan and Li (2002), Su et al. (2013), Lee et al. (2015), Yao and Ullah (2013), Chen and Pouzo (2015) among others). In addition, the kernel weighting enables the local approximation of the conditional moment.

Once constructing a test, I introduce a Wild bootstrap procedure using Mammen's distribution to improve the finite-sample performance of a test. Wild bootstrap is resampling residuals using a two-point distribution, which allows heteroskedasticity as well as non-*i.i.d.* structure (See Wu (1983), Liu (1988), Mammen (1993), Davidson and Flachaire (2008) among others). Since Wild bootstrap is more robust than a pair bootstrap and resampling bootstrap (See Efron (1979), Horowitz (2001, 2003) among others), it has been used in many previous studies (See Li and Wang (1998), Fan and Li (2002) among others), but not in a simultaneous equations model.

There are two main contributions of this paper to econometrics. For one part is in the estimation in that I take the control function approach in a nonparametric triangular simultaneous equations model. By applying the nonparametric estimation, this nonparametric test overcomes the chronic limitations of a misspecification of the functional form. The nonparametric estimation improves the power of the test because a correct model specification under the alternative ensures the consistency of a test. In addition, I can convert the conditional moment of interest $E[U|X, Z]$ to $E[U|V]$ by taking the control function approach. I reduce the dimension of a conditional moment of interest, which is important to resolve a potential problem of curse of dimensionality. This converted moment condition has not been used for testing in the current literature and makes the estimation of the conditional moment simpler.

The other part of the contributions lies in the simple construction and implementation of the test by using a kernel weighting. Even though there are some test statistics that are difficult to implement despite their advantages in accuracy, my test statistic is easy to implement because it only requires the null hypothesis. Furthermore, I can capture nonlinear correlations with the kernel weighting. Using the kernel increases the accuracy for detecting endogeneity as local approximation of the correlation between U and V becomes possible. In addition to the improvement in accuracy, it can be a useful test due to its simplicity since it follows the standard normal under the null. I also introduce a Wild bootstrap procedure to enhance the finite-sample performance.

A large literature has been developed on estimation methods of nonparametric simultaneous equations (See Newey et al. (1999), Su and Ullah (2008), Matzkin (2008), Berry and Haile (2018), Hahn et al. (2018), Imbens and Newey (2009) among others). However, as my test requires only the null hypothesis, I do not need to implement these nonparametric two-stage estimation methods. Rather, I can apply conventional nonparametric estimation methods to obtain Nadaraya-Watson estimator or local linear estimator (See Pagan and Ullah (1999), and Li and Racine (2007) among others).

For the current parametric tests for endogeneity, the Hausman test and Wu test are the most popular endogeneity test in a parametric regression setting (See Wu (1973), Hausman (1978)). Even though both tests are constructed in a different way, they are analogous in that the model specification under the alternative is confined to parametric estimation. As noted earlier, the power of a test inevitably declines if the model specification under the alternative is incorrect. Acknowledging this, many empirical papers instead report the difference between OLS estimates and 2SLS estimates as an alternative (Angrist and Evans (1998), Autor et al. (2013), Coglianese et al. (2017), Semykina (2018) among others).

There have been a few papers on nonparametric tests for endogeneity (See Blundell and Horowitz (2007), Breunig (2015)). The advantage of these tests lies in their great performance in size and power in a finite sample and they are good for a global approximation. Meanwhile, a kernel weighting method is suited for the local approximation. If the data have nonlinear elements in a small range of a variable, then local approximation with the kernel can capture the endogeneity more accurately. The use of kernel weighting in constructing my test can require additional computational burden. However, as I reduce the dimension in the conditional moment of interest, such computational burden of using the kernel function is lessened. Also, my test can enhance the finite-sample performance with bootstrapping.

The paper is organized as follows. Section 3.2 introduces a model, hypotheses, and the test statistic for endogeneity. In Section 3.3, I conduct Monte Carlo simulations with other current test statistics for endogeneity. Once introducing the extension of the conditional moment test to include other exogenous variables in Section 3.4, I then apply the test to the empirical data in Section 3.5. I will test the endogeneity of Chinese import shock to the US with the US local unemployment share using Autor et al. (2013). Section 3.6 concludes the paper.

3.2 A Consistent Nonparametric Test for Endogeneity

In this section, I introduce a triangular simultaneous equations model and hypotheses. Then, I propose the test statistic for endogeneity and its asymptotic properties. Lastly, a wild bootstrap procedure is proposed to improve the test's performance in a finite sample using Mammen's distribution.

3.2.1 Model and Hypotheses

In constructing a test for endogeneity, I consider a triangular nonparametric simultaneous equations model of Newey et al. (1999) and Su and Ullah (2008), which is given as

$$\begin{cases} y_i & = m\left(x_i\right) + u_i \\ x_i & = g\left(z_i\right) + v_i, \end{cases} \tag{3.1}$$

for $i = 1, \ldots, n$, where y_i is an observable scalar random variable, x_i is a $d_x \times 1$ vector of regressors, and z_i is a $d_z \times 1$ vector of instrumental variables with the unknown functions $m : \mathbb{R}^{d_x} \to \mathbb{R}^1$ and $g : \mathbb{R}^{d_z} \to \mathbb{R}^{d_x}$. All the variables are *i.i.d.* over i. u_i and v_i are disturbances such that $E[u_i \mid x_i, z_i] = 0$ and $E[v_i \mid z_i] = 0$ are satisfied. From equation

(3.1), the moment condition of interest in the structural equation is $E[u_i \mid x_i,\, z_i] = 0$. Assuming the exogeneity of instrumental variables, the moment condition itself can be used to test for endogeneity (Blundell and Horowitz (2007) and Breunig (2015)).

As an alternative to estimate this triangular simultaneous equations model, the structural equation can be re-written as follows by taking the control function approach (Blundell and Powell (2003), Das et al. (2003), Horowitz (2011), Kasy (2011), Murtazashvilli and Wooldridge (2016) among others):

$$\begin{aligned} y_i &= m(x_i) + u_i \\ &= m(x_i) + E[u_i \mid v_i] + u_i - E[u_i \mid v_i] \\ &= m(x_i) + h(v_i) + \epsilon_i, \text{ where } \epsilon_i = u_i - E[u_i \mid v_i] \\ &= m_1(x_i, v_i) + \epsilon_i \end{aligned} \tag{3.2}$$

It is easy to show $E[\epsilon_i \mid v_i] = E[u_i - E[u_i \mid v_i] \mid v_i] = E[u_i \mid v_i] - E[u_i \mid v_i] = 0$. More importantly, for testing endogeneity, the moment condition $E[u_i \mid x_i,\, z_i]$ can be also expressed from the equation (3.2) as

$$\begin{aligned} E[u_i \mid x_i,\, z_i] &= E[u_i \mid x_i - g(z_i),\, z_i] \\ &= E[u_i \mid v_i] \end{aligned}$$

The two moment conditions are equivalent under the exogeneity of Z. But the latter condition can be useful because this conversion reduces the dimension, which can resolve curse of dimensionality to some degree. Based on the re-written model, I develop a direct test for the endogeneity *under the assumptions above*, i.e., whether $E[u_i|v_i] = 0$ a.s. or not. The testing hypotheses are as follows:

$$\begin{cases} \mathbb{H}_0 & : \Pr\left(E\left[u_i|v_i\right] = 0\right) = 1 \\ \mathbb{H}_1 & : \Pr\left(E\left[u_i|v_i\right] = 0\right) < 1 \end{cases}$$

Under H_0, if $E[u_i \mid v_i] = 0$, this implies that there is no endogeneity in the model. If not, there exists endogeneity. The test results will suggest which estimation strategy can give a consistent and most efficient estimator.

I will first show $E\left[u_i|v_i\right] = 0$ is equivalent to $E[f\left(v_i\right) u_i E[u_i|v_i]] = 0$ since I use the latter moment condition in constructing a test statistic. This has been used in other nonparametric test literature as well (Li and Wang (1998), Hsiao and Li (2001), Henderson et al. (2008), Wang et al. (2018) among others). Showing the equivalence of two moment conditions is given in Theorem 2. Even though they are equivalent, the latter condition has an advantage by simplifying the form of the test statistic by cancelling out the marginal density of v_i in the denominator for estimating $E\left[u_i|v_i\right]$. I will use the moment condition in Theorem 2 in constructing a test for endogeneity.

Theorem 2 $E[u_i|v_i] = 0$ *iff* $E[f\left(v_i\right) u_i E[u_i|v_i]] = 0$*, where* $f\left(\cdot\right)$ *is the density function of* v_i *that is bounded away from zero for all* v*.*

Proof of Theorem 2 I let $f\left(v\right)$ and $f(u|v)$ be the marginal density of v_i and the conditional density of u_i given $v_i = v$, respectively.

$$\begin{aligned}
E\left[f\left(v_i\right)u_i E[u_i|v_i]\right] &= \iint u_1\left(\int u_2 f\left(v,u_2\right)du_2\right)f\left(v,u_1\right)du_1 dv \\
&= \int\left(\int u_1 f\left(u_1|v\right)du_1\right)\left(\int u_2 f\left(u_2|v\right)du_2\right)f^2\left(v\right)dv \\
&= \int\left(\int u f\left(u|v\right)du\right)^2 f^2\left(v\right)dv,
\end{aligned}$$

since u_i is $i.i.d.$ over i. Therefore, $E[u_i|v_i] = \int u f\left(u|v\right)du = 0$ iff $E[f\left(v_i\right)u_i E[u_i|v_i]] = 0$ since $f\left(v\right) > 0$. ■

3.2.2 Test Statistic and Asymptotic Properties

Define the probability density function of $\hat{v}_i$ as $\hat{f}(\hat{v}_i) = \frac{1}{n-1}\sum_{j\neq i}^{n} K\left(H_v^{-1}\left(\hat{v}_j - \hat{v}_i\right)\right)$. The sample analogue of $\mathrm{E}[u_i \mathrm{E}[u_i \mid v_i] f(v_i)] = 0$ can be derived as follows:

$$\begin{aligned}
I_n &= \widehat{E}[\widehat{f}\left(\widehat{v}_i\right)\widehat{u}_i\widehat{E}[\widehat{u}_i|\widehat{v}_i]] \\
&= \frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i\widehat{f}\left(\widehat{v}_i\right)\left\{\frac{1}{\left(n-1\right)\left|H_v\right|\widehat{f}\left(\widehat{v}_i\right)}\sum_{j\neq i}^{n}\widehat{u}_j K\left(H_v^{-1}\left(\widehat{v}_j-\widehat{v}_i\right)\right)\right\} \qquad (3.3) \\
&= \frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i\left\{\frac{1}{\left(n-1\right)\left|H_v\right|}\sum_{j\neq i}^{n}\widehat{u}_j K\left(H_v^{-1}\left(\widehat{v}_j-\widehat{v}_i\right)\right)\right\} \\
&= \frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\widehat{u}_i\widehat{u}_j K\left(H_v^{-1}\left(\widehat{v}_j-\widehat{v}_i\right)\right),
\end{aligned}$$

where K is a non-negative d_x-variate kernel function, and H_v is a $d_x \times d_x$ bandwidth matrix that is symmetric and positive definite; $|H_v|$ is the determinant of H_v.

The difference from the Li-Wang type test is that I use generated regressors inside the kernel function[1]. Then, the estimates can be obtained as below.

$$\widehat{u}_i = y_i - \widehat{m}(x_i),$$
$$\widehat{v}_i = x_i - \widehat{g}(z_i)$$

where $\widehat{m}(\cdot)$ and $\widehat{g}(\cdot)$ are consistent estimators under $\mathbb{H}_0$ using the conventional nonparametric regression (either local constant or local linear). Since the test statistic is constructed under the null hypothesis, I do not consider instrumental variable estimation. For characterizing the asymptotic distribution, the following assumptions will be used.

(A1) $\{y_i, X_i, Z_i\}_{i=1}^n$ is independently and identically distributed (IID).

(A2) $E[u \mid z] = 0$, $\sigma^2(v) = E[u^2 \mid v]$, $\sigma^2(v)$ is continuous at v and $E[\sigma^2(v)] < \infty$.

The model assumes the *i.i.d.* distribution of $\{y_i, X_i, Z_i\}_{i=1}^n$. Also, as my interest lies in testing the endogeneity of X, I assume the exogeneity of Z. The conditional variance $\sigma^2(v)$ is continuous at v and its expectation is finite. I do not assume the homoskedasticity for the conditional variance.

(A3) $f(x)$ is uniformly continuous at $x, \forall x \in G, G$ compact subset of $\mathbb{R}$, $0 < f(x) \leq B_f < \infty$, and $|f(x) - f(x')| < m_f\,|x - x'|$ for some $0 < m_f < \infty$ is satisfied.

[1]Li and Wang (1998) use fixed regressors X inside the kernel function. Theorem 5 from Hsiao and Li (2001) suggests a test using generated regressors inside the kernel function, but those generated regressors are from the parametric estimation.

(A4) The kernel function $K(\cdot)$ is bounded and symmetric density function with compact support such that $\int K(\psi)d\psi = 1$. For $\forall x \in \mathbb{R}$, $|K(x)| < B_k < \infty$. I assume $\left|K^j(u) - K^j(v)\right| \leq C_1 |u-v|$, for $j = 0, 1, 2, 3$.

In (A3), the conditional density $f(x)$ satisfies the Lipschitz continuous condition. In addition, as it is smooth and bounded, a Taylor expansion can be applied. When constructing the test, I use the kernel function as a weighting function. Regarding properties of the kernel function, it is bounded and symmetric. As in $f(x)$, the kernel function satisfies the Lipschitz continuous function.

(A5) As $n \to \infty$, each element of $H_v, H_z, H_x \to 0$. i) It satisfies $n^{1/2} |H_z| |H_v|^{1/2} / \ln n \to \infty$, $n |H_v|^2 \to \infty$, and $n |H_v|^6 \to 0$. ii) $n |H_z| / \ln n \to \infty$ and $n |H_x| / \ln n \to \infty$.

This assumption is on the restriction of the bandwidth. A5-i) is for the asymptotic properties of the proposed test statistic and A5-ii) is the standard assumptions for nonparametric estimation. (A5)-i) are the additional assumptions because I apply the nonparametric estimation to obtain the residuals inside the kernel function. For parametric residuals, this assumption is not needed as seen in Hsiao and Li (2001).

(A6) $m(\cdot)$ and $g(\cdot)$ are continuous and twice differentiable in X and Z respectively.

The last assumption (A6) is to allow for the differentiability of $g(\cdot)$ is necessary in terms of applying a Taylor expansion inside the kernel function of the test statistic for all the following theorems. Based on these assumptions, I standardize the estimator,

$$J_n = \sqrt{n^2 |H_v|} I_n / \sqrt{\hat{\Omega}}, \text{where } \hat{\Omega} = \frac{2}{n(n-1)|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{u}_i^2 \hat{u}_j^2 K\left(H_v^{-1}\left(\hat{v}_j - \hat{v}_i\right)\right).$$

Theorem 3 shows that the asymptotic distribution of this test statistic follows the standard normal distribution. Based on this, the asymptotic critical value of the test can be calculated. Therefore, when J_n is large enough to exceed the critical value of $N(0,1)$ at α-percent level, then I reject the null hypothesis, meaning that the variable of our interest is not endogenous. Otherwise, I accept the null hypothesis.

Theorem 3 *Under* $\mathbb{H}_0$, *as* $\hat{\Omega}$ *is a consistent estimator of* $\Omega = 2[\int K^2(\psi)d\psi]E[\sigma^4(v)f(v)]$,

$$J_n \to N(0,1) \text{ as } n \to \infty.$$

For the asymptotic properties under the alternative, I introduce the Pitman local alternatives as follows:

$$\mathbb{H}_1(\delta_n) : m_1(x_i, v_i) = m(x_i) + \delta_n l(v_i)\,,$$

where $l(\cdot)$ is continuously differentiable and bounded, and $\delta_n = n^{-1/2}\,|H_v|^{-1/4}$. Based on the equation (2), note that $l(\cdot)$ does not include the elements of x_i because $m_1(x_i, v_i)$ is separable by construction of the model[2].

Theorem 4 *Under the Pitman local alternative, if* $\delta_n = n^{-1/2}\,|H_v|^{-1/4}$, *then*

$$J_n \xrightarrow{d} N(E[l(v_i)^2 f(v_i)]/\sqrt{\Omega}, 1) \text{ as } n \to \infty.$$

[2]This is different from Yao and Ullah (2013) which tests for a relevant variable. Under the alternative, as they do not assume the separability of two sets of variables x_{1i} and x_{2i} , $m(x_{1i}, x_{2i}) = m(x_{1i}) + \delta_n\ l(x_{1i}, x_{2i})$ under the alternative.

Then, as the magnitude of $E[l(v_i)^2 f(v_i)]/\sqrt{\Omega}$ increases, the test statistic deviates farther from the zero mean, and the local power increases. However, the variance remains at one for both hypotheses.

Theorem 5 *Assuming (A1)-(A6) and under* $\mathbb{H}_1$*,* $Pr[\hat{J}_n > B_n] \to 1$ *for any non-stochastic sequence* $\{B_n : B_n = o(\sqrt{n^2 |H_v|})\}$*. Under* H_1*,* $\hat{I}_n = I_n + o_p((n |H_v|^{1/2})^{-1})$*, where* $I_n = E[(h(v_i))^2 f(v_i)]$*, and* $\hat{\Omega} = \Omega + o_p(1)$*.*

Theorem 5 suggests the consistency of the test statistic. Under $\mathbb{H}_1$, the probability of rejecting the null will converge to 1.

3.2.3 Bootstrap Method

As the asymptotic normal approximation does not perform well in small sample settings, I propose a wild bootstrap test as an alternative. Hardle and Mammen (1993) proposed a wild bootstrap method using two-point distribution. Wild bootstrap method has advantages among different bootstrap methods in that it can generate the non-*i.i.d.* samples as well as allowing heterogeneity in the sample. Among the choices for a two-point distribution, I use Mammen's distribution rather than Rademacher distribution because it does not require the symmetry of a distribution. In this regard, I apply a wild bootstrap method using Mammen's distribution. Steps to get a bootstrap test statistic are given below.

1. Estimate $\hat{g}(z_i)$ and $\hat{m}(x_i)$ by a nonparametric kernel estimation (either LCLS or LLLS) for a structural and reduced-form equation respectively. Note that this is not an instrumental variable (IV) estimation.

2. Generate u_i^* as the wild bootstrap error. I construct $u_i^* = \frac{1-\sqrt{5}}{2}\hat{u}_i$ with the probability of $\frac{1+\sqrt{5}}{2}$ and $u_i^* = \frac{1+\sqrt{5}}{2}\hat{u}_i$ with the probability of $1-\frac{1+\sqrt{5}}{2}$. It is easy to show $\mathrm{E}[u_i^*] = 0$, $E\left[u_i^{*2}\right] = \hat{u}_i^2$, and $\mathrm{E}\left[u_i^{*3}\right] = \hat{u}_i^3$.

3. Generate y_i^*, where $y_i^* = \hat{m}(x_i) + \hat{u}_i^*$ under the null hypothesis.

4. Using the bootstrap sample $\{y_i^*, x_i, z_i\}_{i=1}^n$, regress y_i^* on x_i^* to obtain $\hat{m}^*(x_i^*)$, and get $\hat{u}_i^* = y_i^* - \hat{m}^*(x_i)$. Under the null, the variable of interest lies in the structural equation by taking a control function approach. Thus, I do not generate the wild bootstrap sample on $\{x_i, z_i\}_{i=1}^n$. Therefore, $\hat{v}_i^* = \hat{v}_i$.

5. With $\{\hat{u}_i^*, \hat{v}_i^*\}_{i=1}^n$, compute the bootstrap test statistic J_n^* and repeat above procedure for B times. In the simulation, the number of bootstrapping used is 399.

6. Based on the empirical distribution of J_n^*, calculate the critical value c^* and obtain the p-value, which is $P(J_n \geq c^*)$. If p-value is less than 0.05 at 5% significance level, the null is rejected.

Following these bootstrap procedures, I can obtain the asymptotic distribution of J_n^*. I will show how the bootstrap test performs in the Monte Carlo Simulations. The asymptotic distribution of bootstrap test under the null is shown in Theorem 6.

Theorem 6 *Let the bootstrap statistic be defined as* $J_n^* = n\left|H_v\right|^{1/2} I_n^*/\sqrt{\hat{\Omega}^*}$*, where* $\hat{\Omega}^* = \frac{2}{n(n-1)|H_v|}\sum_{i=1}^n \sum_{j\neq i}^n \hat{u}_i^{*2}\hat{u}_j^{*2} K(H_v^{-1}(\hat{v}_j^* - \hat{v}_i^*))$*. Under* $\mathbb{H}_0^*$*, as* $\hat{\Omega}^*$ *is a consistent estimator of* $\Omega = 2[\int K^2(\psi)d\psi]E[f(v)\sigma^4(v)]$*,* $J_n^* \to N(0,1)$ *in distribution as* $n \to \infty$*.*

The proofs of Theorem 6 will follow similarly to those of Theorem 3. Under the $H_1(\delta_n)$, $P(J_n > c^*) \to 1$ asymptotically, where c^* denotes the bootstrap critical value based on the boostrap samples. This shows the consistency of the bootstrap test statistic.

3.3 Simulations

3.3.1 Data Generating Processes

Now, I perform the test for endogeneity using three different data generating processes. For DGP_1, I followed data generating process from Newey and Powell (2003). Here, $\{Y_i, X_i, Z_i\}_{i=1}^n$ does not have a bounded support.

$$\text{DGP}_1\text{:}\begin{cases} Y_i = m(X_i) + U_i = \log(\mid X_i - 1 \mid +1) sgn(X_i - 1) + U_i \\ X_i = g(Z_i) + V_i = Z_i + V_i \end{cases},$$

where $i = 1, ..., n$, and errors $U_i\, V_i$, and Z_i are generated as

$$\begin{pmatrix} U_i \\ V_i \\ Z_i \end{pmatrix} \sim \ i.i.d.\ \text{N} \begin{pmatrix} 1 & \theta & 0 \\ \theta & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Next, I do the simulations where $\{Y_i, X_i, Z_i\}_{i=1}^n$ has a bounded support in DGP_2 and DGP_3. DGP_2 is from Su and Ullah (2008) and DGP_3 is modified from DGP_2.

$$\begin{cases} Y_i = 1 + 2\exp(X_i)/(1 + \exp(X_i)) + U_i \\ X_i = Z_i + V_i \end{cases},$$

where $i = 1, ..., n$, errors $V_i,$ and Z_i are generated as

$$V_i = 0.5w_i + 0.2v_x,\ Z_i = 1 + 0.5v_z,$$

in which $v_{y,} v_x$, w_i are i.i.d. sum of 48 independent random variables each uniformly distributed on [-0.25,0.25].

$$\text{DGP}_2 : U_i = \theta w_i + 0.3 v_y$$

$$\text{DGP}_3 : U_i = \theta(w_i + 2w_i^2) + 0.3 v_y$$

For all three data generating processes, $\theta = 0, 0.2, 0.5$, and 0.8, which indicates no endogeneity, weak endogeneity, medium endogeneity and strong endogeneity, respectively. In particular, when $\theta = 0$, note that it refers to no endogeneity and DGP_2 and DGP_3 become identical. The main difference between two data generating processes is how U_i and V_i are correlated in terms of a functional form in the presence of endogeneity while the model is still correctly specified. In this regard, the simulation results for DGP$_3$ will present how my nonparametric test captures such nonlinear terms under the alternative.

For bandwidth selection, I use rule-of-thumb bandwidths for both the estimation and the test. For the estimation, I use local linear estimation with a second-order Epanechnikov kernel by using a rule-of-thumb bandwidth[3], which is $h_x = 2.34 std(x_i) n^{-1/5}$ and $h_z = 2.34 std(z_i) n^{-1/5}$ for the structural equation and reduced-form equation respectively. I obtain $\hat{m}(x_i) = \hat{\alpha}$ from $(\hat{\alpha}, \hat{\beta}) = \arg\max_{\alpha,\beta} \sum_{t=1}^{n} (y_t - \alpha - \beta(x_i - x_t))^2$. For the test bandwidth, $h_v = c \cdot std(x_i) n^{-1/5}$ and $c = 0.5,\ 1.06,\ 1.5$. For the Blundell and Horowitz (2007) test, I use the cross-validation bandwidth for estimating the joint density of (X, Z), and obtain the bandwidth by multiplying $n^{1/5-7/24}$ times the cross-validation bandwidth. For the Breunig

[3] This rule-of-thumb bandwidth when using a second-order Epanechnikov kernel is suggested by Henderson et al. (2012).

test (2015), I use series estimation based on his simulation settings. Since both tests use a Fourier series as a basis function, and I implement cosine basis functions given by $f_j(t) = \sqrt{2}\cos(j\pi t)$ for $j = 1, 2, ...M$. I set $M = 40$ and the smoothing parameter as $\tau_j = j^{-1}$. The number of repetition is 1000 for the sample size of 100, and 500 for the sample size of 400. The number of bootstrap repetitions is 399 for both sample sizes.

3.3.2 Simulation Results

For each data generating process, both the size and the power are estimated by changing the strength of endogeneity (the value of θ). I then compare my test's performance with the Hausman test, the Blundell and Horowitz (2007) test (BH_n), and the Breunig (2015) test (B_n). The Hausman test is a parametric test, where it measures the difference between OLS and 2SLS estimates. While Blundell and Horowitz (2007) apply a kernel-based estimation and Breunig (2015) applies a series-based estimation, both use Fourier series in constructing a test statistic.

Table 3.1-3.4 represent both size and power for each data generating process with different values of bandwidth. Other than my conditional moment test (J_n and J_n^*), all other tests' performance does not vary with the bandwidth.[4] In addition, I apply bootstrap procedure to the conditional moment test (J_n and J_n^*) as its asymptotic distribution follows the standard normal as in Theorem 6 to improve a finite-sample performance. However, as the asymptotic distribution for both BH_n and B_n is not pivotal, I do not apply bootstrap procedure.

[4] Even though Blundell and Horowitz (2007) applies a kernel-based estimation, they construct a series-based test. Thus, different bandwidths for the test only applied to the conditional moment test.

Table 3.1 presents the estimated size for all cases. As mentioned earlier, DGP_2 and DGP_3 results are identical when there exists no endogeneity. The bootstrap size of my conditional moment test is close to the correct size at each significance level although J_n is undersized in the asymptotic test[5]. For different bandwidths, their estimated size is close to the nominal size in all significance levels and its performance improves with the increase in size. Overall, the size of the Hausman test is close to the correct size for other significance levels. In addition, the BH_n and B_n tests are undersized, but their performance is better in the bounded support of $\{Y_i, X_i, Z_i\}_{i=1}^n$ since both tests assume the bounded support.

In Table 3.2, the power of DGP_1 is shown for a different level of endogeneity under an unbounded support of $\{Y_i, X_i, Z_i\}_{i=1}^n$. With weak endogeneity, the power of the test is slightly over the nominal size. As the strength of endogeneity increases, the test becomes more powerful. Furthermore, the power increases as the bandwidth increases, which can be explained by Theorem 4. In all sample sizes, the Hausman test performs better than the conditional moment test except when there is a stronger endogeneity ($\theta = 0.8$). While the rejection probabilities of BH_n, and B_n test rise as the sample size as well as the level of endogeneity increase, it does not perform as well as the conditional moment test.

The power of DGP_2 is presented in Table 3.3. In a small sample, my conditional moment test is the most powerful at each level of endogeneity among all tests. The power of the conditional moment test is almost equal to 1 with the sample size of 400 even in the presence of weak endogeneity with the bounded support case. Hausman test performs equally as well as the conditional moment test except when under the weak endogeneity for both small and large sample sizes. Overall, BH_n test performs better mostly in a large

[5]This underestimation of the size has been also noted in Li and Wang (1998) and Hsiao and Li (2001).

sample. However, the power of BH_n is slightly over the nominal size in a small sample. BH_n test outperforms B_n both in small and large samples.

Table 3.4 presents the power of DGP_3, where a nonlinear correlation between U_i and V_i is present. Considering that my nonparametric test can capture the nonlinear relationship between U_i and V_i, the test performance compared to Hausman test is noticeable in estimating the power for all sample sizes. First, due to a presence of the nonlinear term in the data generating process, my test's power reaches almost 1 even in a small sample size as well as the weak endogeneity. In contrast, as Hausman test cannot capture the nonlinear relationship, power falls compared to DGP_2 in a small sample. In addition, the Hausman test's power is less than its performance with the presence of a linear correlation. This implies that the Hausman test can be inconsistent if the model specification is incorrect under the alternative.

In summary, even though I observed the undersized test for J_n using asymptotic critical values, the estimated size based on bootstrap procedure is close to the nominal size for all the data generating processes. As the strength of endogeneity increases, the test becomes more powerful. At the same time, as the sample size increases, the tests become more powerful for all cases. Compared to the Hausman test, BH_n, and B_n tests, my conditional moment test performs the best in that it can detect the nonlinear relationship between U_i and V_i and it is robust to any choice of bandwidth.

Furthermore, I can compare clearly the nonparametric tests' performance between a kernel method and a series-based method. BH_n test uses a series method for the test but still applies a kernel method in the estimation while B_n test is solely on series-based estimator. As kernel-based method performs well in a local approximation, BH_n test is better than B_n test overall. However, the current test dominates all the other tests in both sample sizes in that I use kernel techniques in running an estimation and constructing a test to capture the local correlation.

3.4 Extension: The Case with Other Exogenous Variables

Extending the previous model, I consider a case with both endogenous and exogenous regressors.

$$\begin{cases} y_i & = m\left(x_{i1}, x_{i2}\right) + u_i \\ x_i & = g\left(z_i, x_{i2}\right) + v_i, \end{cases}$$

where $i = 1, ..., n$, y_i is an observable scalar random variable, $m(\cdot)$ denotes a structural function of unknown form, x_{i1} is a $d_{x_1} \times 1$ vector of endogenous regressors, and x_{i2} is a $d_{x_2} \times 1$ vector of exogenous regressors. $g(\cdot)$ is a $d_{x_1} \times 1$ vector of functions of the instruments. z_i is a $d_z \times 1$ vector of instrumental variables. u_i and v_i are disturbances such that $E[u_i \mid z_i, x_{i1}, x_{i2}] = 0$ and $E[v_i \mid z_i, x_{i1}, x_{i2}] = 0$ are satisfied. Additional assumptions are needed for this extension.

(B1) $\{Y_i, X_{1i}, X_{2i}, Z_i\}_{i=1}^n$ is independent and identically distributed.

(B2) $E\left[u\right| z, x_1, x_2] = 0.\sigma^2(v) = E[u^2 \mid v]$, $\sigma^2(v)$ is continuous at v and $E[\sigma^2(v)] < \infty$.

(B3) $f(x)$ is differentiable, $0 < f(x) \leq B_f < \infty,$ and $|f(x) - f(x)| < m_f\left|x - x'\right|$ for some $0 < m_f < \infty$ is satisfied.

Table 3.1: Size of Each Test

			DGP$_1$			DGP$_2$			DGP$_3$		
		c	1%	5%	10%	1%	5%	10%	1%	5%	10%
$n = 100$	J_n^*	0.5	0.011	0.048	0.093	0.015	0.053	0.102	0.015	0.053	0.102
		1.06	0.011	0.045	0.084	0.009	0.050	0.106	0.009	0.050	0.106
		1.5	0.009	0.043	0.089	0.005	0.048	0.089	0.005	0.048	0.089
	J_n	0.5	0.007	0.022	0.042	0.013	0.028	0.053	0.013	0.028	0.053
		1.06	0.006	0.010	0.018	0.008	0.014	0.022	0.008	0.014	0.022
		1.5	0.004	0.005	0.010	0.002	0.007	0.011	0.002	0.007	0.011
	BH_n	–	0.000	0.002	0.004	0.004	0.005	0.008	0.004	0.005	0.008
	B_n	–	0.026	0.040	0.053	0.034	0.051	0.061	0.034	0.051	0.061
	H_n	–	0.009	0.047	0.099	0.007	0.055	0.103	0.007	0.055	0.103
$n = 400$	J_n^*	0.5	0.010	0.058	0.090	0.008	0.048	0.098	0.008	0.048	0.098
		1.06	0.008	0.046	0.098	0.010	0.044	0.104	0.010	0.044	0.104
		1.5	0.012	0.046	0.100	0.010	0.030	0.088	0.010	0.030	0.088
	J_n	0.5	0.010	0.026	0.058	0.004	0.014	0.050	0.004	0.014	0.050
		1.06	0.010	0.018	0.028	0.002	0.008	0.016	0.002	0.008	0.016
		1.5	0.004	0.010	0.018	0.002	0.006	0.006	0.002	0.006	0.006
	BH_n	–	0.000	0.006	0.022	0.004	0.020	0.042	0.004	0.020	0.042
	B_n	–	0.000	0.002	0.008	0.000	0.004	0.004	0.000	0.004	0.004
	H_n	–	0.018	0.050	0.100	0.014	0.050	0.098	0.014	0.050	0.098

Note: Note that there is no difference in size between DGP$_2$ and DGP$_3$ because the difference between the two data generating processes comes from the non-zero value of θ. Except for J_n and J_n^*, all other tests are not constructed based on the kernel techniques. Therefore, the test performance does not vary with the bandwidth choice.

Table 3.2: Power of Each Test of DGP_1

			$\theta = 0.2$			$\theta = 0.5$			$\theta = 0.8$		
n		c	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	J_n^*	0.5	0.009	0.054	0.116	0.084	0.206	0.313	0.675	0.884	0.936
		1.06	0.013	0.065	0.121	0.174	0.379	0.506	0.901	0.981	0.996
		1.5	0.017	0.085	0.138	0.254	0.491	0.621	0.958	0.998	0.998
	J_n	0.5	0.007	0.023	0.055	0.091	0.174	0.233	0.772	0.888	0.933
		1.06	0.005	0.020	0.033	0.138	0.224	0.290	0.931	0.973	0.981
		1.5	0.005	0.015	0.023	0.135	0.227	0.302	0.955	0.981	0.990
	BH_n	–	0.015	0.067	0.141	0.029	0.107	0.190	0.008	0.016	0.022
	B_n	–	0.020	0.034	0.043	0.018	0.037	0.045	0.078	0.110	0.144
	H_n	–	0.497	0.635	0.711	0.558	0.678	0.752	1.000	1.000	1.000
400	J_n^*	0.5	0.030	0.118	0.190	0.708	0.908	0.966	1.000	1.000	1.000
		1.06	0.068	0.192	0.300	0.948	0.994	0.998	1.000	1.000	1.000
		1.5	0.100	0.272	0.360	0.978	1.000	1.000	1.000	1.000	1.000
	J_n	0.5	0.042	0.094	0.146	0.788	0.908	0.950	1.000	1.000	1.000
		1.06	0.070	0.128	0.180	0.954	0.982	0.994	1.000	1.000	1.000
		1.5	0.070	0.144	0.188	0.972	0.994	1.000	1.000	1.000	1.000
	BH_n	–	0.472	0.636	0.730	0.564	0.724	0.798	0.052	0.068	0.090
	B_n	–	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.028	0.046
	H_n	–	0.936	0.960	0.974	0.946	0.976	0.982	1.000	1.000	1.000

Note: Except for J_n and J_n^*, all other tests are not constructed based on the kernel techniques. Therefore, the test performance does not vary with the bandwidth choice.

Table 3.3: Power of Each Test of DGP_2

n		c	$\theta = 0.2$			$\theta = 0.5$			$\theta = 0.8$		
			1%	5%	10%	1%	5%	10%	1%	5%	10%
100	J_n^*	0.5	0.372	0.621	0.730	0.995	1.000	1.000	1.000	1.000	1.000
		1.06	0.601	0.827	0.888	1.000	1.000	1.000	1.000	1.000	1.000
		1.5	0.722	0.882	0.939	1.000	1.000	1.000	1.000	1.000	1.000
	J_n	0.5	0.443	0.597	0.677	0.997	0.999	1.000	1.000	1.000	1.000
		1.06	0.604	0.756	0.808	1.000	1.000	1.000	1.000	1.000	1.000
		1.5	0.640	0.780	0.839	1.000	1.000	1.000	1.000	1.000	1.000
	BH_n	–	0.004	0.005	0.008	0.016	0.036	0.056	0.026	0.048	0.077
	B_n	–	0.034	0.051	0.061	0.062	0.091	0.114	0.065	0.100	0.128
	H_n	–	0.007	0.055	0.103	1.000	1.000	1.000	1.000	1.000	1.000
400	J_n^*	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.06	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	J_n	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.06	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	BH_n	–	0.004	0.020	0.042	0.726	0.756	0.802	0.744	0.802	0.830
	B_n	–	0.000	0.004	0.004	0.006	0.016	0.044	0.010	0.024	0.048
	H_n	–	0.014	0.050	0.098	1.000	1.000	1.000	1.000	1.000	1.000

Note: Except for J_n and J_n^*, all other tests are not constructed based on the kernel techniques. Therefore, the test performance does not vary with the bandwidth choice.

Table 3.4: Power of Each Test of DGP_3

			$\theta = 0.2$			$\theta = 0.5$			$\theta = 0.8$		
n		c	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	J_n^*	0.5	0.993	0.996	0.998	0.999	1.000	1.000	0.999	1.000	1.000
		1.06	0.994	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000
		1.5	0.995	1.000	1.000	0.999	1.000	1.000	0.999	1.000	1.000
	J_n	0.5	0.997	0.998	0.999	1.000	1.000	1.000	1.000	1.000	1.000
		1.06	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.5	0.999	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	BH_n	–	0.004	0.005	0.008	0.029	0.107	0.190	0.034	0.116	0.206
	B_n	–	0.034	0.051	0.061	0.018	0.037	0.045	0.020	0.038	0.047
	H_n	–	0.007	0.055	0.103	0.558	0.678	0.752	0.562	0.691	0.748
400	J_n^*	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.06	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	J_n	0.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.06	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
		1.5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	BH_n	–	0.472	0.636	0.730	0.564	0.724	0.798	0.578	0.734	0.810
	B_n	–	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	H_n	–	0.936	0.960	0.974	0.946	0.976	0.982	0.952	0.976	0.982

Note: Except for J_n and J_n^*, all other tests are not constructed based on the kernel techniques. Therefore, the test performance does not vary with the bandwidth choice.

(B4) The kernel function $K(\cdot)$ is bounded and symmetric density function with compact support such that $\int K(\psi)d\psi = 1$. For $\forall x \in \mathbb{R}$, $|K(x)| < B_k < \infty$. I assume $\left|K^j(u) - K^j(v)\right| \leq C_1 |u - v|$, for $j = 0, 1, 2, 3$.

(B5) As $n \to \infty$, $|H_v|, |H_z|, |H_{x_1}|, |H_{x_2}| \to 0$. i) It satisfies $n^{1/2} |H_z| |H_v|^{1/2} / \ln n \to \infty$, $n |H_v|^2 \to \infty$, and $n |H_v|^6 \to 0$. ii) $n |H_z| / \ln n \to \infty$ and $n |H_x| / \ln n \to \infty$.

(B6) $m(\cdot)$ and $g(\cdot)$ are continuous and twice differentiable in X and Z respectively.

Note that the moment condition of interest with other exogenous variables are identical to the previous case.

$$
\begin{aligned}
E[u_i \mid x_{1i}, x_{2i},\, z_i] &= E[u_i \mid x_{1i} - g(z_i, x_{2i}), x_{2i},\, z_i] \\
&= E[u_i \mid v_i]
\end{aligned}
$$

The test statistic will be written as follows:

$$
I_n = \frac{1}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{u}_i \hat{u}_j K\left(H_v^{-1}\left(\hat{v}_j - \hat{v}_i\right)\right),
$$

where $\hat{u}_i = y_i - \hat{m}(x_{i1,}x_{i2})$, $\hat{v}_i = x_{1i} - \hat{g}(z_i, x_{i2})$, and both $\hat{m}(\cdot)$ and $\hat{g}(\cdot)$ are nonparametric estimates.Then, The standardized test statistic is

$$
J_n = \sqrt{n^2\,|H_v|}\,I_n / \sqrt{\hat{\Omega}},
$$

where $\hat{\Omega} = \frac{2}{n(n-1)|H_v|} \sum_{i=1}^{n} \sum_{j}^{n} \hat{u}_i^2 \hat{u}_j^2 K\left(H^{-1}\left(\hat{v}_j - \hat{v}_i\right)\right)$. The extension of the test statistic is not complicated because the test statistic does not change as it analyzes the correlation between u_i and v_i. The only difference is how the residuals are obtained from the estimation, where $\hat{u}_i = y_i - \hat{m}(x_i)$ and $\hat{v}_i = x_i - \hat{g}(z_i)$. In the next section, I will apply the test for endogeneity using the extension.

3.5 Application

By extending the empirical analysis of Autor, Dorn, and Hanson (ADH, 2013), I apply my test for endogneity. In their paper, they analyze the impact of Chinese import exposure on the US local labor market outcomes including employment share and wages. I mainly focus on the US local employment share in manufacturing. The triangular simultaneous equations are set up is as follows.

$$\begin{cases} \Delta L_{it}^m & = \alpha_t + \beta_1 \Delta IPW_{uit} + X_{it}' \beta_2 + u_{it} \\ \Delta IPW_{uit} & = \gamma_t + \delta_1 \Delta IPW_{oit} + X_{it}' \delta_2 + v_{it}, \end{cases}$$

where ΔL_{it}^m is decadel change in the manufacturing share of the working-age population in commuting zone i. ΔIPW_{uit} is the change in import exposure to the US. ΔIPW_{oit} is the change in import exposure to other high-income markets. Following the same model specifications given in ADH (2013), my main interest is on testing the endogeneity of US trade exposure, ΔIPW_{uit}.

For the testing, two model specifications are considered: One (Model 1) is including only a time dummy, and the other (Model 2) is including Δ(imports from China to US)/worker, percentage of employment in manufacturing in the previous period, and census division dummies. For the parametric estimation, the pooled 2SLS is used as in ADH

(2013) for constructing the Hausman test. As I did in simulations, I present the results of my test in comparison with the Blundell and Horowitz (2007) test (BH_n), the Breunig (2015) test (B_n), and the Hausman test (H_n). For the nonparametric tests, local linear estimation is used and its bandwidth is chosen with cross-validation. For Breunig (B_n) test, I run the local polynomial estimation in getting the residuals. I report both asymptotic and bootstrap p-values. The number of bootstrap is 399. For a test statistic, the rule of thumb bandwidth is used for constructing a test statistic using a Gaussian kernel.

Before presenting the test results, Figure 3.1 and Figure 3.2 gives an idea how the residuals u_i and v_i are correlated when they are estimated differently either in parametric or in nonparametric estimation. The dotted line is to denote the 95% confidence interval. If the zero line is inside the confidence interval, it implies no significant correlation. For Figure 3.1, both parametric and nonparametric residuals present a positive significant correlation. In terms of nonparametric residuals, the positive correlation is more present where the data are concentrated. However, the correlation between parametric residuals and nonparametric residuals is shown differently for Model 2 in Figure 3.2. While I can observe a positive correlation using parametric residuals, I do not see any significant correlation in nonparametric residuals $u_{i,NP}$ and $v_{i,NP}$. This contradicting pattern of the correlation can imply a potential problem of misspecification.

The test results are given in Table 3.5. For the test bandwidth of my test, I use $h_v = c \cdot std(v_i)n^{-1/5}$ and $c = 0.5,\ 1.06,\ 1.5$. For Model 1, I reject the null hypothesis both in asymptotic and bootstrap test at 5% significance level. This result is also consistent with the Hausman test, the Blundell and Horowitz (2007) test, and the Breunig (2015) test. However,

Figure 3.1: Correlation between $\hat{u}_i$ and $\hat{v}_i$ in Model 1

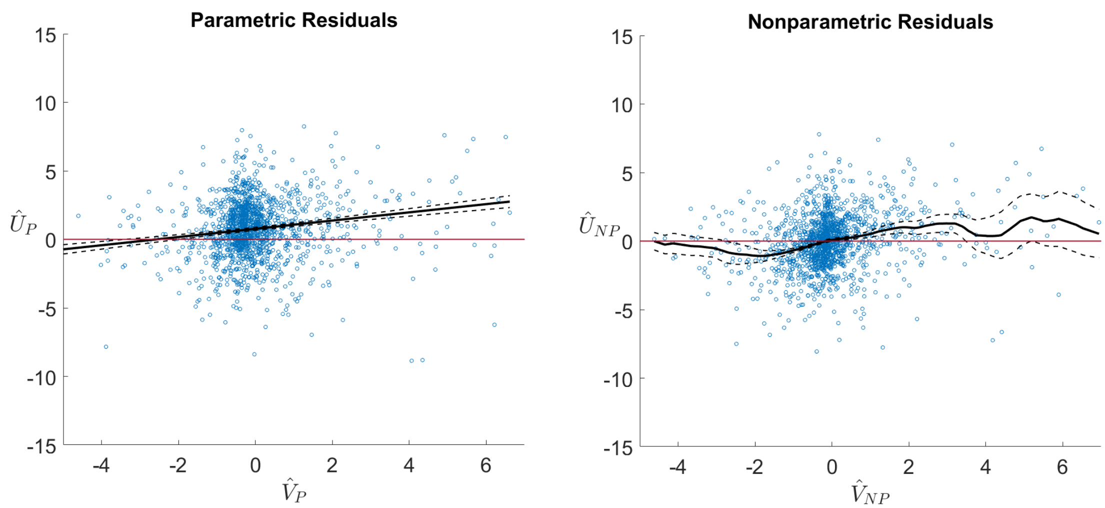
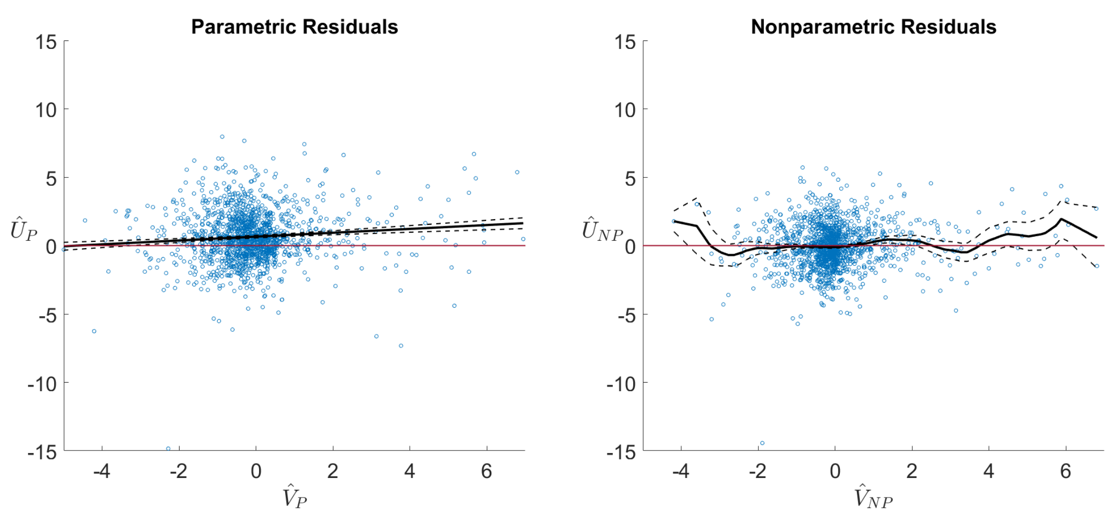


Figure 3.2: Correlation between $\hat{u}_i$ and $\hat{v}_i$ in Model 2



I have a contradicting test result with Hausman test in Model 2. All nonparametric tests do not reject the null hypothesis with the high p-values while the Hausman test still rejects the null hypothesis at 5% significance level.

Table 3.5: P-Values of Each Model

	J_n^*			J_n			BH_n	B_n	H_n
c	0.5	1.06	1.5	0.5	1.06	1.5	–	–	–
Model 1	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.034	0.000
Model 2	0.115	0.499	0.679	0.232	0.487	0.616	0.120	1.000	0.000

Note: Except for J_n and J_n^*, all other tests are not constructed based on the kernel techniques. Therefore, the test performance does not vary with the bandwidth choice.

There can be two possible explanations why I have such a contradicting test results for endogeneity. Considering that Model 2 is estimated by adding other exogenous variables, some factors which cause endogeneity of Chinese import variable might have been filtered out by those variables. In addition, there could the misspecification of the model in terms of the functional form. While the nonparametric tests are not confined to a functional form of the model, parametric tests are. Thus, the misspecification of the functional form can result in the inaccurate detection of endogeneity.

Based on the test results, I further estimate the marginal effect of the Chinese import exposure to US local employment share in manufacturing to discuss the potential bias for both models. The estimation results are given in Figure 3.3. For Model 1, the nonparametric instrumental variable estimation is applied following Darolles et al. (2011). The global nonparametric estimate is -0.883 while the estimate in ADH (2013) is -0.746. This implies that the estimate is overestimated in ADH (2013).

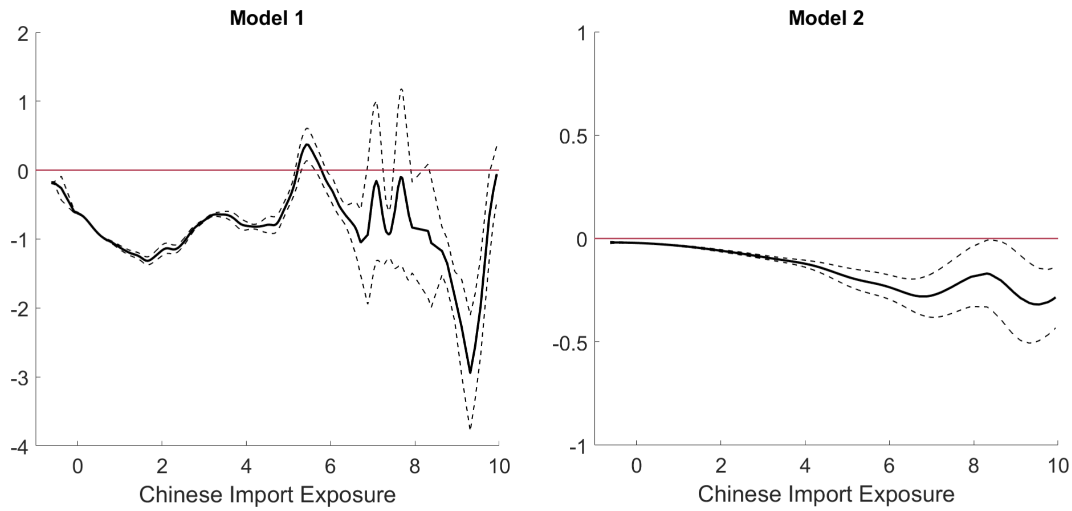
For Model 2, I apply conventional nonparametric estimation because the null hypothesis is not rejected. The global nonparametric estimate is -0.06 and the 2SLS estimate in ADH (2013) is -0.538. The difference between the global nonparametric estimate and 2SLS estimate becomes larger because I do not implement the nonparametric instrumental variable estimation. Even with the parametric OLS estimate, -0.183, it is underestimated in this model specification. In brief, this estimation result implies that the presence of endogeneity is a preliminary step and then the functional form of the estimation strategy is the secondary step in reducing the potential bias of an estimator.

The economic intuition why Chinese import exposure may not be endogenous lies in the inclusion of a variable for the percentage of employment in manufacturing in the previous period. By controlling the percentage of employment in manufacturing in the previous period, this can reflect the shift in the US demand curve to the left, which accompanies the decrease in income. Then, the Chinese import exposure shifts the domestic supply to the left, but it may not further increase Chinese imports because of a decrease domestic demand, which can cut down the simultaneous causality of Chinese import exposure and the current US local employment share.

3.6 Conclusion

Endogeneity is commonly observed in economics by assumption and estimated with instrumental variables in many applied economic papers (Angrist and Evans (1998), Autor et al. (2013), among others). However, testing for the presence of endogeneity cannot be underestimated due to consistency and efficiency issues. Moreover, not every variable which has

Figure 3.3: Marginal Effect of Chinese Import Exposure



been believed to be endogenous is endogenous in every context. Even though there is a large literature on how to deal with the endogeneity in the estimation, testing for endogeneity should be a priority for a more efficient estimator. In this paper, I propose a consistent nonparametric test for endogeneity.

By introducing an alternative way of using the conditional moment in a triangular equations model by taking the control function approach, I can convert the conditional moment for endogeneity test $E[U \mid X, Z] = 0$ to $E[U \mid V] = 0$. As the dimension of V is smaller than that of X and Z, it suffers less from the curse of dimensionality. Based on the modified moment condition, I construct a Li-Wang type test. The advantages of the current test are; i) it follows the standard normal distribution under the null hypothesis, and ii) it can capture the nonlinear correlation between the disturbances U and V aside from the advantage of a nonparametric estimation over a parametric estimation.

As with other nonparametric conditional moment tests (Zheng (1996), Li and Wang (1998), Hsiao and Li (2001), among others), I introduce a wild bootstrap method using Mammen's distribution to improve the finite-sample performance. In simulations, I show that my bootstrap test performs better in finite samples than the asymptotic test for both size and power. In particular, when I have a bounded support for $\{Y_i, X_i, Z_i\}_{i=1}^n$, my test statistic performed better both in estimating size and power than having a unbounded support. Compared to the Hausman, the Blundell and Horowitz (2007), and the Breunig (2015) test, my test statistic outperforms them when the error terms are nonlinearly correlated with each other at all levels of endogeneity. When the error terms are nonlinearly correlated, it seems that the test statistic using a kernel method is better than the statistics using a series estimator.

I also apply this test to the empirical analysis of Autor, Dorn, and Hanson (2013) to test endogeneity of Chinese import exposure with the US local employment. As this estimation includes other exogenous variables, I use the extension of the test statistic given in Section IV. When including a variable for the percentage of employment in manufacturing in the previous period, I obtain a contradicting result between the Hausman test and my test. There can be the cases where the model might have a functional form misspecification or the nonlinear correlation between the error terms U and V. For these two possible reasons, my nonparametric test can be used more accurately for testing for endogeneity. With the comparison in the estimates based on the test results, the potential bias can occur.

In summary, the proposed nonparametric test can be useful in many aspects since endogeneity can be present in different forms and contexts. In addition to a triangular simultaneous equations setting, I can observe endogeneity when having measurement errors in the data or when misspecifying the model by omitting a variable. In this regard, my test statistic can provide a generic approach to test endogeneity in other econometric problems in future research.

Chapter 4

A Nonparametric Panel Estimation for Random Effects and a Consistent Nonparametric Test for Endogeneity

4.1 Introduction

There are many interesting testing problems in panel data for model specification such as cross-sectional independence (Chen et al. (2012)), the time trend (Zhang et al. (2012)), and linearity (Lin et al. (2014)). Among them, endogeneity is important in determining the model specification. In a panel data setting, there are two strands of endogeneity—one is the endogeneity of a variable, and the other is the endogeneity of an individual fixed effect.

When ruling out the possibility of the former, testing for endogeneity in an individual fixed effect plays an important role in determining whether to use the fixed effects or the random effects model. Regarding this issue, there is a trade-off between consistency and efficiency. Under the null hypothesis, which is no endogeneity, the random effects model will be used. With the presence of endogeneity, the fixed effects model will be chosen, giving a consistent but less efficient estimator.

Despite the importance of detecting endogeneity of an individual fixed effect, there has been a difficulty in testing this hypothesis mainly because these individual fixed effects under the random effects model cannot be obtained. Unlike the fixed effects model where the individual fixed effects can be separately estimated, the random effects model is estimated by including individual fixed effects into the error terms. However, if the objective function is defined differently as in Huang et al. (2019), the individual fixed effects model can be estimated even under the random effects.

There are two main contributions of this paper. First, this paper develops a new estimation method to obtain individual-specific components under the random effects by extending Huang et al. (2019). While Huang et al. (2019) impose a parametric specification, I implement the nonparametric random effects panel estimation. This minimizes the chance of having endogeneity of a variable, which can be caused by the misspecification of a functional form. As I can obtain the individual-specific components under the random effects model, the accuracy of a test can be improved instead of testing with the residuals from the random effects model as a unity. Second, I construct a test by converting the moment condition of interest. Because the individual-specific components can be obtained, I construct a test over the i index instead of the i index and time. Then, the rate of convergence can be faster than the previous test statistics.

I set up the null hypothesis as no endogeneity against the presence of endogeneity. Then I construct a conditional moment test using kernel weighting (Li and Wang (1998), Hsiao and Li (2001), Henderson et al. (2008), Wang et al. (2018) among others). One difference is that I use the mean of a fixed variable over time instead of a fixed variable itself. The conditional moment test is simple to construct as it only requires the null hypothesis. If I construct a test using an estimation under the alternative, the nonparametric fixed effects estimation must be applied. Once constructing a test, I introduce a wild cluster bootstrap procedure using Mammen's distribution in a nonparametric random effects model to improve the finite-sample performance of a test. Wild cluster bootstrap is similar to a wild bootstrap, but it resamples clusters of residuals using a two-point distribution (See Wu (1983), Liu (1988), Mammen (1993), Davidson and Flachaire (2008) among others). There has been literature on wild cluster bootstrap such as its asymptotics as well as its procedure with different cluster sizes (See Cameron et al. (2008), Mackinnon et al. (2017), Djogbenou et al. (2019) among others).

As mentioned earlier, there is a trade-off between the random effects and the fixed effects model. Since the fixed effects estimators are consistent both under the null and the alternative, the nonparametric methods estimation methods are heavily based on fixed effects estimation. Regarding the nonparametric panel fixed effects estimation, there has been extensive literature from a standard panel estimation to a dynamic panel estimation (See Lin and Carroll (2000), Su and Ullah (2006), Su and Lu (2013), Lee and Robinson

(2015), Lee et al. (2019) among others). For the random effects model, the nonparametric panel random effects estimation methods have evolved toward obtaining a more efficient estimator (See Henderson and Ullah (2005), Mukherjee (2006), Ma et al. (2015) among others). While there exist extensive estimation methods for the panel, this paper's test for endogeneity can play a role as a preliminary step for the model specification.

Regarding the endogeneity tests, the most popular parametric test is the Hausman test. This compares the random effects estimates and the fixed effects estimates under the parametric specification. With the chronic problem of the Hausman test which suffers from the misspecification of a functional form, Henderson et al. (2008) constructed a Li-Wang type test. In constructing their nonparametric test, the residuals which include the individual fixed effects are used due to the difficulty of obtaining the individual-specific components under the null. This implies that this test improves the accuracy of a test compared to the Hausman test but may fail when the endogeneity of a variable occurs.

The paper is organized as follows. Section II introduces a model, hypotheses, estimation method, and the test statistic for endogeneity. In Section III, I conduct Monte Carlo simulations and compare the test results with the currently developed endogeneity tests. I then apply the test to the empirical data in Section IV. I test the endogeneity of an individual fixed effect by implementing it to the US state-level public capital data. Section V concludes the paper.

4.2 Model and Hypotheses

Consider a nonparametric random effects model for panel data as follows:

$$y_{it} = m(X_{it}) + u_{it}, \text{ where } u_{it} = \alpha_i + v_{it}, \; i = 1, ..., n \text{ and } t = 1, ..., T \tag{4.1}$$

In this model, y_{it} is an observable scalar random variable, $m(\cdot)$ is an unknown function, X_{it} is a $d_x \times 1$ vector of regressors, and u_{it} is the random disturbance such that $E[u_{it} \mid X_{it}] = 0$. The data is independent across the i index. For the disturbance term, I assume that $\alpha_i \sim i.i.d.\, N(0, \sigma_\alpha^2)$, $v_{it} \sim i.i.d.\, N(0, \sigma_v^2)$, α_i and v_{jt} are uncorrelated for $i, j = 1, \cdots, n$, and $\Sigma \equiv E[u_i u_i'] = \sigma_\alpha^2 I_T + \sigma_v^2 i_T i_T$, where I_T is a $T \times T$ identity matrix and i_T is a $T \times 1$ vector of ones.

Assuming the exogeneity of the random errors where $E[v_{it} \mid X_{i1}, ..., X_{iT}] = 0$ is satisfied, the assumption of interest for testing in this paper is whether the individual specific effects, α_i, is exogenous or not. The testing hypotheses are given as follows:

$$\begin{cases} \mathbb{H}_0 : E[\alpha_i \mid X_{it}] = 0 \\ \mathbb{H}_1 : E[\alpha_i \mid X_{it}] \neq 0 \end{cases}$$

Under the null, the random effects estimation can be implemented as there is no endogeneity problem in the estimation. Under the alternative, the fixed effects estimation will be applied due to an inconsistency problem of the random effects estimators. Testing for its presence is important because there is a consistency/efficiency trade-off between the random effects and the fixed effects estimators.

Since the individual-specific component varies over the i index only, the moment condition of interest under the null can be converted to another form, which is $E[\alpha_i \mid \bar{X}_i] = 0$, where $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$. The equivalence of the two conditions can be simply shown as below. Suppose $\bar{X}_i = r(X_{it})$ and there exists $g(\cdot)$ such that $E[\alpha_i \mid X_{it}] = g(r(X_{it}))$. Then,

$$
\begin{aligned}
E[\alpha_i \mid \bar{X}_i] &= E[E[\alpha_i \mid r(X_{it}), X_{it}] \mid r(X_{it})] \qquad (4.2)\\
&= E[E[\alpha_i \mid X_{it}] \mid r(X_{it})]\\
&= E[g(r(X_{it})) \mid r(X_{it})]\\
&= g(r(X_{it}))\\
&= E[\alpha_i \mid X_{it}]
\end{aligned}
$$

The second line of the equation (4.2) holds by the tower property of the conditional expectation as $\sigma(\bar{X}_i) \subset \sigma(X_{it})$. With this converted moment condition, I can set up new hypotheses for testing endogeneity of individual-specific components.

$$
\begin{cases}
\mathbb{H}_0 : E[\alpha_i \mid \bar{X}_i] = 0\\
\mathbb{H}_1 : E[\alpha_i \mid \bar{X}_i] \neq 0
\end{cases}
$$

One obstacle in directly testing for this null hypothesis is the difficulty in estimating α_i in the random effects model unlike the fixed effects model. Therefore, both the Hausman test and the nonparametric test of Henderson et al. (2008) for endogeneity are constructed by assuming $E[u_{it} \mid X_{it}] = 0$. However, Huang et al. (2019) set up a new objective function to obtain a random effects model estimator in a parametric context, which allows the derivation of the individual-specific component in a random effects panel model. By extending their estimation strategy to a nonparametric context, I can obtain the individual specific components in a nonparametric random effects panel model.

4.2.1 Estimation

Extending Huang et al. (2019), I define a new objective function to derive $(\alpha_i, m(x))$ under a nonparametric random effects panel data model. First, I re-write the model in the previous section as follows:

$$y = m(\underline{x}) + U, \text{ where } U = D\alpha + V \tag{4.3}$$

$y = (y_{11}, \cdots, y_{1T}, \cdots, y_{n1}, \cdots, y_{nT})$, $\underline{x} = (x_{11}, \cdots, x_{1T}, \cdots, x_{n1}, \cdots, x_{nT})$, $D = I_n \otimes \iota_T$, $\alpha = (\alpha_1, \cdots, \alpha_n)$, and $V = (v_{11}, \cdots, v_{1T}, \cdots, v_{n1}, \cdots, v_{nT})$. Then the objective function is introduced as

$$\begin{aligned}
&\underset{\alpha_i, m(x)}{\text{Min}} \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} \alpha_i^2 K\left(H^{-1}(X_{it} - x)\right)}{T\sigma_\alpha^2} + \frac{\sum_{i=1}^{n}\sum_{t=1}^{T} (y_{it} - m(x) - \alpha_i)^2 K\left(H^{-1}(X_{it} - x)\right)}{\sigma_v^2} \\
\Leftrightarrow &\underset{\alpha_i, m(x)}{\text{Min}} \frac{1}{T\sigma_\alpha^2} \alpha' D' W(x) D\alpha + \frac{1}{\sigma_v^2} \left(y - m(x) - D\alpha\right)' W(x) \left(y - m(x) - D\alpha\right),
\end{aligned}$$

where K is a non-negative d_x-variate kernel function, $H = diag(h_1, \cdots, h_{d_x})$ is a $d_x \times d_x$ bandwidth matrix that is symmetric and positive definite, $W(x) = diag(K(H^{-1}(X_{11} - x)), \cdots, K(H^{-1}(X_{1T} - x)), \cdots, K(H^{-1}(X_{n1} - x)), \cdots, K(H^{-1}(X_{nT} - x)))$, $\iota = (1, \cdots, 1)'$ is a $n \times 1$ matrix, $|H|$ is the determinant of H, and $m(x)$ is a nonparametric estimator.

There are three steps to obtain $\hat{m}(x)$. As the variance structure is unknown, there are additional steps to follow to estimate it. The procedure is as follows:

1. From the equation (4.3), obtain the usual LCLS estimator, $\tilde{m}(x)$, and $\tilde{u}_{it} = y_{it} - \tilde{m}(x)$, where $\tilde{m}(x) = (\iota' W(x)\iota)^{-1}\iota' W(x)y$. Note that this estimator is not a random effects estimator.

2. Define $\sigma_1^2 = \sigma_v^2 + T\sigma_\alpha^2$. Using the residuals $\tilde{u}_{it}$'s, both $\hat{\sigma}_1^2$ and $\hat{\sigma}_v^2$ can be simply obtained as follows[1]:

$$\begin{aligned}
\hat{\sigma}_1^2 &= \frac{T}{n}\sum_{i=1}^{n}\bar{\tilde{u}}_i^2 \\
\hat{\sigma}_v^2 &= \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\tilde{u}_{it} - \bar{\tilde{u}}_i\right)^2,
\end{aligned}$$

   where $\bar{\tilde{u}}_i = \sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{u}_{it}K(X_{it} - X_{is}) / \sum_{t=1}^{T}\sum_{s=1}^{T}K(X_{it} - X_{is})$. Once estimating $\hat{\sigma}_1^2$ and $\hat{\sigma}_v^2$, I can obtain $\hat{\alpha}_i$ using $\hat{\sigma}_\alpha^2 = \frac{1}{T}(\hat{\sigma}_1^2 - \hat{\sigma}_v^2)$.

3. With $\hat{\Sigma} = \hat{\sigma}_\alpha^2 I_T + \hat{\sigma}_v^2 i_T i_T$, the nonparametric random effects estimator $\hat{m}(x)$[2] and the individual specific effects $\hat{\alpha}$ can be estimated, which can be easily derived from the first order conditions from the objective function above.

$$\begin{aligned}
\hat{u}_{it} &= y_{it} - \hat{m}_{RE}(x) \\
\hat{\alpha}_i &= \frac{T\hat{\sigma}_\alpha^2}{T\hat{\sigma}_\alpha^2 + \hat{\sigma}_v^2}\bar{\hat{u}}_i, \text{ where } \bar{\hat{u}}_i = \frac{\sum_{t=1}^{T}\sum_{s=1}^{T}\hat{u}_{it}K(X_{it} - X_{is})}{\sum_{t=1}^{T}\sum_{s=1}^{T}K(X_{it} - X_{is})}.
\end{aligned}$$

[1] The consistency of these variance has been shown in Henderson and Ullah (2005).

[2] This can be obtained either from Henderson and Ullah (2005) or Su and Ullah (2007). For a more efficient estimator, I apply Su and Ullah (2007) random effects estimator.

4.2.2 Test Statistic and Its Asymptotic Properties

In this section, I construct a test statistic and derive its asymptotic properties.

$$
\begin{aligned}
y_{it} &= m(X_{it}) + u_{it} \\
&= m(X_{it}) + E[\alpha_i \mid X_{it}] + u_{it} - E[\alpha_i \mid X_{it}] \\
&= m(X_{it}) + E[\alpha_i \mid \bar{X}_i] + u_{it} - E[\alpha_i \mid \bar{X}_i] \\
&= m_1(X_{it}, \bar{X}_i) + \eta_{it}, \text{ where } \eta_{it} = u_{it} - E[\alpha_i \mid \bar{X}_i] \qquad (4.4)
\end{aligned}
$$

It is easy to show $E[\eta_{it} \mid X_{it}, \bar{X}_i] = E[\eta_{it} \mid X_{it}] = E[u_{it} - E[\alpha_{it} \mid \bar{X}_i] \mid X_{it}] = E[v_{it} \mid X_{it}] = 0$. The last equality holds by assumption. Using the equivalence of the moment condition from the previous section, I can construct a Li-Wang type test statistic. $E[\alpha_i \mid \bar{X}_i] = 0$ implies $E[\alpha_i E[\alpha_i \mid \bar{X}_i]] = 0$. Then, the conditional moment condition under the null implies $E[\alpha_i E[\alpha_i \mid \bar{X}_i] f(\bar{X}_i)] = 0$.

Theorem 7 *$E[\alpha_i \mid \bar{X}_i] = 0$ iff $E[f\left(\bar{X}_i\right) \alpha_i E[\alpha_i \mid \bar{X}_i]] = 0$, where $f(\cdot)$ is the density function of $\bar{X}_i$ that is bounded away from zero for all $\bar{X}_i$.*

Proof of Theorem I let $f\left(\bar{X}_i\right)$ and $f(\alpha|\bar{x})$ be the marginal density of $\bar{X}_i$ and the conditional density of α_i given $\bar{X}_i = \bar{x}$, respectively.

$$
\begin{aligned}
E\left[f\left(\bar{X}_i\right) \alpha_i E[\alpha_i|\bar{X}_i]\right] &= \iint \alpha_1 \left(\int \alpha_2 f\left(\bar{x}, \alpha_2\right) d\alpha_2\right) f\left(\bar{x}, \alpha_1\right) d\alpha_1 d\bar{x} \\
&= \int \left(\int \alpha_1 f\left(\alpha_1|\bar{x}\right) d\alpha_1\right) \left(\int \alpha_2 f\left(\alpha_2|\bar{x}\right) d\alpha_2\right) f^2\left(\bar{x}\right) d\bar{x} \\
&= \int \left(\int \alpha f\left(\alpha|\bar{x}\right) d\alpha\right)^2 f^2\left(\bar{x}\right) d\bar{x},
\end{aligned}
$$

since α_i is *i.i.d.* over i. Therefore, $E[\alpha_i \mid \bar{X}_i] = \int \alpha f(\alpha|\bar{x})\, d\bar{x} = 0$ iff $E[f(\bar{X}_i)\, \alpha_i E[\alpha_i \mid \bar{X}_i]] = 0$ since $f(\bar{x}) > 0$. ■

Define $f(\bar{X}_i) = \frac{1}{n|\bar{H}|}\sum_{j=1}^{n} K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)$, where $\bar{H} = diag(\bar{h}_1, \cdots, \bar{h}_{d_x})$ is a $d_x \times d_x$ bandwidth matrix that is symmetric and positive definite, $|\bar{H}|$ is the determinant of $\bar{H}$. These bandwidths are different from the bandwidths for the estimation because the regressors are $\bar{X}_i$, not X_{it}. Throughout the paper, the bandwidths for testing are calculated using the rule-of-thumb. The test statistic I_n is obtained as follows:

$$
\begin{aligned}
I_n &= E[\hat{\alpha}_i E[\hat{\alpha}_i \mid \bar{X}_i] f(\bar{X}_i)] \\
&= \frac{1}{n}\sum_{i=1}^{n}\hat{\alpha}_i\left(\frac{1}{(n-1)\left|\bar{H}\right|}\sum_{j\neq i}^{n}\hat{\alpha}_j K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)\right) \\
&= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{\alpha}_i\hat{\alpha}_j K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)
\end{aligned}
$$

For characterizing the asymptotic distribution, the following assumptions will be used.

(A1) $\{U_i, X_i\}_{i=1}^{n}$ is *i.i.d.*, where $U_i = (u_{i1}, \cdots, u_{iT})'$ and $X = (X_{i1}, \cdots, X_{iT})'$.

(A2) $E[u \mid x] = 0$, $\sigma^2(x) = E[u^2 \mid x]$, $\sigma^2(x)$ is continuous at x and $E[\sigma^2(x)] < \infty$.

The model assumes the *i.i.d.* distribution of $\{U_i, X_i\}_{i=1}^{n}$. Also, as my interest lies in testing the endogeneity of α, I assume the exogeneity of V. The conditional variance $\sigma^2(x)$ is continuous at x and its expectation is finite. I assume the homoskedasticity for the conditional variance.

(A3) $f(x)$ is differentiable, $0 < f(x) \leq B_f < \infty$, and $|f(x) - f(x')| < m_f |x - x'|$ for some $0 < m_f < \infty$ is satisfied.

(A4) The kernel function $K(\cdot)$ is bounded and symmetric density function with compact support such that $\int K(\psi)d\psi = 1$. For $\forall x \in \mathbb{R}$, $|K(x)| < B_k < \infty$. I assume $\left|K^j(u) - K^j(v)\right| \leq C_1 |u - v|$, for $j = 0, 1, 2, 3$.

In (A3), the conditional density $f(x)$ satisfies the Lipschitz continuous condition. In addition, as it is smooth and bounded, a Taylor expansion can be applied. In constructing my test, I use the kernel function as a weighting function. Regarding properties of the kernel function, it is bounded and symmetric. As in $f(x)$, the kernel function satisfies the Lipschitz continuous function.

(A5) As $n \to \infty$, $|H|, \left|\bar{H}\right| \to 0$. It satisfies $n\left|\bar{H}\right|^2 \to \infty$, $n\left|\bar{H}\right|^6 \to 0$, and $n|H|/\ln n \to \infty$.

(A6) $m(\cdot)$ is continuous and twice differentiable in X respectively.

The assumption (A5) is on the restriction of the bandwidth for the asymptotic properties of the proposed test statistic and the standard assumptions for nonparametric estimation. The test statistic then can be standardized as follows:

$$T_n = n\left|\bar{H}\right|^{1/2} I_n/\sqrt{\hat{\Omega}}, \text{where } \hat{\Omega} = \frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{\alpha}_i^2\hat{\alpha}_j^2 K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right).$$

Theorem 8 *Under $\mathbb{H}_0$, as $\hat{\Omega}$ is a consistent estimator of $\Omega = 2[\int K^2(\psi)d\psi]E[\sigma^4(\bar{X})f(\bar{X})]$,*

$$T_n \to N(0,1) \textit{ as } n \to \infty.$$

For the asymptotic properties under the alternative, I introduce the Pitman local alternatives as follows:

$$\mathbb{H}_1(\delta_n) : m_1(X_{it}\bar{X}_i) = m(X_{it}) + \delta_n l(\bar{X}_i)\,,$$

where $l(\cdot)$ is continuously differentiable and bounded, and $\delta_n = n^{-1/2}\left|\bar{H}\right|^{-1/4}$. Based on the equation (4.4), note that $l(\cdot)$ does not include the elements of X_{it} because $m_1(X_{it}, \bar{X}_i)$ is separable by construction of the model.

Theorem 9 *Under the Pitman local alternative, if* $\delta_n = n^{-1/2}\left|\bar{H}\right|^{-1/4}$*, we have*

$$T_n \xrightarrow{d} N(E[l(\bar{X}_i)^2 f(\bar{X}_i)]/\sqrt{\Omega}, 1) \ \text{as } n \to \infty.$$

Then, as the magnitude of $E[l(\bar{X}_i)^2 f(\bar{X}_i)]/\sqrt{\Omega}$ increases, the test statistic deviates farther from the zero mean, and the local power increases. However, the variance remains at one for both hypotheses.

Theorem 10 *Assuming (A1)-(A6) and under* $\mathbb{H}_1$*, Pr[*$\hat{T}_n > B_n$*]* $\to 1$ *for any non-stochastic sequence* $\{B_n : B_n = o(\sqrt{n^2\left|\bar{H}\right|})\}$*.* $\hat{I}_n = I_n + o_p((n\left|\bar{H}\right|^{1/2})^{-1})$*, where* $I_n = E[(l(\bar{X}_i))^2 f(\bar{X}_i)]$*, and* $\hat{\Omega} = \Omega + o_p(1)$*.*

Theorem 10 suggests the consistency of the test statistic. Under $\mathbb{H}_1$, the probability of rejecting the null will converge to 1.

4.2.3 Bootstrap Procedure

In order to increase the test's performance in a finite sample, I introduce a wild cluster bootstrap for the test statistic as below.

1. Estimate $\hat{m}(x)$ and $\hat{\alpha}_i$ for a nonparametric random effects model following the estimation method given in Section 4.2.1.

2. Generate u_i^* as the wild bootstrap error, where $u_i^* = (u_{i1}^*, \cdots, u_{iT}^*)'$. I construct $u_i^* = \frac{1-\sqrt{5}}{2}\hat{u}_i$ with the probability of $\frac{1+\sqrt{5}}{2\sqrt{5}}$ and $u_i^* = \frac{1+\sqrt{5}}{2}\hat{u}_i$ with the probability of $1 - \frac{1+\sqrt{5}}{2\sqrt{5}}$, where $\hat{u}_i = (\hat{u}_{i1}, \cdots, \hat{u}_{iT})'$ and $\hat{u}_{it} = y_{it} - \hat{m}(x)$. It is easy to show $E\left[u_{it}^*\right] = 0$, $E\left[u_{it}^{*2}\right] = \hat{u}_{it}^2$, and $E\left[u_{it}^{*3}\right] = \hat{u}_{it}^3$.

3. Generate y_{it}^*, where $y_{it}^* = \hat{m}(x) + \hat{u}_{it}^*$ under the null hypothesis.

4. Use the bootstrap sample $\{y_i^*, X_i\}_{i=1}^n$ to obtain $\hat{m}^*(x)$, and get $\hat{\alpha}_i^* = (T\hat{\sigma}_\alpha^2/(T\hat{\sigma}_\alpha^2 + \hat{\sigma}_v^2))\bar{\hat{u}}_i^*$.

5. With $\{\hat{\alpha}_i^*, X_i\}_{i=1}^n$, compute the bootstrap test statistic T_n^* and repeat above procedure for B times. In this simulation, the number of bootstrapping is 300.

6. Based on the empirical distribution of T_n^*, calculate the critical value c^* and obtain the p-value, which is $P(T_n \geq c^*)$. If p-value is less than 0.05 at 5% significance level, we reject the null.

Following these bootstrap procedures, I can obtain the asymptotic distribution of T_n^*. I will show how the bootstrap test performs in the Monte Carlo Simulations. The asymptotic distribution of bootstrap test under the null is shown in Theorem 11.

Theorem 11 *Define* $\hat{\Omega}^* = \frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^n \sum_j^n \hat{\alpha}_i^{*2} \hat{\alpha}_j^{*K} \left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)$. *Let the bootstrap test statistic be* $T_n^* = n\left|\bar{H}\right|^{1/2} I_n^*/\sqrt{\hat{\Omega}^*}$, *where Under* $\mathbb{H}_0^*$, *as* $\hat{\Omega}^*$ *is a consistent estimator of* $\Omega = 2[\int K^2(\psi)d\psi]E[f(\bar{x})\sigma^4(\bar{x})]$, $T_n^* \to N(0,1)$ *in distribution as* $n \to \infty$.

The proofs of Theorem 11 will follow similarly to those of Theorem 8. In addition, when the null hypothesis is false, $P(T_n > T_n^*) \to 1$ asymptotically, which shows the consistency of the bootstrap test statistic.

4.3 Simulations

4.3.1 Data Generating Processes

I perform the test for endogeneity using two different data generating processes. I follow Henderson et al. (2008) for DGP_2.

$$DGP_1 : \begin{cases} Y_{it} = 1 + 2X_{it} - X_{it}^2 + \alpha_i + v_{it}, v_{it} \sim i.i.d.N(0,1) \\ \alpha_i = \xi_i + \theta(\bar{X}_i + 2\bar{X}_i)^3 \end{cases}$$

$$DGP_2 : \begin{cases} Y_{it} = \sin(2X_{it}) + \alpha_i + v_{it}, v_{it} \sim i.i.d.N(0,1) \\ \alpha_i = \xi_i + \theta\bar{X}_i, \end{cases}$$

where X_{it} is $i.i.d.$ $U[-1,1]$, $\xi_i \sim i.i.d.U(0,1)$, and $\bar{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$.

For both data generating processes, $\theta = 0$, 0.4, and 0.8, which indicates no endogeneity, weak endogeneity, and strong endogeneity, respectively. For DGP_1, the model is correctly specified in both the parametric and nonparametric specification but the correlation between α_i and $\bar{X}_i$ is nonlinear. For DGP_2 follows the specification from Wang (2003). In this case, the model is nonlinearly specificed while the correlation between α_i and $\bar{X}_i$ is linear. For bandwidth selection, I use rule-of-thumb bandwidths for both the estimation and the test. For the estimation, I use local linear estimation with a second-order

Gaussian kernel by using a rule-of-thumb bandwidth, which is $h = 1.06std(X_{it})(nT)^{-1/5}$. I obtain $\hat{m}(X_i)$ from $\hat{m}(x) = \arg\max_{m(x)} \frac{1}{T\sigma_\alpha^2}\sum_{i=1}^n \sum_{t=1}^n \alpha_i^2 K(\frac{X_{it}-x}{h}) + \frac{1}{\sigma_v^2}\sum_{i=1}^n \sum_{t=1}^n (y_{it} - m(x) - \alpha_i)^2 K(\frac{X_{it}-x}{h})$. For the test bandwidth, $\bar{h} = c \cdot std(X_i)n^{-1/5}$ and $c = 0.5,\ 1.06,\ 1.5$. The number of repetition is 1000 for both sample sizes of 50 and 100. T is fixed as 3 throughout the simulations. The number of bootstrap repetitions is 300 for both samples.

4.3.2 Simulation Results

For each data generating process, both the size and the power are estimated by changing the strength of endogeneity (the value of θ). I then compare my test's performance with the Hausman test (H_n) and the Li-Wang type test (J_n) using the α_i's from the nonparametric fixed effects estimation. In this paper, I use the Su and Ullah (2006) nonparametric fixed effects estimation method. The Hausman test is a parametric test, where it measures the difference between fixed effects panel and random effects estimates. First, Table 4.1-4.3 represent both size and power for each data generating process with different values of bandwidth. Both nonparametric tests' performance varies with the bandwidth. In addition, I apply a wild cluster bootstrap procedure for both test statistic.

Table 4.1 and 4.2 presents the size and power when the model is correctly specified but the endogenous correlation is nonlinear. The bootstrap size of my conditional moment test is close to the correct size at each significance level although T_n is undersized in the asymptotic test. For different bandwidths, their estimated size is close to the nominal size in all significance levels and its performance improves with the increase in size. For the fixed

Table 4.1: Size of Each Test

			1%			5%			10%		
	n	c	0.5	1.06	1.5	0.5	1.06	1.5	0.5	1.06	1.5
DGP_1	50	T_n^*	0.012	0.010	0.020	0.044	0.054	0.048	0.082	0.094	0.094
		T_n	0.010	0.004	0.004	0.022	0.010	0.008	0.038	0.016	0.010
		J_n^*	0.002	0.004	0.004	0.032	0.028	0.022	0.072	0.070	0.072
		J_n	0.004	0.006	0.006	0.020	0.006	0.008	0.038	0.014	0.010
		H_n	0.024	0.024	0.024	0.052	0.052	0.052	0.094	0.094	0.094
	100	T_n^*	0.016	0.014	0.014	0.062	0.070	0.064	0.116	0.122	0.138
		T_n	0.010	0.004	0.002	0.022	0.010	0.004	0.060	0.016	0.012
		J_n^*	0.004	0.002	0.000	0.040	0.046	0.038	0.096	0.096	0.080
		J_n	0.002	0.002	0.002	0.020	0.014	0.006	0.050	0.030	0.010
		H_n	0.002	0.002	0.002	0.036	0.036	0.036	0.084	0.084	0.084
DGP_2	50	T_n^*	0.012	0.006	0.012	0.034	0.036	0.030	0.082	0.080	0.070
		T_n	0.002	0.000	0.000	0.018	0.002	0.002	0.034	0.004	0.002
		J_n^*	0.002	0.004	0.004	0.030	0.032	0.024	0.076	0.076	0.080
		J_n	0.004	0.006	0.006	0.020	0.006	0.008	0.038	0.014	0.010
		H_n	0.014	0.014	0.014	0.060	0.060	0.060	0.098	0.098	0.098
	100	T_n^*	0.012	0.014	0.004	0.042	0.050	0.050	0.112	0.110	0.110
		T_n	0.004	0.000	0.000	0.018	0.000	0.000	0.044	0.008	0.000
		J_n^*	0.004	0.004	0.004	0.040	0.046	0.040	0.088	0.094	0.090
		J_n	0.002	0.002	0.002	0.020	0.016	0.006	0.050	0.030	0.010
		H_n	0.002	0.002	0.002	0.034	0.034	0.034	0.088	0.088	0.088

Note: c is for different bandwidth sizes, T_n^* is a bootstrap test, T_n is an asymptotic test, J_n^* is a bootstrap test for fixed-effects α_i, J_n is an asymptotic test for fixed-effects α_i, and H_n is the Hausman test.

Table 4.2: Power of Each Test under DGP_1

			1%			5%			10%		
θ	n	c	0.5	1.06	1.5	0.5	1.06	1.5	0.5	1.06	1.5
0.4	50	T_n^*	0.018	0.022	0.024	0.056	0.068	0.090	0.096	0.138	0.156
		T_n	0.008	0.006	0.004	0.034	0.018	0.014	0.060	0.026	0.028
		J_n^*	0.012	0.028	0.034	0.062	0.090	0.102	0.132	0.180	0.216
		J_n	0.026	0.034	0.032	0.062	0.066	0.060	0.102	0.098	0.082
		H_n	0.148	0.148	0.148	0.298	0.298	0.298	0.382	0.382	0.382
	100	T_n^*	0.018	0.026	0.042	0.074	0.088	0.112	0.122	0.154	0.190
		T_n	0.016	0.014	0.010	0.048	0.040	0.032	0.078	0.062	0.060
		J_n^*	0.070	0.116	0.122	0.134	0.220	0.250	0.224	0.286	0.362
		J_n	0.082	0.118	0.128	0.142	0.172	0.178	0.204	0.222	0.224
		H_n	0.312	0.312	0.312	0.504	0.504	0.504	0.640	0.640	0.640
0.8	50	T_n^*	0.034	0.060	0.098	0.104	0.176	0.208	0.170	0.272	0.328
		T_n	0.036	0.032	0.032	0.070	0.074	0.064	0.102	0.114	0.106
		J_n^*	0.108	0.200	0.252	0.300	0.422	0.498	0.416	0.528	0.590
		J_n	0.206	0.290	0.300	0.298	0.386	0.394	0.386	0.452	0.452
		H_n	0.620	0.620	0.620	0.778	0.778	0.778	0.846	0.846	0.846
	100	T_n^*	0.066	0.148	0.216	0.190	0.352	0.444	0.298	0.498	0.584
		T_n	0.076	0.118	0.130	0.158	0.220	0.238	0.224	0.294	0.332
		J_n^*	0.402	0.602	0.706	0.638	0.810	0.874	0.744	0.908	0.924
		J_n	0.504	0.674	0.708	0.648	0.778	0.814	0.712	0.840	0.864
		H_n	0.954	0.954	0.954	0.982	0.982	0.982	0.990	0.990	0.990

Note: c is for different bandwidth sizes, T_n^* is a bootstrap test, T_n is an asymptotic test, J_n^* is a bootstrap test for fixed-effects α_i, J_n is an asymptotic test for fixed-effects α_i, and H_n is the Hausman test.

Table 4.3: Power of Each Test under DGP_2

			1%			5%			10%		
θ	n	c	0.5	1.06	1.5	0.5	1.06	1.5	0.5	1.06	1.5
0.4	50	T_n^*	0.016	0.022	0.032	0.044	0.066	0.092	0.098	0.132	0.148
		T_n	0.002	0.000	0.000	0.024	0.012	0.006	0.052	0.020	0.016
		J_n^*	0.008	0.014	0.016	0.064	0.050	0.058	0.104	0.102	0.114
		J_n	0.010	0.018	0.014	0.038	0.032	0.028	0.068	0.050	0.042
		H_n	0.052	0.052	0.052	0.146	0.146	0.146	0.228	0.228	0.228
	100	T_n^*	0.018	0.028	0.058	0.078	0.114	0.138	0.118	0.182	0.224
		T_n	0.014	0.012	0.002	0.046	0.034	0.026	0.066	0.048	0.040
		J_n^*	0.034	0.064	0.068	0.090	0.138	0.150	0.152	0.196	0.220
		J_n	0.052	0.068	0.062	0.080	0.094	0.092	0.110	0.130	0.116
		H_n	0.100	0.100	0.100	0.220	0.220	0.220	0.338	0.338	0.338
0.8	50	T_n^*	0.030	0.056	0.084	0.078	0.150	0.214	0.146	0.250	0.320
		T_n	0.014	0.016	0.014	0.054	0.042	0.034	0.076	0.064	0.056
		J_n^*	0.046	0.066	0.088	0.142	0.224	0.274	0.230	0.336	0.380
		J_n	0.082	0.100	0.084	0.152	0.178	0.178	0.204	0.250	0.234
		H_n	0.254	0.254	0.254	0.414	0.414	0.414	0.546	0.546	0.546
	100	T_n^*	0.060	0.136	0.216	0.160	0.306	0.406	0.272	0.430	0.522
		T_n	0.058	0.076	0.066	0.104	0.160	0.148	0.164	0.202	0.202
		J_n^*	0.164	0.274	0.332	0.316	0.472	0.562	0.422	0.604	0.660
		J_n	0.206	0.298	0.310	0.320	0.420	0.430	0.386	0.486	0.522
		H_n	0.524	0.524	0.524	0.742	0.742	0.742	0.842	0.842	0.842

Note: c is for different bandwidth sizes, T_n^* is a bootstrap test, T_n is an asymptotic test, J_n^* is a bootstrap test for fixed-effects α_i, J_n is an asymptotic test for fixed-effects α_i, and H_n is the Hausman test.

effects based test, the test is oversized. For the Hausman test, the size of the test is close to the correct size, but still undersized. As the level of endogeneity becomes stronger, the power of all test increases. My conditional moment test does not reject the null hypothesis as much as the other two tests do, and this may lie in the fact that the individual-specific components are obtained with the $T = 3$ observations.

Table 4.1 and 4.3 presents the size and power when the model is mis-specified but the endogenous correlation is linear. In this case, my nonparametric bootstrap size is close to the nominal size when the sample size is 100. As the strength of endogeneity increases, the test becomes more powerful. However, as the endogenous correlation is linear, the Hausman test performs the best among the three tests.

In summary, even though I observed the undersized test for T_n using asymptotic critical values, the estimated size based on bootstrap procedure is close to the nominal size for all the data generating processes. As the strength of endogeneity increases, the test becomes more powerful. Furthermore, I can compare clearly the nonparametric tests' performance when constructing a test between using individual-specific components only and using random effects residuals. In this regard, the current test dominates all the other tests in terms of accuracy in capturing the endogeneity of individual-specific components by kernel techniques to capture the local correlation.

4.4 Application

In this section, I analyze the productivity of public capital in the economy. In a large literature, there has been a deviate over whether the public capital productivity contributes to the private sector, but the focus of my application lies in whether the model is correctly specified–between fixed effects and random effects.

Following Baltagi and Pinnoi (1995), Henderson and Ullah (2005), and Su et al. (2013), I consider the following one-way random effects nonparametric model:

$$\log(Y_{it}) = m(\log(KG_{it}), \log(KPR_{it}), \log(L_{it}), UNEM_{it}) + \alpha_i + v_{it},$$

where $i = 1, \cdots, 48$, $t = 1, \cdots, 17$, Y_{it} denotes the GDP of state i in period t, KG denotes public capital, KPR denotes the prival captial stock estimated from the Bureau of Economics Analysis, L is employment, and UNEM stands for the unemployment rate used to control for business cycle effects. The panel data is for the US 48 continuous states over the period 1970-1986. For the test bandwidth of my test, I use $h = c \cdot std(\bar{X}_i)n^{-1/5}$ and $c = 0.5,\ 1.06,\ 1.5$ for using the second-order Gaussian kernel. The number of bootstrap repetition is 300.

Before presenting the test results, Figure 4.1 and Figure 4.2 show the correlation between individual-specific components and each variable from the panel random effects estimation. From both figures, it is easily to note that there is almost no correlation detected. In Figure 4.3, there seems to be nonzero correlation between the individual-specific components and the unemployment rate, the correlation is almost zero in value.

The test results are given in Table 4.4. The null hypothesis is rejected both in asymptotic and bootstrap test at 5% significance level. This result is also consistent with the test using $\hat{\alpha}_{FE}$'s. However, I have a contradicting test result with Hausman test. All nonparametric tests do not reject the null hypothesis with the high p-values while the Hausman test still rejects the null hypothesis at 5% significance level.

Table 4.4: P-Values of Each Test

c	T_n^*	T_n	J_n^*	J_n	H_n
0.5	0.470	0.677	0.357	0.481	0.000
1.06	0.520	0.578	0.547	0.840	0.000
1.5	0.567	0.530	0.560	0.842	0.000

Note: c is for different bandwidth sizes, T_n^* is a bootstrap test, T_n is an asymptotic test, J_n^* is a bootstrap test for fixed-effects α_i, J_n is an asymptotic test for fixed-effects α_i, and H_n is the Hausman test.

4.5 Conclusion

In a panel data, testing for endogeneity in an individual-specific component plays an important role in determining whether to use the fixed effects or the random effects model. In this regard, testing for the presence of endogeneity of individual-specific components cannot be underestimated due to consistency and efficiency issues. In this paper, I propose a consistent nonparametric test for endogeneity.

By introducing a new objective function to obtain a nonparametric random effects estimator, I can extract the individual-specific components in the random effects and construct a new nonparametric test for endogeneity. I can convert the conditional moment for endogeneity test $E[\alpha \mid X] = 0$ to $E[\alpha \mid \bar{X}] = 0$. Based on the modified moment condition, I construct a Li-Wang type test which follows the standard normal distribution asymptotically under the null hypothesis.

Figure 4.1: Correlation between Individual-Specific Effects and Each Variable from Parametric Regression

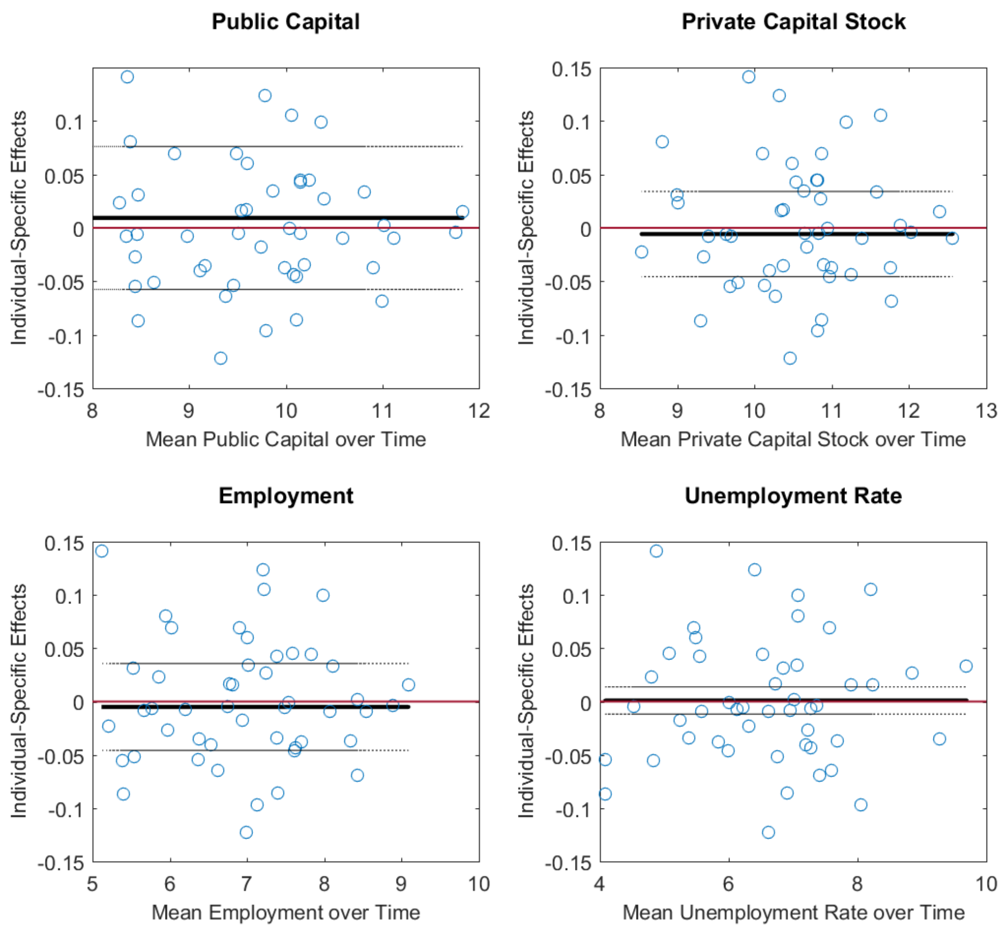


Figure 4.2: Correlation between Individual-Specific Effects and Each Variable from Parametric Regression

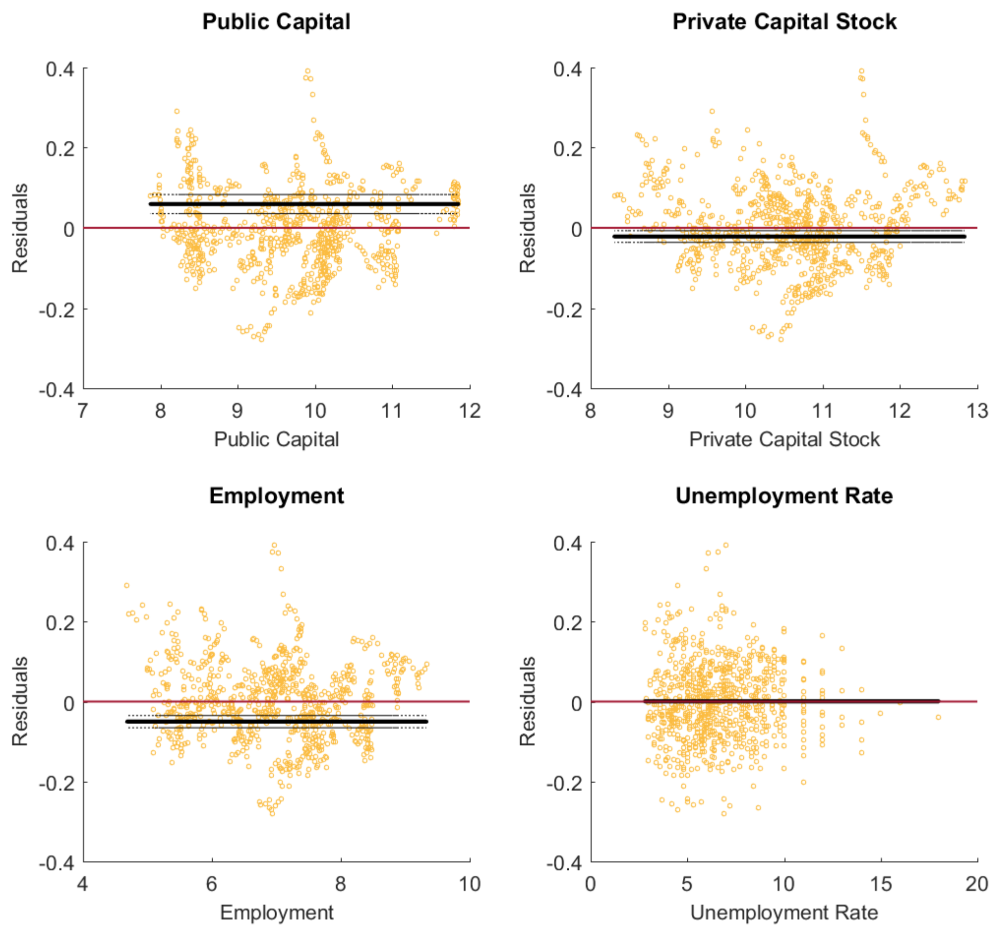
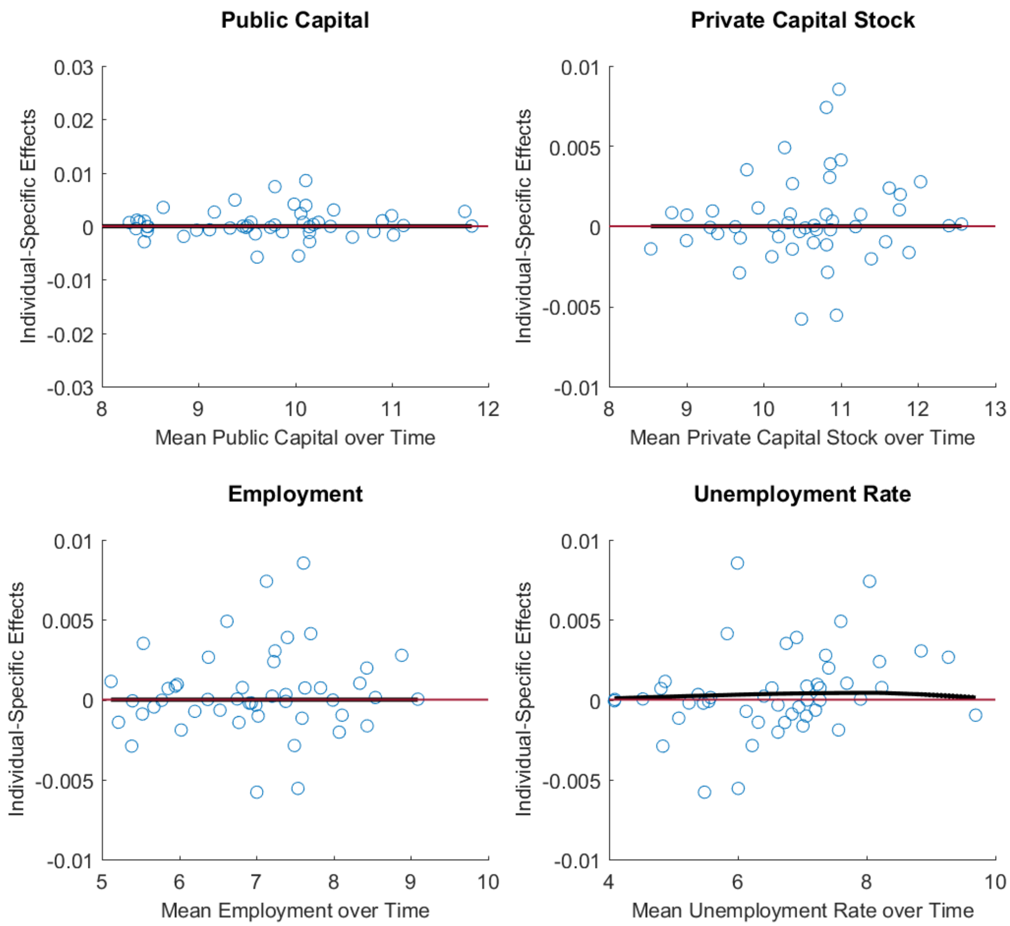


Figure 4.3: Correlation between Individual-Specific Effects and Each Variable from Non-parametric Regression



As with other nonparametric conditional moment tests (Zheng (1996), Li and Wang (1998), Hsiao and Li (2001), among others), I introduce a wild cluster bootstrap method using Mammen's distribution to improve the finite-sample performance. In simulations, I show that my bootstrap test performs better in finite samples than the asymptotic test for both size and power. Compared to the Hausman and fixed-effects individual specific components tests, my test statistic outperforms them when the individual-specific terms are not correlated but error terms are correlated with a variable. This addresses the advantage of my test by using the estimated individual-specific components instead of the residuals which include both individual-specific components as well as the random errors.

I also apply this test to the empirical analysis anayzing the effect of public capital to the economy. Following the same specification of Su et al. (2013), I tested for endogeneity and I obtain a contradicting result between the Hausman test and the nonparametric tests including my test. There can be the cases where the model might have a functional form misspecification or the nonlinear correlation between the individual-specific components and the variables a and X. For these two possible reasons, my nonparametric test can be used more accurately for testing for endogeneity.

Chapter 5

Conclusions

A hypothesis testing is one of the key components of the econometric analysis to analyze the effect with the statistical significance. However, nonparametric hypothesis testing has not been developed systematically and the current nonparametric tests are ad-hoc. The objective of this dissertation is to explore these issues and provide results on testing linearity as an illustration and develop new nonparametric tests for endogeneity in a cross-sectional data and a panel data respectively.

In Chapter 2, I analyze the relationship of the current nonparametric tests for linearity. By imposing some conditions, I obtained which test is locally most powerful both analytically and numerically. I can compare the local power of each test asymptotically. Furthermore, by developing a nonparametric Rao-Score test for the model specification, I show its equivalence to Su-Ullah type goodness-of-fit test.

Chapter 3 develops a consistent nonparametric test for endogeneity under a triangular simultaneous equations model. After taking the control function approach, I use nonparametric residuals in constructing a test to obtain the consistency of the test. My test has strengths in that it is easy to implement as its asymptotic distribution is the standard normal and it can capture the locally nonlinear correlation with kernel weighting.

Chapter 4 proposes a new estimation method for the nonparametric panel random effects model and develops a new nonparametric test for endogeneity. Extending Huang et al. (2019) to a nonparametric context, the individual fixed effects model can be estimated under the random effects and they are used to construct a test. With a large T, the test performs well in terms of size and power.

Bibliography

[1] Joshua Angrist and William Evans. Children and their parents' labor supply: Evidence from exogenous variation in family size. *American Economic Review*, 88(3):450–77, 1998.

[2] H Autor, David Dorn, and Gordon H Hanson. American economic review, forthcoming the china syndrome: Local labor market effects of import competition in the united states.

[3] Badi H Baltagi and Nat Pinnoi. Public capital stock and state productivity growth: Further evidence from an error components model. *Empirical Economics*, 20(2):351–359, 1995.

[4] Steven T Berry and Philip A Haile. Identification of nonparametric simultaneous equations models with a residual index structure. *Econometrica*, 86(1):289–315, 2018.

[5] Richard Blundell and Joel L Horowitz. A non-parametric test of exogeneity. *The Review of Economic Studies*, 74(4):1035–1058, 2007.

[6] Richard Blundell and James L Powell. Endogeneity in nonparametric and semiparametric regression models. *Econometric society monographs*, 36:312–357, 2003.

[7] Christoph Breunig. Goodness-of-fit tests based on series estimators in nonparametric instrumental regression. *Journal of Econometrics*, 184(2):328–346, 2015.

[8] A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008.

[9] David Card. Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160, 2001.

[10] Jia Chen, Jiti Gao, and Degui Li. A new diagnostic test for cross-section uncorrelatedness in nonparametric panel data models. *Econometric Theory*, 28(5):1144–1163, 2012.

[11] Xiaohong Chen and Demian Pouzo. Sieve wald and qlr inferences on semi/nonparametric conditional moment models. *Econometrica*, 83(3):1013–1079, 2015.

[12] John Coglianese, Lucas W Davis, Lutz Kilian, and James H Stock. Anticipation, tax avoidance, and the price elasticity of gasoline demand. *Journal of Applied Econometrics*, 32(1):1–15, 2017.

[13] Serge Darolles, Yanqin Fan, Jean-Pierre Florens, and Eric Renault. Nonparametric instrumental regression. *Econometrica*, 79(5):1541–1565, 2011.

[14] Mitali Das, Whitney K Newey, and Francis Vella. Nonparametric estimation of sample selection models. *The Review of Economic Studies*, 70(1):33–58, 2003.

[15] Russell Davidson and Emmanuel Flachaire. The wild bootstrap, tamed at last. *Journal of Econometrics*, 146(1):162–169, 2008.

[16] Russell Davidson and James G MacKinnon. Wild bootstrap tests for iv regression. *Journal of Business & Economic Statistics*, 28(1):128–144, 2010.

[17] Antoine A Djogbenou, James G MacKinnon, and Morten Ørregaard Nielsen. Asymptotic theory and wild bootstrap inference with clustered errors. *Journal of Econometrics*, 212(2):393–412, 2019.

[18] Kjell Doksum, Alexander Samarov, et al. Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *The Annals of Statistics*, 23(5):1443–1473, 1995.

[19] B Efron. Bootstrap methods: Another look at the jackknife. ann. stat. 1979.

[20] Jianqing Fan and Irene Gijbels. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, volume 66. CRC Press, 1996.

[21] Jianqing Fan and Jiancheng Jiang. Nonparametric inference with generalized likelihood ratio tests. *Test*, 16(3):409–444, 2007.

[22] Jianqing Fan and Qiwei Yao. *Nonlinear time series: nonparametric and parametric methods*. Springer Science & Business Media, 2008.

[23] Jianqing Fan, Chunming Zhang, Jian Zhang, et al. Generalized likelihood ratio statistics and wilks phenomenon. *The Annals of Statistics*, 29(1):153–193, 2001.

[24] Yanqin Fan and Qi Li. A consistent model specification test based on the kernel sum of squares of residuals. *Econometric Reviews*, 21(3):337–352, 2002.

[25] Yanqin Fan and Aman Ullah. On goodness-of-fit tests for weakly dependent processes using kernel method. *Journal of Nonparametric Statistics*, 11(1-3):337–360, 1999.

[26] Pedro L Gozalo. A consistent model specification test for nonparametric estimation of regression function models. *Econometric Theory*, 9(3):451–477, 1993.

[27] William H Greene. *Econometric analysis*. Pearson Education India, 2012.

[28] Jinyong Hahn, Zhipeng Liao, and Geert Ridder. Nonparametric two-step sieve m estimation and inference. *Econometric Theory*, pages 1–44, 2018.

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

[30] Wolfgang Hardle, Enno Mammen, et al. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*, 21(4):1926–1947, 1993.

[31] Jerry A Hausman. Specification tests in econometrics. *Econometrica: Journal of the econometric society*, pages 1251–1271, 1978.

[32] Daniel J Henderson, Raymond J Carroll, and Qi Li. Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics*, 144(1):257–275, 2008.

[33] Daniel J Henderson and Aman Ullah. A nonparametric random effects estimator. *Economics Letters*, 88(3):403–407, 2005.

[34] Javier Hidalgo. A nonparametric conditional moment test for structural stability. *Econometric Theory*, 11(4):671–698, 1995.

[35] Yongmiao Hong, Yoon-Jin Lee, et al. A loss function approach to model specification testing and its relative efficiency. *The Annals of Statistics*, 41(3):1166–1203, 2013.

[36] Yongmiao Hong, Halbert White, et al. Consistent specification testing via nonparametric series regression. *ECONOMETRICA-EVANSTON ILL-*, 63:1133–1133, 1995.

[37] Joel L Horowitz. A smoothed maximum score estimator for the binary response model. *Econometrica: journal of the Econometric Society*, pages 505–531, 1992.

[38] Joel L Horowitz. The bootstrap. In *Handbook of econometrics*, volume 5, pages 3159–3228. Elsevier, 2001.

[39] Joel L Horowitz. The bootstrap in econometrics. *Statistical Science*, pages 211–218, 2003.

[40] Joel L Horowitz. Applied nonparametric instrumental variables estimation. *Econometrica*, 79(2):347–394, 2011.

[41] Joel L Horowitz and Wolfgang Härdle. Testing a parametric model against a semiparametric alternative. *Econometric theory*, 10(5):821–848, 1994.

[42] Joel L Horowitz and Vladimir G Spokoiny. An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, 69(3):599–631, 2001.

[43] Cheng Hsiao and Qi Li. A consistent test for conditional heteroskedasticity in time-series regression models. *Econometric Theory*, 17(1):188–221, 2001.

[44] Bai Huang, Tae-Hwy Lee, and Aman Ullah. A combined random effect and fixed effect forecast for panel data models. *Journal of Management Science and Engineering*, 4(1):28–44, 2019.

[45] Li-Shan Huang, Jianwei Chen, et al. Analysis of variance, coefficient of determination and f-test for local polynomial regression. *The Annals of Statistics*, 36(5):2085–2109, 2008.

[46] Hidehiko Ichimura. Semiparametric least squares (sls) and weighted sls estimation of single-index models. 1991.

[47] Guido W Imbens and Whitney K Newey. Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512, 2009.

[48] Maximilian Kasy. Identification in triangular systems using control functions. *Econometric Theory*, 27(3):663–671, 2011.

[49] Roger W Klein and Richard H Spady. An efficient semiparametric estimator for binary response models. *Econometrica: Journal of the Econometric Society*, pages 387–421, 1993.

[50] Narayana R Kocherlakota and Kei-Mu Yi. A simple time series test of endogenous vs. exogenous growth models: An application to the united states. *The Review of Economics and Statistics*, pages 126–134, 1996.

[51] Pascal Lavergne and Quang Vuong. Nonparametric significance testing. *Econometric Theory*, 16(4):576–601, 2000.

[52] Jungyoon Lee and Peter M Robinson. Panel nonparametric regression with fixed effects. *Journal of econometrics*, 188(2):346–362, 2015.

[53] Justin Lee. U-statistics: Theory and practice. 1990.

[54] Tae-Hwy Lee, Yundong Tu, and Aman Ullah. Forecasting equity premium: Global historical average versus local historical average and constraints. *Journal of Business & Economic Statistics*, 33(3):393–402, 2015.

[55] Yoonseok Lee, Debasri Mukherjee, and Aman Ullah. Nonparametric estimation of the marginal effect in fixed-effect panel data models. *Journal of Multivariate Analysis*, 171:53–67, 2019.

[56] Qi Li and Jeff Racine. Nonparametric estimation of distributions with categorical and continuous data. *journal of multivariate analysis*, 86(2):266–292, 2003.

[57] Qi Li and Jeffrey Scott Racine. *Nonparametric econometrics: theory and practice.* Princeton University Press, 2007.

[58] Qi Li and Suojin Wang. A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics*, 87(1):145–165, 1998.

[59] Xihong Lin and Raymond J Carroll. Nonparametric function estimation for clustered data when the predictor is measured without/with error. *Journal of the American statistical Association*, 95(450):520–534, 2000.

[60] Zhongjian Lin, Qi Li, and Yiguo Sun. A consistent nonparametric test of parametric regression functional form in fixed effects panel data models. *Journal of Econometrics*, 178:167–179, 2014.

[61] O Linton and JP Nielson. A kernel method of estimating nonparametric structured regression based on a marginal distribution. *Biometrika*, 82(93):100, 1995.

[62] Oliver Linton, Esfandiar Maasoumi, and Yoon-Jae Whang. Consistent testing for stochastic dominance under general sampling schemes. *The Review of Economic Studies*, 72(3):735–765, 2005.

[63] Regina Y Liu et al. Bootstrap procedures under some non-iid models. *The Annals of Statistics*, 16(4):1696–1708, 1988.

[64] Shujie Ma, Jeffrey S Racine, Aman Ullah, et al. *Nonparametric Regression-Spline Random Effects Models*. Department of Economics, McMaster Univ., 2015.

[65] James G MacKinnon and Matthew D Webb. Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, 32(2):233–254, 2017.

[66] Enno Mammen et al. Bootstrap and wild bootstrap for high dimensional linear models. *The annals of statistics*, 21(1):255–285, 1993.

[67] Carlos Martins-Filho and Feng Yao. Nonparametric frontier estimation via local linear regression. *Journal of Econometrics*, 141(1):283–319, 2007.

[68] Rosa L Matzkin. Identification in nonparametric simultaneous equations models. *Econometrica*, 76(5):945–978, 2008.

[69] Edward Miguel, Shanker Satyanath, and Ernest Sergenti. Economic shocks and civil conflict: An instrumental variables approach. *Journal of political Economy*, 112(4):725–753, 2004.

[70] Santosh Mishra, Liangjun Su, and Aman Ullah. Semiparametric estimator of time series conditional variance. *Journal of Business & Economic Statistics*, 28(2):256–274, 2010.

[71] Debasri Mukherjee. Feasible estimation of nonparametric generalized random effect model: An application. *Journal of Quantitative Economics*, 4(1):41–48, 2006.

[72] Irina Murtazashvili and Jeffrey M Wooldridge. A control function approach to estimating switching regression models with endogenous explanatory variables and endogenous switching. *Journal of econometrics*, 190(2):252–266, 2016.

[73] Whitney K Newey, James L Powell, and Francis Vella. Nonparametric estimation of triangular simultaneous equations models. *Econometrica*, 67(3):565–603, 1999.

[74] Adrian Pagan and Aman Ullah. *Nonparametric econometrics*. Cambridge university press, 1999.

[75] Christopher F Parmeter, Daniel J Henderson, and Subal C Kumbhakar. Nonparametric estimation of a hedonic price function. *Journal of Applied Econometrics*, 22(3):695–699, 2007.

[76] C Radakrishna Rao. Differential metrics in probability spaces based on entropy and divergence measures. Technical report, PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS, 1985.

[77] Murray Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pages 832–837, 1956.

[78] Anastasia Semykina. Self-employment among women: Do children matter more than we previously thought? *Journal of Applied Econometrics*, 33(3):416–434, 2018.

[79] Bernard W Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

[80] Liangjun Su and Xun Lu. Nonparametric dynamic panel data models: kernel estimation and specification testing. *Journal of Econometrics*, 176(2):112–133, 2013.

[81] Liangjun Su, Irina Murtazashvili, and Aman Ullah. Local linear gmm estimation of functional coefficient iv models with an application to estimating the rate of return to schooling. *Journal of Business & Economic Statistics*, 31(2):184–207, 2013.

[82] Liangjun Su and Aman Ullah. Profile likelihood estimation of partially linear panel data models with fixed effects. *Economics Letters*, 92(1):75–81, 2006.

[83] Liangjun Su and Aman Ullah. More efficient estimation of nonparametric panel data models with random effects. *Economics Letters*, 96(3):375–380, 2007.

[84] Liangjun Su and Aman Ullah. Local polynomial estimation of nonparametric simultaneous equations models. *Journal of Econometrics*, 144(1):193–218, 2008.

[85] Liangjun Su and Aman Ullah. A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. *Econometric Theory*, 29(1):187–212, 2013.

[86] Yun Sun and Qi Li. An alternative series based consistent model specification test. *Economics Letters*, 93(1):37–44, 2006.

[87] Aman Ullah. Specification analysis of econometric models. *Journal of quantitative economics*, 1:187–209, 1985.

[88] Aman Ullah. Non-parametric estimation of econometric functionals. *Canadian Journal of Economics*, pages 625–658, 1988.

[89] Aman Ullah and Victoria Zinde-Walsh. On the robustness of lm, lr, and w tests in regression models. *Econometrica: Journal of the Econometric Society*, pages 1055–1066, 1984.

[90] Naisyin Wang. Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika*, 90(1):43–52, 2003.

[91] Yoon-Jae Whang and Donald WK Andrews. Tests of specification for parametric and semiparametric models. *Journal of Econometrics*, 57(1-3):277–318, 1993.

[92] Chien-Fu Jeff Wu et al. Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics*, 14(4):1261–1295, 1986.

[93] De-Min Wu. Alternative tests of independence between stochastic regressors and disturbances. *Econometrica (pre-1986)*, 41(4):733, 1973.

[94] Feng Yao and Aman Ullah. A nonparametric r2 test for the presence of relevant variables. *Journal of Statistical Planning and Inference*, 143(9):1527–1547, 2013.

[95] Adonis John Yatchew. Nonparametric regression tests based on least squares. *Econometric Theory*, 8(4):435–451, 1992.

[96] Yonghui Zhang, Liangjun Su, and Peter CB Phillips. Testing for common trends in semi-parametric panel data models with fixed effects. *The Econometrics Journal*, 15(1):56–100, 2012.

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

Appendix A

Appendix for Chapter 2

This Appendix is for the derivation of the comparison between test statistics for 2.4.5. Note that it is done when $p = 1$, and the Gaussian kernel is used. Define $K(u)$ as the standard Gaussian kernel function and $\kappa(u)$ as the convolution kernel as follows.

$$K(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right), \quad \kappa(u) = \frac{1}{\sqrt{4\pi}} \exp\left(-\frac{u^2}{4}\right)$$

For the following derivations/calculations of the test statistics, please note that only the kernel components are extracted for simplicity.

A.1 Li-Wang Type Test

$$\int K^2(u)du = \int \frac{1}{2\pi}\exp\left(-u^2\right)du = \frac{\sqrt{\pi}}{2\pi}\underbrace{\int \frac{1}{\sqrt{\pi}}\exp\left(-u^2\right)du}_{\sim N(0,\frac{1}{2})}$$

$$= \frac{1}{2\sqrt{\pi}}$$

$$= 0.2821$$

A.2 Su-Ullah Type Test

For a local constant estimator,

$$\int \kappa^2(u)du = \int \frac{1}{4\pi}\exp\left(-\frac{u^2}{2}\right)du$$

$$= \frac{\sqrt{2\pi}}{4\pi}\underbrace{\int \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du}_{\sim N(0,1)}$$

$$= \frac{1}{2\sqrt{2\pi}}$$

$$= 0.1995$$

For a local linear estimator,

$$\int\left(\int \kappa(z)(1+z(x+z))dz\right)^2 dx$$

$$= \int\left(\int \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z^2}{2}\right)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{(z+x)^2}{2}\right)(1+z(x+z))dz\right)^2 dx$$

$$
\begin{aligned}
=&\int\left(\frac{1}{2\pi}\exp\left(-\frac{1}{4}x^2\right)\left[\sqrt{\pi}\int\frac{1}{\sqrt{\pi}}\exp\left(-\left(z+\frac{1}{2}x\right)^2\right)dz\right.\right.\\
&\left.\left.+\sqrt{\pi}\int\frac{1}{\sqrt{\pi}}z^2\exp\left(-\left(z+\frac{1}{2}x\right)^2\right)dz+\sqrt{\pi}x\int\frac{1}{\sqrt{\pi}}z\exp\left(-\left(z+\frac{1}{2}x\right)^2\right)dz\right]\right)^2dx\\
=&\int\left(\frac{1}{2\sqrt{\pi}}\exp\left(-\frac{1}{4}x^2\right)\left[\frac{3}{2}-\frac{1}{4}x^2\right]\right)^2dx\\
=&\int\frac{1}{4\pi}\exp\left(-\frac{x^2}{2}\right)\left[\frac{9}{4}-\frac{3}{4}x^2+\frac{1}{16}x^4\right]dx\\
=&\frac{9}{8\sqrt{2\pi}}-\frac{3}{8\sqrt{2\pi}}+\frac{3}{32\sqrt{2\pi}}\\
=&0.3366
\end{aligned}
$$

A.3 Yao-Ullah Type Test and F-Type Test

$$
\begin{aligned}
&\int(\kappa(u)-2K(u))^2du\\
=&\int(\kappa^2(u)-2\kappa(u)K(u)+4K^2(u))du\\
=&\int\frac{1}{4\pi}\exp\left(-\frac{u^2}{2}\right)du-4\int\frac{1}{2\sqrt{2}\pi}\exp\left(-\frac{3u^2}{4}\right)du+4\int\frac{1}{2\pi}\exp\left(-u^2\right)du\\
=&\frac{\sqrt{2\pi}}{4\pi}\underbrace{\int\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du}_{\sim N(0,1)}-\frac{\sqrt{2}\sqrt{4\pi/3}}{\pi}\underbrace{\int\frac{1}{\sqrt{4\pi/3}}\exp\left(-\frac{3u^2}{4}\right)du}_{\sim N(0,\frac{2}{3})}\\
&+\frac{4\sqrt{\pi}}{2\pi}\underbrace{\int\frac{1}{\sqrt{\pi}}\exp\left(-u^2\right)du}_{\sim N(0,\frac{1}{2})}\\
=&\frac{1}{2\sqrt{2\pi}}-\frac{2\sqrt{2}}{\sqrt{3\pi}}+\frac{2}{\sqrt{\pi}}\\
=&0.4065
\end{aligned}
$$

A.4 Proof of Theorem 1

Let $\hat{f}(x) = \hat{\varepsilon}' H_x \hat{\varepsilon}$ and $\hat{g}(x) = n^{-1}\hat{\varepsilon}' W_x \hat{\varepsilon}$. I claim

$$\text{plim} \int \frac{\hat{f}(x)}{\hat{g}(x)} dx = \text{plim} \frac{\int \hat{f}(x)dx}{\int \hat{g}(x)dx}, \text{ and } \lim_{n\to\infty} E\left[\frac{\hat{f}(x)}{\hat{g}(x)}\right] = \lim_{n\to\infty} \frac{E[\hat{f}(x)]}{E[\hat{g}(x)]}.$$

By Taylor expansion,

$$\begin{aligned}
E\left[\frac{\hat{f}(x)}{\hat{g}(x)}\right] &= E\left[\frac{\hat{f}(x)}{\hat{g}(x) - E[\hat{g}(x)] + E[\hat{g}(x)]}\right] \\
&= E\left[\frac{\hat{f}(x)}{E[\hat{g}(x)]} \frac{E[\hat{g}(x)]}{(\hat{g}(x) - E[\hat{g}(x)] + E[\hat{g}(x)])}\right] \\
&= E\left[\frac{\hat{f}(x)}{E[\hat{g}(x)]} \left(\frac{\hat{g}(x) - E[\hat{g}(x)] + E[\hat{g}(x)]}{E[\hat{g}(x)]}\right)^{-1}\right] \\
&= E\left[\frac{\hat{f}(x)}{E[\hat{g}(x)]} \left(1 + \frac{\hat{g}(x) - E[\hat{g}(x)]}{E[\hat{g}(x)]}\right)^{-1}\right] \\
&= E\left[\frac{\hat{f}(x)}{E[\hat{g}(x)]} \left(1 - \frac{\hat{g}(x) - E[\hat{g}(x)]}{E[\hat{g}(x)]} + \left(\frac{\hat{g}(x) - E[\hat{g}(x)]}{E[\hat{g}(x)]}\right)^2 - \cdots\right)\right] \\
&\simeq \frac{E[\hat{f}(x)]}{E[\hat{g}(x)]} - \frac{E\left[\hat{f}(x)(\hat{g}(x) - E[\hat{g}(x)])\right]}{(E[\hat{g}(x)])^2} + \frac{E\left[\hat{f}(x)(\hat{g}(x) - E[\hat{g}(x)])^2\right]}{(E[\hat{g}(x)])^3}
\end{aligned}$$

By Cauchy-Schwarz Inequality,

$$\begin{aligned}
E\left[\hat{f}(x)(\hat{g}(x) - E[\hat{g}(x)])\right] &\leq \sqrt{E[\hat{f}^2(x)]E[(\hat{g}(x) - E[\hat{g}(x)])^2]} = \sqrt{E[\hat{f}^2(x)]\mathrm{V}(\hat{g}(x))} \\
E\left[\hat{f}(x)(\hat{g}(x) - E[\hat{g}(x)])^2\right] &\leq \sqrt{E[\hat{f}^2(x)]E[(\hat{g}(x) - E[\hat{g}(x)])^4]} = \sqrt{E[\hat{f}^2(x)]}\ \mathrm{V}(\hat{g}(x))
\end{aligned}$$

$$\therefore \quad \left|E\left[\frac{\hat{f}(x)}{\hat{g}(x)}\right] - \frac{E[\hat{f}(x)]}{E[\hat{g}(x)]}\right| \leq \frac{1}{(E[\hat{g}(x)])^2}\left|-\sqrt{E[\hat{f}^2(x)]\mathrm{V}(\hat{g}(x))} + \sqrt{E[\hat{f}^2(x)]}\ \mathrm{V}(\hat{g}(x))\right|$$

It will be sufficient to show that the above claim holds by showing

$$\begin{aligned}
\hat{f}(x) &= \hat{\varepsilon}' H_x \hat{\varepsilon} = (\hat{\varepsilon} - \varepsilon + \varepsilon)' H_x (\hat{\varepsilon} - \varepsilon + \varepsilon) \\
&= (\varepsilon - (g(X,\hat{\theta}) - g(X,\theta)))' H_x (\varepsilon - (g(X,\hat{\theta}) - g(X,\theta))) \\
&= \varepsilon' H_x \varepsilon - (g(X,\hat{\theta}) - g(X,\theta))' H_x \varepsilon - \varepsilon' H_x (g(X,\hat{\theta}) - g(X,\theta)) \\
&+ (g(X,\hat{\theta}) - g(X,\theta))' H_x (g(X,\hat{\theta}) - g(X,\theta)) \\
&= \varepsilon' H_x \varepsilon + O((nh^{p/2})^{-1}) \\
&= n^{-1} \varepsilon' \bar{H}_x \varepsilon + \varepsilon' (H_x - n^{-1} \bar{H}_x) \varepsilon + o_p(1), \text{ where } \bar{H}_x = Z_x W_x (E[Z_x' W_x Z_x])^{-1} W_x Z_x \\
&= n^{-1} \varepsilon' \bar{H}_x \varepsilon + o_p(1)
\end{aligned}$$

$$\begin{aligned}
\hat{g}(x) &= n^{-1} \hat{\varepsilon}' W_x \hat{\varepsilon} \\
&= n^{-1} (\hat{\varepsilon} - \varepsilon + \varepsilon)' W_x (\hat{\varepsilon} - \varepsilon + \varepsilon) \\
&= n^{-1} \varepsilon' W_x \varepsilon - n^{-1} (g(X,\hat{\theta}) - g(X,\theta))' W_x \varepsilon - n^{-1} \varepsilon' W_x (g(X,\hat{\theta}) - g(X,\theta)) \\
&+ n^{-1} (g(X,\hat{\theta}) - g(X,\theta))' W_x (g(X,\hat{\theta}) - g(X,\theta)) \\
&= \varepsilon' W_x \varepsilon + O(n^{-1})
\end{aligned}$$

$$\therefore \hat{f}(x) - E[\hat{f}(x)] = o_p(1), \text{ and } g(x) - E[g(x)] = o_p(1)$$

$$\begin{aligned}
\lim_{n\to\infty} \left| E\left[\frac{\hat{f}(x)}{\hat{g}(x)}\right] - \frac{E[\hat{f}(x)]}{E[\hat{g}(x)]} \right| &\leq \lim_{n\to\infty} \frac{1}{(E[\hat{g}(x)])^2} \left| -\sqrt{E[\hat{f}^2(x)] \mathrm{V}(\hat{g}(x))} + \sqrt{E[\hat{f}^2(x)]}\ \mathrm{V}(\hat{g}(x)) \right| \\
&= 0 \qquad \blacksquare
\end{aligned}$$

Appendix B

Appendix for Chapter 3

B.1 Proof of Theorem 3

I let

$$\widehat{\Delta}_{ij} = \widehat{v}_j - \widehat{v}_i \quad \text{and} \quad \Delta_{ij} = v_j - v_i.$$

I also let Δ_{ij}^* be a vector between $\widehat{\Delta}_{ij}$ and Δ_{ij}. Note that

$$\widehat{\Delta}_{ij} - \Delta_{ij} = \left(\widehat{g}\left(z_i\right) - g\left(z_i\right)\right) - \left(\widehat{g}\left(z_j\right) - g\left(z_j\right)\right),$$

$$\widehat{u}_i = y_i - \widehat{m}\left(x_i\right) = u_i - \left(\widehat{m}\left(x_i\right) - m\left(x_i\right)\right).$$

Assuming that $K(\cdot)$ is twice continuously differentiable, I expand:

$$\begin{aligned} K\left(H_v^{-1}\widehat{\Delta}_{ij}\right) &= K\left(H_v^{-1}\Delta_{ij}\right) + K^{(1)}\left(H_v^{-1}\Delta_{ij}\right)' H_v^{-1}\left(\widehat{\Delta}_{ij} - \Delta_{ij}\right) \\ &\quad + \frac{1}{2}\left(\widehat{\Delta}_{ij} - \Delta_{ij}\right)' H_v^{-1} K^{(2)}\left(H_v^{-1}\Delta_{ij}^*\right) H_v^{-1}\left(\widehat{\Delta}_{ij} - \Delta_{ij}\right), \end{aligned}$$

The test statistic I_n can be written as follows:

$$\begin{aligned}
I_n &= \frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\widehat{u}_i\widehat{u}_jK\left(H_v^{-1}\widehat{\Delta}_{ij}\right) \\
&= \frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\widehat{u}_i\widehat{u}_jK\left(H_v^{-1}\Delta_{ij}\right) \\
&+ \frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\widehat{u}_i\widehat{u}_jK^{(1)}\left(H_v^{-1}\Delta_{ij}\right)'H_v^{-1}\left(\widehat{\Delta}_{ij}-\Delta_{ij}\right) \\
&+ \frac{1}{2n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\widehat{u}_i\widehat{u}_j\left(\widehat{\Delta}_{ij}-\Delta_{ij}\right)'H_v^{-1}K^{(2)}\left(H_v^{-1}\Delta_{ij}^{*}\right)H_v^{-1}\left(\widehat{\Delta}_{ij}-\Delta_{ij}\right) \\
&\equiv I_{1n}+I_{2n}+I_{3n}.
\end{aligned}$$

To derive the asymptotic distribution of I_n, I will show

(I) $\sqrt{n^2\left|H\right|}I_{1n}\xrightarrow{d}N(0,\Omega)$

(II) $I_{2n}=O\left(\left(\frac{\ln n}{n^{3/2}\sqrt{|H_v|^{1/2}|H_z|}}\right)^{\frac{1}{2}}\right)+O\left(\frac{|H_z|^2}{n|H_v|^{1/2}}\right)=o_p((\sqrt{n^2\left|H_v\right|})^{-1})$ by (A5)

(III) $I_{3n}=O\left(\frac{\ln n}{n^2\sqrt{|H_v|^{1/2}|H_z|}}\right)+O\left(\frac{|H_z|^4}{\sqrt{n^2|H_v|}}\right)+O\left(\frac{(\ln n)^{1/2}|H_z|^{3/2}}{n^{3/2}|H_v|^{1/2}}\right)=o_p((\sqrt{n^2\left|H_v\right|})^{-1})$

(IV) $\hat{\Omega}=\Omega+o_p(1)$

(I) Let $\mu_n\left(x\right)=\widehat{m}\left(x\right)-m\left(x\right)$. I first decompose

$$\begin{aligned}
I_{1n} &= \frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(u_i-\mu_n\left(x_i\right)\right)\left(u_j-\mu_n\left(x_j\right)\right)K\left(H^{-1}\Delta_{ij}\right) \\
&= \frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}u_iu_jK\left(H_v^{-1}\Delta_{ij}\right) \\
&+ \frac{2}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}u_i\mu_n\left(x_j\right)K\left(H_v^{-1}\Delta_{ij}\right)
\end{aligned}$$

$$+\frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\mu_n\left(x_i\right)\mu_n\left(x_j\right)K\left(H_v^{-1}\Delta_{ij}\right)$$

$$\equiv I_{11n}+I_{12n}+I_{13n}.$$

(I)-(A) Following Lemma 1 from Yao and Ullah (2013), define second-order U-statistic $U_n=\frac{1}{n(n-1)}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\phi_n(X_i,X_j)$, where $\phi_n(X_i,X_j)$ is symmetric function of X_j and X_i, where $\{X_i\}_{i=1}^{n}$ is a sequence of IID random variables. $E[\phi_n(X_i,X_j)\mid X_j]=0$. I can easily verify that I_{11n} is a degenerated second order U-statistic.

$$\begin{aligned}
I_{11n}&=\frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}u_iu_jK\left(H_v^{-1}\Delta_{ij}\right)\\
&=\frac{1}{n(n-1)\left|H_v\right|}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}[u_iu_jK\left(H_v^{-1}\Delta_{ij}\right)+u_ju_iK\left(H_v^{-1}\Delta_{ij}\right)]\\
&=\frac{1}{n(n-1)\left|H_v\right|}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}[\psi_n(W_i,W_j)+\psi_n(W_j,W_i)],\ \text{where } W_i=(u_i,v_i)\\
&=\frac{1}{n(n-1)\left|H_v\right|}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\phi_n(W_i,W_j)
\end{aligned}$$

As $\mathrm{E}[\phi_n^2(W_i,W_j)]=\mathrm{E}[\psi_n^2(W_i,W_j)]+\mathrm{E}[\psi_n^2(W_j,W_i)]+2\mathrm{E}[\psi_n(W_i,W_j)\psi_n(W_j,W_i)]$,

$$\begin{aligned}
\frac{1}{|H_v|}E[\psi_n^2(W_i,W_j)]&=\frac{1}{|H_v|}E[\psi_n^2(W_j,W_i)]\\
&=\frac{1}{|H_v|}E[K^2\left(H_v^{-1}\Delta_{ij}\right)u_i^2u_j^2]\\
&=\frac{1}{|H_v|}E[K^2\left(H_v^{-1}\Delta_{ij}\right)\sigma^2(v_i)\sigma^2(v_j)]\\
&=\frac{1}{|H_v|}\int K^2\left(H_v^{-1}\Delta_{ij}\right)\sigma^2(v_i)\sigma^2(v_j)f(v_i)f(v_j)dv_idv_j
\end{aligned}$$

$$
\begin{aligned}
&= \int K^2(\psi)\sigma^2(v_i + H_v\psi)\sigma^2(v_i)f(v_i)f(v_i + H_v\psi)dv_i d\psi \\
&\to \int K^2(\psi)d\psi E[\sigma^4(v_i)f(v_i)] < \infty
\end{aligned}
$$

$$
\begin{aligned}
\frac{2}{|H_v|}E[\psi_n(W_i, W_j)\psi_n(W_j, W_i)] &= \frac{2}{|H_v|}E[K^2\left(H_v^{-1}\Delta_{ij}\right)\sigma^4(v_i)f(v_i)] \\
&\to 2\int K^2(\psi)d\psi E[\sigma^4(v_i)f(v_i)] < \infty
\end{aligned}
$$

Then, I have $\left(E[\phi_n^2(W_i, W_j)]\right)^2 = O(|H_v|^2)$. By Cr inequality,

$$
E[\phi_n^4(W_i, W_j)] \leq C\left[E[\psi_n^4(W_i, W_j)] + E[\psi_n^4(W_j, W_i)]\right].
$$

$$
\begin{aligned}
\frac{1}{|H_v|}E[\psi_n^4(W_i, W_j)] &= \frac{1}{|H_v|}E\left[E\left[K^4\left(H_v^{-1}\Delta_{ij}\right)u_i^4 u_j^4 \mid v_i, v_j\right]\right] \\
&= \frac{1}{|H_v|}E\left[\sigma^4(v_i)\sigma^4(v_j)K^4\left(H_v^{-1}\Delta_{ij}\right)\right] \\
&= \int K^4(\psi)\sigma^4(v_i)\sigma^4(v_i + H_v\psi)f(v_i)f(v_i + H_v\psi)dv_i d\psi \\
&= \int K^4(\psi)\sigma^8(v_i)f^2(v_i)dv_i d\psi \\
&= \left(\int K^4(\psi)d\psi\right)\left(\int \sigma^8(v_i)f^2(v_i)dv_i\right)
\end{aligned}
$$

Here, I have $\frac{1}{n}\mathrm{E}[\phi_n^4(W_i, W_j)] = O(n^{-1}|H_v|)$.

$$
\begin{aligned}
G_n(W_i,W_j) =& E\left[\phi_n(W_t,W_i)\phi_n(W_t,W_j) \mid W_i,W_j\right] \\
=& E\left[(\psi_n(W_t,W_i)+\psi_n(W_i,W_t))(\psi_n(W_t,W_j)+\psi_n(W_j,W_i)) \mid W_i,W_j\right] \\
=& E[\psi_n(W_t,W_i)\psi_n(W_t,W_j) \mid W_i,W_j] + E[\psi_n(W_t,W_i)\psi_n(W_j,W_t) \mid W_i,W_j] \\
&+ E[\psi_n(W_i,W_t)\psi_n(W_t,W_j) \mid W_i,W_j] + E[\psi_n(W_i,W_t)\psi_n(W_j,W_t) \mid W_i,W_j] \\
=& G_1(W_i,W_j)+G_2(W_i,W_j)+G_3(W_i,W_j)+G_4(W_i,W_j)
\end{aligned}
$$

By Cr inequality,

$$
\begin{aligned}
E\left[G_n^2(W_i,W_j)\right] &= E[(G_1(W_i,W_j)+G_2(W_i,W_j)+G_3(W_i,W_j)+G_4(W_i,W_j))^2] \\
&\leq C\left[G_1^2(W_i,W_j)+G_2^2(W_i,W_j)+G_3^2(W_i,W_j)+G_4^2(W_i,W_j)\right]
\end{aligned}
$$

$$
\begin{aligned}
& E[G_1^2(W_i,W_j)] \\
&= E\left[E\left[\psi_n(W_t,W_i)\psi_n(W_t,W_j) \mid W_i,W_j\right]^2\right] \\
&= E\left[E\left[K\left(H_v^{-1}\Delta_{it}\right)K\left(H_v^{-1}\Delta_{jt}\right)u_iu_ju_t^2 \mid v_i,v_j\right]^2\right] \\
&= E\left[\left(u_iu_jE\left[K\left(H_v^{-1}\Delta_{it}\right)K\left(H_v^{-1}\Delta_{jt}\right)\sigma^2(v_t) \mid v_i,v_j\right]\right)^2\right] \\
&= E\left[\left(u_iu_j\int K\left(H_v^{-1}\Delta_{it}\right)K\left(H_v^{-1}\Delta_{jt}\right)\sigma^2(v_t)f(v_t)dv_t\right)^2\right] \\
&= E\left[E\left[u_i^2u_j^2\left(\int K\left(\psi_1\right)K\left(\psi_1+H_v^{-1}\Delta_{ij}\right)\sigma^2(v_i+H_v\psi_1)f(v_i+H_v\psi_1)\left|H_v\right|d\psi_1\right)^2 \mid v_i,v_j\right]\right] \\
&= E\left[\sigma^2(v_i)\sigma^2(v_j)\left(\int K\left(\psi_1\right)K\left(\psi_1+H_v^{-1}\Delta_{ij}\right)\sigma^2(v_i+H_v\psi_1)f(v_i+H_v\psi_1)\left|H_v\right|d\psi_1\right)^2\right]
\end{aligned}
$$

$$= |H_v|^2 \int \sigma^2(v_i)\sigma^2(v_j) \left(\int K(\psi_1) K\left(\psi_1 + H_v^{-1}\Delta_{ij}\right) \sigma^2(v_i + H_v\psi_1) f(v_i + H_v\psi_1) d\psi_1 \right)^2$$

$$\times f(v_i) f(v_j) dv_i dv_j$$

$$= |H_v|^2 \int \sigma^2(v_i)\sigma^2(v_i - H_v\psi_2) \left(\int K(\psi_1) K(\psi_1 + \psi_2) \sigma^2(v_i + H_v\psi_1) f(v_i + H_v\psi_1) d\psi_1 \right)^2$$

$$\times f(v_i) f(v_i - H_v\psi_2) |H_v| dv_i d\psi_2$$

$$= |H_v|^3 \int \sigma^4(v_i) \left(\int K(\psi_1) K(\psi_1 + \psi_2) \sigma^2(v_i) f(v_i) d\psi_1 \right)^2 f^2(v_i) dv_i d\psi_2$$

$$= |H_v|^3 \left(\int \sigma^8(v_i) f^4(v_i) dv_i \right) \int \left(\int K(\psi_1) K(\psi_1 + \psi_2) d\psi_1 \right)^2 d\psi_2$$

$$= O(|H_v|^3)$$

I can conclude that

$$\begin{aligned} \frac{E[G_n^2(W_i,W_j)] + n^{-1}E[\phi_n^4(W_i,W_j)]}{(E[\phi_n^2(W_i,W_j)])^2} &= \frac{O(|H_v|^2) + n^{-1}O(|H_v|)}{O(|H_v|^2)} \\ &= O(|H_v|) + O((n|H_v|)^{-1}) \\ &\to 0 \text{ as } n \to \infty \end{aligned}$$

Also, I have $\frac{1}{|H_v|}\mathrm{E}[\phi_n^2(W_i, W_j)] \to \Omega$, where $\Omega \equiv 2\int K^2(\psi)d\psi\mathrm{E}[\sigma^4(v_i)f(v_i)]$. By applying Hall's Central Limit Theorem,

$$\sqrt{n^2 |H_v|} I_{11n} \xrightarrow{d} N(0, \Omega)$$

(I)-(B) As $\hat{m}(x_i)$ is a local linear estimator and $z_t' = (1, \frac{x_t - x_i}{h_x})$,

$$\begin{aligned}
&\hat{m}(x_i) - m(x_i) \\
=& e_1' \frac{1}{2} \left(\sum_{t=1}^{n} z_t' K\left(H_x^{-1}(x_t - x_i)\right) z_t \right)^{-1} \sum_{t=1}^{n} z_t' K\left(H_x^{-1}(x_t - x_i)\right) (x_i - x_t) m^{(2)}(x_{it})(x_i - x_t)' \\
+& e_1' \left(\sum_{t=1}^{n} z_t' K\left(H_x^{-1}(x_t - x_i)\right) z_t \right)^{-1} \sum_{t=1}^{n} z_t' K\left(H_x^{-1}(x_t - x_i)\right) u_t \\
=& e_1' \left(\sum_{t=1}^{n} z_t' K\left(H_x^{-1}(x_t - x_i)\right) z_t \right)^{-1} \sum_{t=1}^{n} z_t' K\left(H_x^{-1}(x_t - x_i)\right) \left(\frac{1}{2} m^*(x_{it}) + u_t \right),
\end{aligned}$$

where $m^*(x_{it}) = (x_i - x_t) m^{(2)}(x_{it})(x_i - x_t)'$ and $x_{it} = \lambda x_i + (1 - \lambda) x_t$. Define

$$\hat{\mu}_n(x_i) = \hat{m}(x_i) - m(x_i) = \mu_n(x_i) + \hat{\mu}_n(x_i) - \mu_n(x_i),$$

where $\mu_n(x_i) = \frac{1}{n|H_x|f(x_i)} \sum_{t=1}^{n} K\left(H_x^{-1}(x_t - x_i)\right) (\frac{1}{2} m^*(x_{it}) + u_t)$.

$$\begin{aligned}
I_{12n} &= -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j (\hat{m}(x_i) - m(x_i)) K\left(H_v^{-1}(v_j - v_i)\right) \\
&= -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j \Bigg[e_1' \left(\sum_{t=1}^{n} z_t K\left(H_x^{-1}(x_t - x_i)\right) z_t \right)^{-1} \\
&\times \sum_{t=1}^{n} z_t K\left(H_x^{-1}(x_t - x_i)\right) \left(\frac{1}{2} m^*(x_{it}) + u_t \right) \Bigg] K\left(H_v^{-1}(v_j - v_i)\right) \\
&= -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j \ \hat{\mu}_n(x_i) K\left(H_v^{-1}(v_j - v_i)\right) \\
&= -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j \left(\mu_n(x_i) + \hat{\mu}_n(x_i) - \mu_n(x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right) \\
&= -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j \left(\frac{1}{n|H_x|f(x_i)} \sum_{t=1}^{n} K\left(H_x^{-1}(x_t - x_i)\right) \frac{1}{2} m^*(x_{it}) \right)
\end{aligned}$$

$$\times K\left(H_v^{-1}(v_j - v_i)\right) - \frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_j \left(\frac{1}{n|H_x|f(x_i)}\sum_{t=1}^{n} K\left(H_x^{-1}(x_t - x_i)\right) u_t\right)$$

$$\times K\left(H_v^{-1}(v_j - v_i)\right) + \frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_j \left(\hat{\mu}_n(x_i) - \mu_n(x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right)$$

$$\equiv S_{1n} + S_{2n} + S_{3n}$$

(I)-(B)-(i)

$$S_{1n} = -\frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_j \left(\frac{1}{n\,|H_x|\,f(x_i)}\sum_{t=1}^{n} K\left(H_x^{-1}(x_t - x_i)\right)\frac{1}{2}m^*(x_{it})\right)$$

$$\times K\left(H_v^{-1}(v_j - v_i)\right)$$

$$= -\frac{2}{n^2(n-1)\,|H_v|\,|H_x|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{n}\frac{1}{2f(x_i)}u_j m^*(x_{it}) K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right)$$

$$= -\frac{2}{n^2(n-1)\,|H_v|\,|H_x|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{n}\frac{1}{2f(x_i)}u_j m^*(x_{it}) K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right)$$

1) $i = t$

$$S_{1n} = -\frac{2}{n(n-1)\,|H_v|\,|H_x|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{2f(x_i)}u_j m^*(x_i) K(0) K\left(H_v^{-1}(v_j - v_i)\right) = 0$$

$$\because m^*(x_i) = (x_i - x_i) m^{(2)}(x_i)(x_i - x_i) = 0$$

2) $j = t,\ j \neq i,\ t \neq i$

$$S_{1n} = -\frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{2f(x_i)}u_j m^*(x_{ij}) K\left(H_x^{-1}(x_t - x_i), H_v^{-1}(v_j - v_i)\right)$$

Given $E[u \mid x, v] = 0$,

$$S_{1n} = -\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left[\frac{1}{2f(x_i)\,|H_v|} u_j m^*(x_{ij}) K\left(H_x^{-1}(x_t - x_i), H_v^{-1}(v_j - v_i)\right) \right.$$
$$\left. + \frac{1}{2f(x_j)\,|H_v|} u_i m^*(x_{ji}) K\left(H_x^{-1}(x_t - x_i), H_v^{-1}(v_j - v_i)\right) \right]$$

By letting $W_i = (u_i, x_i, v_i)$,

$$\begin{aligned} S_{1n} &= -\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} [\psi(W_i, W_j) + \psi(W_j, W_i)] \\ &= -\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \phi_n(W_i, W_j) \\ &= -\frac{2}{n(n-1)} \sum_{\substack{i=1 \\ i<j}}^{n} \sum_{j=1}^{n} \phi_n(W_i, W_j) \end{aligned}$$

By applying Lemma 1 of Yao and Ullah (2013),

$$S_{1n} - \left[\frac{2}{n(n-1)} \sum_{i=1}^{n} \int \phi_n(W_i, W_j) dP(W_j) - n^{-1} E[\phi_n(W_i, W_j)] \right]$$
$$= O_p\left(n^{-1} (E[\phi_n^2(W_i, W_j)]^{\frac{1}{2}} \right)$$

Note that $\mathrm{E}[\phi_n(W_i, W_j)] = 0$ ($\because$ $\mathrm{E}[u_i \mid x_i, v_i] = 0$). By Lipschitz-condition,

$$E[\phi_n^2(W_i, W_j)] \leq C\left[E[\psi^2(W_i, W_j)] + E[\psi^2(W_j, W_i)] \right].$$

$$E[\psi^2(W_i,W_j)]$$

$$= E\left[\frac{1}{4f(x_i)^2\left|H_v\right|^2}u_j^2(m^*(x_{ij}))^2K^2\left(H_x^{-1}(x_t-x_i),H_v^{-1}(v_j-v_i)\right)\right]$$

$$= E\left[\frac{\left|H_x\right|^4}{4f(x_i)^2\left|H_v\right|^2}u_j^2\left(H_x^{-1}(x_j-x_i)\right)^2(m^{(2)}(x_{ij}))^2\left(H_x^{-1}(x_j-x_i)\right)^2\right.$$

$$\left.K^2\left(H_x^{-1}(x_t-x_i),H_v^{-1}(v_j-v_i)\right)\right]$$

$$= \int\frac{\left|H_x\right|^4}{4f(x_i)^2\left|H_v\right|^2}\sigma^2(w_j)\left(H_x^{-1}(x_j-x_i)\right)^2(m^{(2)}(x_{ij}))^2\left(H_x^{-1}(x_j-x_i)\right)^2$$

$$\times K^2\left(H_x^{-1}(x_t-x_i),H_v^{-1}(v_j-v_i)\right)f(w_i)f(w_j)dw_idw_j$$

Let $\psi_x = H_x^{-1}(x_j-x_i), \psi_v = H_v^{-1}(v_j-v_i),$ and $\psi = (\psi_x,\psi_v)$.

$$= \int\frac{\left|H_x\right|^4}{4f(x_j+H_x\psi_x)^2\left|H\right|^2}\sigma^2(w_j)\psi_x^2(m^{(2)}(x_j+(1-\lambda)H_x\psi_x))^2\psi_x^2K^2\left(\psi_x,\psi_v\right)$$

$$\times f(\psi)f(w_j+H_xH_v\psi)\left|H\right|d\psi dw_j$$

$$= \int\frac{\left|H_x\right|^4}{4f(x_j)^2\left|H_v\right|}\sigma^2(w_j)\psi_x^2(m^{(2)}(x_j))^2\psi_x^2K^2\left(\psi_x,\psi_v\right)f(w_j)^2d\psi dw_j+o(1)$$

$$= \frac{\left|H_x\right|^4}{\left|H_v\right|}\left(\int K^2(\psi)\psi_x^4d\psi\right)\left(\int\frac{1}{f(x_j)^2}\sigma^2(w_j)(m^{(2)}(x_j))^2f(w_j)^2dw_j\right)$$

Then, I have

$$n^{-1}\left(E[\phi_n^2(W_i,W_j)]\right)^{\frac{1}{2}} = O_p(n^{-1}\left|H_v\right|^{-1/2}\left|H_x\right|^2).$$

3) $j\neq t,\ t\neq i$

$$S_{1n} = -\frac{2}{n^2(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j\neq t}}^{n}\sum_{t=1}^{n}\frac{1}{2f(x_i)\left|H_v\right|}u_jm^*(x_{it})K\left(H_x^{-1}(x_t-x_i)\right)K\left(H_v^{-1}(v_j-v_i)\right)$$

By letting $W_i = (u_i, x_i, v_i)$,

$$= -\frac{2}{3n^2(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j \neq t}}^{n} \sum_{t=1}^{n} \left[\psi_n(W_i, W_j, W_t) + \psi_n(W_j, W_t, W_i) + \psi_n(W_t, W_i, W_j) \right]$$

$$= -\frac{2}{n^2(n-1)} \sum_{i<j<t}^{n} \sum^{n} \sum^{n} 2\phi_n(W_i, W_j, W_t)$$

$$= \left(-\frac{2}{n^2(n-1)} - \binom{n}{3}^{-1} + \binom{n}{3}^{-1} \right) \sum_{i<j<t}^{n} \sum^{n} \sum^{n} 2\phi_n(W_i, W_j, W_t)$$

$$E[\phi_n^2(W_i, W_j, W_t)] \leq C \left[E[\psi^2(W_i, W_j, W_t)] + E[\psi^2(W_j, W_t, W_i)] + E[\psi^2(W_t, W_i, W_j)] \right]$$

By H-decomposition,

$$E[\phi_n(W_i, W_j, W_t) \mid W_j, W_t]$$

$$= u_j E \left[\frac{1}{2f(x_i)\,|H_v|} m^*(x_{it}) K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right) \mid W_j, W_t \right]$$

$$= \phi_{2n}(W_j, W_t)$$

$$E[\phi_{2n}^2(W_j, W_t)]$$

$$= E\left[\left(u_j E\left[\frac{1}{2f(x_i)\,|H_v|} m^*(x_{it}) K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right) \mid W_j, W_t \right] \right)^2 \right]$$

$$= E\left[u_j \left(E\left[\frac{|H_x|^4}{2f(x_i)\,|H|} \left(H_x^{-1}(x_t - x_i)\right)^2 (m^{(2)}(x_{it}))^2 \left(H_x^{-1}(x_t - x_i)\right)^2 \right.\right.\right.$$

$$\left.\left.\left. \times\ K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right) \mid W_j, W_t \right] \right)^2 \right]$$

$$= E\left[\sigma^2(v_j) \left(\int \frac{|H_x|^4}{2f(x_i)\,|H|} \left(H_x^{-1}(x_t - x_i)\right)^2 (m^{(2)}(x_{it}))^2 \left(H_x^{-1}(x_t - x_i)\right)^2 \right.\right.$$

$$\left.\left. \times K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right) f(w_i) dw_i \right)^2 \right]$$

$$= E\left[\sigma^2(v_j)\left(\int \frac{|H_x|^4}{2f(x_t - H_x\psi_{x,1})\,|H|}\psi_{x,1}^2(m^{(2)}(x_t - \lambda H_x\psi_{x,1}))^2\psi_{x,1}^2 K\left(\psi_{x,1}\right)\right.\right.$$

$$\left.\left.\times\, K\left(\psi_{v,1} + H_v^{-1}(v_j - v_t)\right) f(w_t - HH_x\psi_1)\,|H|\,d\psi_1\right)^2\right]$$

$$= \int \sigma^2(v_j)\left(\int \frac{|H_x|^4}{2f(x_t - h_x\psi_{x,1})}\psi_{x,1}^2(m^{(2)}(x_t - \lambda H_x\psi_{x,1}))^2\psi_{x,1}^2 K\left(\psi_{x,1}\right)\right.$$

$$\left.\times K\left(\psi_{v,1} + H_v^{-1}(v_j - v_t)\right) f(w_t - HH_x\psi_1)d\psi_1\right)^2 f(w_j)f(w_t)dw_j dw_t$$

$$= \int \sigma^2(v_t + H_v\psi_{v,2})\left(\int \frac{|H_x|^4}{2f(x_t - H_x\psi_{x,1})}\psi_{x,1}^2(m^{(2)}(x_t - \lambda H_x\psi_{x,1}))^2\psi_{x,1}^2 K\left(\psi_{x,1}\right)\right.$$

$$\left.\times\, K\left(\psi_{v,1} + \psi_{v,2}\right) f(w_t - H\psi_1)d\psi_1\right)^2 f(w_j)f(w_t)dw_j dw_t$$

$$= \frac{|H_x|^8\,|H|}{4}\int \sigma^2(v_t)\frac{f^4(w_t)}{f^2(x_t)}(m^{(2)}(x_t))^2 dw_t \int\left(\int \psi_{x,1}^4 K\left(\psi_{x,1}\right) K\left(\psi_{v,1} + \psi_{v,2}\right) d\psi_1\right)^2 d\psi_2$$

$$= O(|H_x|^8\,|H|)$$

$$\therefore S_{1n} = O(n^{-2}\,|H|^{1/2}\,|H_x|^4)$$

(I)-(B)-(ii)

$$S_{2n} = -\frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_j\left(\frac{1}{n\,|H_x|\,f(x_i)}\sum_{t=1}^{n} K\left(H_x^{-1}(x_t - x_i)\right)u_t\right) K\left(H_v^{-1}(v_j - v_i)\right)$$

$$= -\frac{2}{n(n-1)\,|H_v|\,|H_x|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{n}\frac{1}{f(x_i)}u_j u_t K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right)$$

$$= -\frac{2}{n(n-1)\,|H_v|\,|H_x|}\sum_{i=1}^{n}\frac{1}{f(x_i)}\left[\sum_{j\neq i}^{n}\sum_{t=1}^{n} u_j u_t K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right)\right.$$

$$\left. + \sum_{j\neq i}^{n}\sum_{t=1}^{n} u_t u_j K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right)\right]$$

1) $i = t$

$$S_{2n} = -\frac{2}{n(n-1)\left|H_v\right|\left|H_x\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{f(x_i)}u_j u_i K\left(H_x^{-1}(x_j - x_i), H_v^{-1}(v_j - v_i)\right)$$

$$= -\frac{2}{n(n-1)\left|H_v\right|\left|H_x\right|}\sum_{i<j}^{n}\sum^{n}\left[\frac{1}{f(x_i)}u_j u_i K\left(H_x^{-1}(x_j - x_i), H_v^{-1}(v_j - v_i)\right)\right.$$

$$\left.+\frac{1}{f(x_j)}u_i u_j K\left(H_x^{-1}(x_j - x_i), H_v^{-1}(v_j - v_i)\right)\right]$$

By letting $W_i = (u_i, x_i, v_i)$,

$$= -\frac{2}{n(n-1)}\sum_{i<j}^{n}\sum^{n}\left[\psi_n(W_i, W_j) + \psi_n(W_j, W_i)\right]$$

$$= -\frac{2}{n(n-1)}\sum_{i<j}^{n}\sum^{n}\phi_n(W_i, W_j)$$

As $\mathrm{E}[\phi_n^2(W_i, W_j)] = \mathrm{E}[\psi_n^2(W_i, W_j)] + \mathrm{E}[\psi_n^2(W_j, W_i)] + 2\mathrm{E}[\psi_n(W_i, W_j)\psi_n(W_j, W_i)]$,

$$\frac{1}{\left|H_v\right|\left|H_x\right|}E[\phi_n^2(W_i, W_j)]$$

$$= \frac{1}{\left|H_v\right|\left|H_x\right|}E[\phi_n^2(Z_j, Z_i)]$$

$$= \frac{1}{\left|H_v\right|\left|H_x\right|}E\left[\frac{1}{f(x_i)^2}u_j^2 u_i^2 K^2\left(H_x^{-1}(x_j - x_i), H_v^{-1}(v_j - v_i)\right)\right]$$

$$= \frac{1}{\left|H_v\right|\left|H_x\right|}\int\frac{1}{f(x_i)^2}\sigma^2(w_i)\sigma^2(w_j)K^2\left(H_x^{-1}(x_j - x_i), H_v^{-1}(v_j - v_i)\right)$$

$$\times f(w_i)f(w_j)dw_i dw_j$$

$$= \frac{1}{\left|H_v\right|\left|H_x\right|}\int\frac{1}{f(x_i)^2}\sigma^2(w_i)\sigma^2(w_i + H\psi)K^2(\psi)f(w_i)f(w_i + H\psi)\left|H\right|\left|H_x\right|dw_i d\psi$$

$$= \int\frac{1}{f(x_i)^2}\sigma^4(w_i)K^2(\psi)f(w_i)^2 dw_i d\psi$$

$$= \left(\int K^2(\psi)d\psi\right)\left(\int\frac{1}{f(x_i)^2}\sigma^4(w_i)f(w_i)^2 dw_i\right)$$

$$\therefore \; n\left(|H_v|\,|H_x|\right)^{1/2} S_{2n} \xrightarrow{d} N(0,\Sigma),$$

where $\Sigma = \left(\int K^2(\psi)d\psi\right)\left(\int \frac{1}{f(x_i)^2}\sigma^4(w_i)f(w_i)^2 dw_i\right)$.

$$n\,|H_v|^{1/2}\, S_{2n} = (n^{-1}\,|H_x|^{-1/2})(n\left(|H_v|\,|H_x|\right)^{1/2} S_{2n})$$

$$\to 0 \text{ as } n \to \infty$$

2) $t = j,\; j \neq i,\; t \neq i$

$$S_{2n} = -\frac{2}{n(n-1)\,|H_v|\,|H_x|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{f(x_i)}u_j^2 K\left(H_x^{-1}(x_j - x_i), H_v^{-1}(v_j - v_i)\right)$$

$$\begin{aligned}
E[\psi_n^2(W_i,W_j)] &= E\left[\frac{1}{f^2(x_i)\left(|H_v|\,|H_x|\right)^2}u_j^2 K^2\left(H_x^{-1}(x_j - x_i), H_v^{-1}(v_j - v_i)\right)\right] \\
&= \int \frac{1}{f^2(x_i)\left(|H_v|\,|H_x|\right)^2}\sigma^2(v_j)K^2\left(H_x^{-1}(x_j - x_i), H_v^{-1}(v_j - v_i)\right) \\
&\times f(w_i)f(w_j)dw_i dw_j \\
&= \int \frac{1}{f^2(x_i)\left(|H_v|\,|H_x|\right)^2}\sigma^2(v_i + H_v\psi_v)K^2\left(\psi\right)f(w_i)f(w_i + HH_x\psi)h dw_i d\psi \\
&= \frac{1}{|H|}\int \frac{1}{f^2(x_i)}\sigma^2(v_i)K^2\left(\psi\right)f^2(w_i)dw_i d\psi \\
&= \frac{1}{|H_v|\,|H_x|}\left(\int K^2\left(\psi\right)d\psi\right)\int \frac{1}{f^2(x_i)}\sigma^2(v_i)f^2(w_i)dw_i \\
&= O((|H_v|\,|H_x|)^{-1})
\end{aligned}$$

$$\therefore \; S_{2n} = n^{-1}\left(E[\psi_n^2(W_i,W_j)]\right)^{1/2} = O(n^{-1}\left(|H_v|\,|H_x|\right)^{-1/2})$$

3) $j \neq t, t \neq i$

$$
\begin{aligned}
S_{2n} &= -\frac{2}{n^2(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j \neq t}}^{n} \sum_{t=1}^{n} \frac{1}{f(x_i)\,|H_v|\,|H_x|} u_j u_t K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right) \\
&= -\frac{2}{3n^2(n-1)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ i \neq j \neq t}}^{n} \sum_{t=1}^{n} \left[\psi_n(W_i, W_j, W_t) + \psi_n(W_j, W_t, W_i) + \psi_n(W_t, W_i, W_j)\right] \\
&= -\frac{2}{n^2(n-1)} \sum \sum_{i<j<t}^{n} \sum 2\phi_n(W_i, W_j, W_t) \\
&= \left(-\frac{2}{n^2(n-1)} - \binom{n}{3}^{-1} + \binom{n}{3}^{-1} \right) \sum \sum_{i<j<t}^{n} \sum 2\phi_n(W_i, W_j, W_t)
\end{aligned}
$$

By H-decomposition,

$$
\begin{aligned}
& E[\phi_n(W_i, W_j, W_t) \mid W_j, W_t] \\
&= u_j u_t E\left[\frac{1}{f(x_i)\,|H_x|\,|H_v|} K\left(H_x^{-1}(x_t - x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right) \mid W_j, W_t \right] \\
&= \phi_{2n}(W_j, W_t)
\end{aligned}
$$

Then, I can write U-statistic

$$
U_n = \frac{6}{n(n-1)} \sum \sum_{j<t}^{n} \phi_{2n}(W_j, W_t) + O_p(H_n^{(3)})
$$

$$
\begin{aligned}
&E[\phi_{2n}^2(W_j,W_t)]\\
&= E\left[\left(u_j u_t E\left[\frac{1}{f(x_i)\,|H_v|\,|H_x|}K\left(H_x^{-1}(x_t-x_i)\right)K\left(H_v^{-1}(v_j-v_i)\right)\mid Z_j,Z_t\right]\right)^2\right]\\
&= E\left[u_j^2u_t^2\left(\int\frac{1}{f(x_i)\,|H_v|\,|H_x|}K\left(H_x^{-1}(x_t-x_i)\right)K\left(H_v^{-1}(v_j-v_i)\right)f(w_i)dw_i\right)^2\right]\\
&= E\Bigg[E\Bigg[u_j^2u_t^2\Bigg(\int\frac{1}{f(x_t-H_x\psi_{x,1})\,|H_v|\,|H_x|}K\left(\psi_{x,1}\right)K\left(\psi_{v,1}+H_v^{-1}(v_j-v_t)\right)\\
&\times f(w_t-H_vH_x\psi_1)\,|H_v|\,|H_x|\,d\psi_1\Bigg)^2\mid Z_j,Z_t\Bigg]\Bigg]\\
&= E\Bigg[\sigma^2(v_j)\sigma^2(v_t)\Bigg(\int\frac{1}{f(x_t-H_x\psi_{x,1})}K\left(\psi_{x,1}\right)K\left(\psi_{v,1}+H_v^{-1}(v_j-v_t)\right)\\
&\times f(w_t-H_vH_x\psi_1)\psi_1\Bigg)^2\Bigg]\\
&= \int\sigma^2(v_j)\sigma^2(v_t)\Bigg(\int\frac{1}{f(x_t-H_x\psi_{x,1})}K\left(\psi_{x,1}\right)K\left(\psi_{v,1}+H_v^{-1}(v_j-v_t)\right)\\
&\times f(w_t-H_vH_x\psi_1)d\psi_1\Bigg)^2 f(w_j)f(w_t)dw_jdw_t\\
&= \int\sigma^2(v_t)\sigma^2(v_t+H\psi_{v,2})\Bigg(\int\frac{1}{f(x_t-H_x\psi_{x,1})}K\left(\psi_{x,1}\right)K\left(\psi_{v,1}+\psi_{v,2}\right)\\
&\times f(w_t-H_vH_x\psi_1)d\psi_1\Bigg)^2 f(w_t)f(w_t+H_vH_x\psi_2)\,|H_v|\,|H_x|\,d\psi_2dw_t\\
&= |H_v|\,|H_x|\left(\int\sigma^4(v_t)\frac{f^4(w_t)}{f^2(x_t)}dw_t\right)\int\left(\int K\left(\psi_{x,1}\right)K\left(\psi_{v,1}+\psi_{v,2}\right)d\psi_1\right)^2d\psi_2\\
&= O(|H_v|\,|H_x|)
\end{aligned}
$$

I have $\sigma_{3n}^2 = O(1)$, $Var(H_n^{(3)}) = O(n^{-3}) = o(n^{-2}\,|H_v|^2)$.

$$\therefore\ S_{2n} = O(n^{-2}\,|H_v|^{1/2})$$

(I)-(B)-(iii)

Note that $\frac{1}{\hat{f}(x_i)} - \frac{1}{f(x_i)} = O_p\left(\left(\frac{\ln n}{n|H_x|}\right)^{1/2}\right) + O_p(|H_x|)$.

$$
\begin{aligned}
S_{3n} &= -\frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_j\left(\hat{\mu}_n(x_i)-\mu_n(x_i)\right)K\left(H_v^{-1}(v_j-v_i)\right)\\
&= -\frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_j\left(\frac{1}{\hat{f}(x_i)}-\frac{1}{f(x_i)}\right)K\left(H_x^{-1}(x_t-x_i)\right)\\
&\times\left(\frac{1}{2}m^*(x_{it})+u_t\right)K\left(H_v^{-1}(v_j-v_i)\right)\\
&\leq\left(O_p\left(\left(\frac{\ln n}{n\,|H_x|}\right)^{1/2}\right)+O_p(|H_x|)\right)\left[-\frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_j\left(\frac{1}{2}m^*(x_{it})+u_t\right)\right.\\
&\left.\times\, K\left(H_x^{-1}(x_t-x_i)\right)K\left(H_v^{-1}(v_j-v_i)\right)\right]\\
&= O_p(n^{-5/2}\,|H_x|^{-1}\,|H_v|^{-1/2}\,(\ln n)^{1/2}) + O_p(n^{-2}\,|H_x|^{1/2}\,|H_v|^{-1/2})\\
&= o((n\,|H_v|^{1/2})^{-1})
\end{aligned}
$$

(I)-(C)

By letting $S_n(z_t) = (\sum_{t=1}^{n} z_t K_{it} z_t)^{-1}$ and $S(z_t) = \begin{pmatrix} f(x_t) & 0 \\ 0 & f(x_t)\sigma_k^2 \end{pmatrix}$,

$$
\begin{aligned}
&|\hat{\mu}_n(x_i)-\mu_n(x_i)|\\
=&\frac{1}{n\,|H_x|}\left|e_1' S_n(z_t)^{-1}\sum_{t=1}^{n} z_t K\left(H_x^{-1}(x_t-x_i)\right)\left(\frac{1}{2}m^*(x_{it})+u_t\right)\right.\\
&\left.-\frac{1}{f(x_i)}\sum_{t=1}^{n} z_t K\left(H_x^{-1}(x_t-x_i)\right)\left(\frac{1}{2}m^*(x_{it})+u_t\right)\right|\\
=&\frac{1}{n\,|H_x|}\left|e_1'(S_n(z_t)^{-1}-S(z_t)^{-1})\sum_{t=1}^{n} z_t K\left(H_x^{-1}(x_t-x_i)\right)\left(\frac{1}{2}m^*(x_{it})+u_t\right)\right|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{1}{|H_x|}((1,0)(S_n(z_t)^{-1} - S(z_t)^{-1})^2(1,0)')^{1/2} \\
&\times \frac{1}{n}\left(\left|\sum_{t=1}^{n} K\left(H_x^{-1}(x_t - x_i)\right)\left(\frac{1}{2}m^*(x_{it}) + u_t\right)\right|\right. \\
&+\left.\left|\sum_{t=1}^{n}\left(H_x^{-1}(x_t - x_i)\right)' K\left(H_x^{-1}(x_t - x_i)\right)\left(\frac{1}{2}m^*(x_{it}) + u_t\right)\right|\right)
\end{aligned}
$$

I follow Lemma 2 of Martins-Filho and Yao (2007) to obtain

$$
|\hat{\mu}_n(x_i) - \mu_n(x_i)| = O\left(|H_x|\left(\frac{\ln n}{n\,|H_x|}\right)^{\frac{1}{2}}\right) + O(|H_x|^3).
$$

Then, I have

$$
\sup|\hat{m}(x_i) - m(x_i)| = O\left(\left(\frac{\ln n}{n\,|H_x|}\right)^{\frac{1}{2}}\right) + O(|H_x|^2)
$$

Using $\hat{m}(x_i) = m^{(1)}(x_{ij})(x_i - x_j)$, where $x_{ij} = \lambda x_i + (1-\lambda)x_j$,

$$
\begin{aligned}
I_{13n} &= \frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{\mu}_n(x_i)'\hat{\mu}_n(x_j)K\left(H_v^{-1}(v_j - v_i)\right) \\
&= \frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\mu_n(x_i) + \hat{\mu}_n(x_i) - \mu_n(x_i))' \\
&\times(\mu_n(x_j) + \hat{\mu}_n(x_j) - \mu_n(x_j))K\left(H_v^{-1}(v_j - v_i)\right)
\end{aligned}
$$

$$
\begin{aligned}
\sup|\mu_n(x_i) + \hat{\mu}_n(x_i) - \mu_n(x_i)| &\leq \sup|\mu_n(x_i)| + \sup|\hat{\mu}_n(x_i) - \mu_n(x_i)| \\
&= O_p\left(\left(\frac{\ln n}{n\,|H_x|}\right)^{1/2}\right) + O_p(|H_x|^2)
\end{aligned}
$$

$$
\begin{aligned}
&\sup\left|(\mu_n(x_i)+\hat{\mu}_n(x_i)-\mu_n(x_i))'(\mu_n(x_j)+\hat{\mu}_n(x_j)-\mu_n(x_j))\right| \\
&\leq (\sup|\mu_n(x_i)+\hat{\mu}_n(x_i)-\mu_n(x_i)|)^2 \\
&= O_p\left(\frac{\ln n}{n\,|H_x|}\right)+O_p\left(|H_x|^4\right)
\end{aligned}
$$

$$
\begin{aligned}
I_{13n} &\leq (\sup|\mu_n(x_i)+\hat{\mu}_n(x_i)-\mu_n(x_i)|)^2\frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}K\left(H_v^{-1}(v_j-v_i)\right) \\
&= O_p\left(n^{-2}\,|H_v|\left(\frac{\ln n}{n\,|H_x|}\right)\right)+O_p\left(n^{-2}\,|H_v|^{-1}\,|H_x|^4\right)
\end{aligned}
$$

(II)

$$
\begin{aligned}
\sup|\hat{g}(z_i)-g(z_i)-(\hat{g}(z_j)-g(z_j))| &\leq \sup|\hat{g}(z_i)-g(z_i)|+\sup|\hat{g}(z_j)-g(z_j)| \\
&= O\left(\left(\frac{\ln n}{n\,|H_z|}\right)^{\frac{1}{2}}\right)+O(|H_z|^2),
\end{aligned}
$$

$$
\hat{I}_{2n}=\frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{u}_i\hat{u}_jK^{(1)}\left(H_v^{-1}(v_j-v_i)\right)'H_v^{-1}(\hat{g}(z_i)-g(z_i)-(\hat{g}(z_j)-g(z_j)))
$$

$$
\begin{aligned}
\left|\hat{I}_{2n}\right| &\leq \sup\left|H_v^{-1}(\hat{g}(z_i)-g(z_i)-(\hat{g}(z_j)-g(z_j)))\right| \\
&\quad\times\frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{u}_i\hat{u}_jK^{(1)}\left(H_v^{-1}(v_j-v_i)\right) \\
&= \left(\left(\frac{\ln n}{n\,|H_z|\,|H_v|^2}\right)^{\frac{1}{2}}\right)+O\left(\frac{|H_z|^2}{n\,|H_v|}\right) \\
&= o((n\,|H_v|^{1/2})^{-1})
\end{aligned}
$$

(III)

$$\begin{aligned}\sup|\hat{g}(z_i)-g(z_i)-(\hat{g}(z_j)-g(z_j))| &\le \sup|\hat{g}(z_i)-g(z_i)|+\sup|\hat{g}(z_j)-g(z_j)|\\ &= O\left(\left(\frac{\ln n}{n\,|H_z|}\right)^{\frac{1}{2}}\right)+O(|H_z|^2),\end{aligned}$$

$$\begin{aligned}\hat{I}_{3n} = \frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{u}_i\hat{u}_j(\hat{g}(z_i)-g(z_i)-(\hat{g}(z_j)-g(z_j)))' \\ \times H_v^{-1}K^{(2)}\left(H_v^{-1}(v_j-v_i)\right)H_v^{-1}(\hat{g}(z_i)-g(z_i)-(\hat{g}(z_j)-g(z_j)))\end{aligned}$$

$$\begin{aligned}\left|\hat{I}_{3n}\right| &\le C\frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left|\hat{u}_i\hat{u}_jG\left(H_v^{-1}\left(v_j^*-v_i^*\right)\right)\right|\cdot(H_v^{-1}\sup|\hat{g}(z_i)-g(z_i)|)^2\\ &= O\left(\frac{\ln n}{n\,|H_z|\,|H_v|^2}\right)+O\left(\frac{|H_z|^4}{n\,|H_v|^2}\right)=o((n\,|H_v|^{1/2})^{-1})\end{aligned}$$

(IV)

$$\begin{aligned}\hat{\Omega} &= \frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{u}_i^2\hat{u}_j^2K^2\left(H_v^{-1}\left(\hat{v}_j-\hat{v}_i\right)\right)\\ &= \frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{u}_i^2\hat{u}_j^2K^2\left(H_v^{-1}\left(v_j-v_i\right)\right)+o_p(1)\\ &= \frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}u_i^2u_j^2K^2\left(H_v^{-1}\left(v_j-v_i\right)\right)+o_p(1)\\ &= \frac{2}{n(n-1)\,|H_v|}\sum^{n}\sum_{i<j}^{n}\left[u_i^2u_j^2K^2\left(H_v^{-1}\left(v_j-v_i\right)\right)+u_j^2u_i^2K^2\left(H_v^{-1}\left(v_i-v_j\right)\right)\right]\end{aligned}$$

By using the properties of U-statistics, $\Omega = 2\int K^2(\psi)d\psi E[\sigma^4(v)f(v)]$.

B.2 Proof of Theorem 4

Under the alternative, $m_1(x_i, v_i) = m(x_i) + \delta_n l(v_i)$. Then, $u_i = \varepsilon_i + \delta_n l(v_i)$, where $\varepsilon_i = y_i - m(x_i, v_i)$ and $\delta_n = n^{-1/2} |H_v|^{-1/4}$.

$$
\begin{aligned}
I_n &= \frac{1}{n(n-1)|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{u}_i \hat{u}_j \hat{K}\left(H_v^{-1}(v_j - v_i)\right) \\
&= \hat{I}_{1nG} + o((n|H_v|^{1/2})^{-1})
\end{aligned}
$$

$$
\begin{aligned}
\hat{I}_{1nG} &= \frac{1}{n(n-1)|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{u}_i \hat{u}_j K\left(H_v^{-1}(v_j - v_i)\right) \\
&= \frac{1}{n(n-1)|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_i u_j K\left(H_v^{-1}(v_j - v_i)\right) \\
&- \frac{2}{n(n-1)|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j(\hat{m}(x_i) - m(x_i)) K\left(H_v^{-1}(v_j - v_i)\right) \\
&+ \frac{1}{n(n-1)|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} (\hat{m}(x_i) - m(x_i))(\hat{m}(x_j) - m(x_j)) K\left(H_v^{-1}(v_j - v_i)\right) \\
&= I_{11nG} + I_{12nG} + I_{13nG}
\end{aligned}
$$

For the following sections, I will show

$$
n|H_v|^{1/2} I_{11nG} \xrightarrow{d} N(0, \Omega)
$$

$$
\begin{aligned}
I_{11nG} &= \frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} u_i u_j K\left(H_v^{-1}\left(v_j - v_i\right)\right) \\
&= \frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} (\varepsilon_i + \delta_n l(v_i))(\varepsilon_j + \delta_n l(v_j)) K\left(H_v^{-1}\left(v_j - v_i\right)\right) \\
&= \frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \varepsilon_i \varepsilon_j K\left(H_v^{-1}\left(v_j - v_i\right)\right) \\
&+ \frac{2\delta_n}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \varepsilon_i l(v_j) K\left(H_v^{-1}\left(v_j - v_i\right)\right) \\
&+ \frac{\delta_n^2}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} l(v_i) l(v_j) K\left(H_v^{-1}\left(v_j - v_i\right)\right) \\
&= Q_{1n} + 2\delta_n Q_{2n} + \delta_n^2 Q_{3n}
\end{aligned}
$$

(I)-(A)

$$
\begin{aligned}
Q_{1n} &= \frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} \varepsilon_i \varepsilon_j K\left(H_v^{-1}\left(v_j - v_i\right)\right) \\
&= \frac{1}{n(n-1)\,|H_v|}\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} \left[\varepsilon_i \varepsilon_j K\left(H_v^{-1}\left(v_j - v_i\right)\right) + \varepsilon_j \varepsilon_i K\left(H_v^{-1}\left(v_i - v_j\right)\right)\right]
\end{aligned}
$$

By letting $W_i = (v_i, \varepsilon_i)$,

$$
\begin{aligned}
Q_{1n} &= \frac{1}{n(n-1)\,|H_v|}\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} \left[\psi_n(W_i, W_j) + \psi_n(W_j, W_i)\right] \\
&= \frac{1}{n(n-1)\,|H_v|}\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n} \phi_n(W_i, W_j)
\end{aligned}
$$

As $\mathrm{E}[\phi_n^2(W_i,W_j)] = \mathrm{E}[\psi_n^2(W_i,W_j)] + \mathrm{E}[\psi_n^2(W_j,W_i)] + 2\mathrm{E}[\psi_n(W_i,W_j)\psi_n(W_j,W_i)]$,

$$\frac{1}{|H_v|}E[\psi_n^2(W_i,W_j)] \to \left(\int K^2(\psi_v)\,d\psi_v\right)E\left[\sigma^4(v_i)f(v_i)\right] < \infty$$

$$\frac{2}{|H_v|}E[\psi_n(W_i,W_j)\psi_n(W_j,W_i)] \to 2\left(\int K^2(\psi_v)\,d\psi_v\right)E\left[\sigma^4(v_i)f(v_i)\right]$$

$$\therefore\ n\,|H_v|^{1/2}\,Q_{1n} \xrightarrow{d} N(0,\Omega),$$

where $\Omega = 2\left(\int K^2(\psi_v)\,d\psi_v\right)E\left[\sigma^4(v_i)f(v_i)\right]$.

(I)-(B)

$$\begin{aligned}
Q_{2n} &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{|H_v|}\varepsilon_i l(v_j)K\left(H_v^{-1}(v_j-v_i)\right) \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\frac{1}{|H_v|}\varepsilon_i l(v_j)K\left(H_v^{-1}(v_j-v_i)\right) + \frac{1}{|H_v|}\varepsilon_j l(v_i)K\left(H_v^{-1}(v_j-v_i)\right)\right] \\
&= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left[\psi_n(W_i,W_j)+\psi_n(W_j,W_i)\right],\ \text{where } W_i=(v_i,\varepsilon_i)
\end{aligned}$$

$E[\psi_n(W_i,W_j)] = 0$. By applying Lipschitz condition,

$$\begin{aligned}
E[\psi_n^2(W_i,W_j)] &= E\left[\frac{1}{|H_v|^2}\varepsilon_i^2 l(v_j)^2K^2\left(H_v^{-1}(v_j-v_i)\right)\right] \\
&= \frac{1}{|H_v|^2}\int \sigma^2(v_i)l(v_j)^2K^2\left(H_v^{-1}(v_j-v_i)\right)f(v_i)f(v_j)dv_idv_j \\
&= \frac{1}{|H_v|}\int \sigma^2(v_i)l(v_i+H_v\psi)^2K^2(\psi)\,f(v_i)f(v_i+H_v\psi)dv_id\psi \\
&= \frac{1}{|H_v|}\left(\int K^2(\psi)\,d\psi\right)\left(\int \sigma^2(v_i)l(v_i)^2f(v_i)^2dv_i\right)
\end{aligned}$$

As $Q_{2n} = n^{-1} \left(E[\phi_n^2(W_i, W_j)]\right)^{\frac{1}{2}} = O(n^{-1} |H_v|^{-1/2})$,

$$\therefore n |H_v|^{1/2} \delta_n Q_{2n} = n^{-1/2} |H_v|^{-1/4} \xrightarrow{p} 0$$

(I)-(C)

$$Q_{3n} = \frac{1}{n(n-1)|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} l(v_i) l(v_j) K\left(H_v^{-1}(v_j - v_i)\right)$$

$$\begin{aligned}
&\frac{1}{|H_v|} E\left[l(v_i) l(v_j) K\left(H_v^{-1}(v_j - v_i)\right)\right] \\
=&\frac{1}{|H_v|} \int l(v_i) l(v_i + H_v\psi) K(\psi) f(v_i) f(v_i + H_v\psi) |H_v| \, dv_i d\psi \\
=&\int l(v_i)^2 K(\psi) f(v_i)^2 dv_i d\psi \\
=&\left(\int K(\psi) \, d\psi\right)\left(\int l(v_i)^2 f(v_i)^2 dv_i\right) \\
=&E[l(v_i)^2 f(v_i)] \\
=&O_p(1)
\end{aligned}$$

$$n |H_v|^{1/2} \delta_n^2 Q_{3n} = n |H_v|^{1/2} (n^{-1} |H_v|^{-1/2}) Q_{3n} = Q_{3n} \xrightarrow{p} E[l(v_i)^2 f(v_i)]$$

$$\therefore \; n |H_v|^{1/2} I_{11nG} \xrightarrow{d} N(E[l(v_i)^2 f(v_i)], \Omega)$$

B.3 Proof of Theorem 5

Under $\mathbb{H}_1$,

$$\begin{aligned}
\hat{u}_i &= y_i - \hat{m}(x_i) \\
&= y_i - m(x_i, v_i) + m(x_i, v_i) - \hat{m}(x_i) \\
&= \varepsilon_i + m(x_i, v_i) - \hat{m}(x_i) \\
&= \varepsilon_i + (m(x_i, v_i) - m(x_i)) - (\hat{m}(x_i) - m(x_i)) \\
&= u_i - (\hat{m}(x_i) - m(x_i))
\end{aligned}$$

The test statistic is then written as follows:

$$\begin{aligned}
I_n &= \frac{1}{n(n-1)\left|H_v\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{u}_i \hat{u}_j K\left(H_v^{-1}\left(\hat{v}_j - \hat{v}_i\right)\right) \\
&= \frac{1}{n(n-1)\left|H_v\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_i u_j K\left(H_v^{-1}\left(v_j - v_i\right)\right) + o(\left(n\left|H_v\right|^{1/2}\right)^{-1}) \\
&= \frac{1}{n(n-1)\left|H_v\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} (\varepsilon_i + h(v_i))(\varepsilon_j + h(v_i)) K\left(H_v^{-1}\left(v_j - v_i\right)\right)
\end{aligned}$$

Define $\phi_n(W_i, W_j) = \frac{1}{|H_v|} h(v_i) h(v_j) K\left(H_v^{-1}\left(v_j - v_i\right)\right)$.

$$\begin{aligned}
E[\phi_n(W_i, W_j)] &= E[E[\phi_n(W_i, W_j) \mid W_i, W_j]] \\
&= E\left[\frac{1}{|H_v|} K\left(H_v^{-1}\left(v_j - v_i\right)\right) h(v_i) h(v_j)\right] \\
&= \frac{1}{|H_v|} \int K\left(H_v^{-1}\left(v_j - v_i\right)\right) h(v_i) h(v_j) f(v_i) f(v_j) dv_i dv_j \\
&= \int K\left(\psi\right) h(v_i) h(v_i + H_v \psi) f(v_i) f(v_i + H_v \psi) dv_i d\psi
\end{aligned}$$

$$= \int K\left(\psi\right)d\psi \int (h(v_i))^2 f(v_i)^2 dv_i$$

$$= E[(h(v_i))^2 f(v_i)]$$

$$\begin{aligned}
\hat{\Omega} &= \frac{2}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{u}_i^2\hat{u}_j^2 K^2\left(H_v^{-1}\left(\hat{v}_j-\hat{v}_i\right)\right) \\
&= \frac{2}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{u}_i^2\hat{u}_j^2 K^2\left(H_v^{-1}\left(v_j-v_i\right)\right) + o\left((n\left|H_v\right|)^{-1}\right)
\end{aligned}$$

$$\begin{aligned}
&E\left[\phi_n(W_i,W_j)\right] = E\left[E\left[\phi_n(W_i,W_j) \mid v_i\right]\right] \\
&= E\left[\frac{1}{\left|H_v\right|}K^2\left(H_v^{-1}\left(v_j-v_i\right)\right)\left(\sigma^2(v_i)+(h(v_i))^2\right)\left(\sigma^2(v_j)+(h(v_j))^2\right)\right] \\
&= \frac{1}{\left|H_v\right|}\int K^2\left(H_v^{-1}\left(v_j-v_i\right)\right)\left(\sigma^2(v_i)+(h(v_i))^2\right)\left(\sigma^2(v_j)+(h(v_j))^2\right) f(v_i)f(v_j)dv_i dv_j \\
&= \int K^2\left(\psi\right)\left(\sigma^2(v_i)+(h(v_i))^2\right)\left(\sigma^2(v_i+H_v\psi)+(h(v_i+H_v\psi))^2\right) f(v_i)f(v_i+H_v\psi)dv_i d\psi \\
&= \left(\int K^2(\psi)d\psi\right)\left(\int \left(\sigma^2(v_i)+(h(v_i))^2\right)^2 f(v_i)^2 dv_i\right) \\
&= \left(\int K^2(\psi)d\psi\right)\left[E\left[\sigma^4(v_i)f(v_i)\right]+2E\left[\sigma^2(v_i)(h(v_i))^2 f(v_i)\right]+E\left[(h(v_i))^2 f(v_i)\right]\right] \\
&= B_1
\end{aligned}$$

$$\hat{\Omega} \xrightarrow{p} 2B_1$$

$$J_n = \frac{n\left|H_v\right|^{1/2} I_n}{\sqrt{\hat{\Omega}}} > c_n = o_p(n\left|H_v\right|^{1/2})$$

Appendix C

Appendix for Chapter 4

C.1 Proof of Theorem 8

An estimated individual-specific component under the random effects model as follows:

$$\hat{\alpha}_i = \frac{T\hat{\sigma}_\alpha^2}{T\hat{\sigma}_\alpha^2 + \hat{\sigma}_v^2} \frac{\sum_{s=1}^T \sum_{t=1}^T \hat{u}_{it}(H^{-1}(X_{is} - X_{it})}{\sum_{s=1}^T \sum_{t=1}^T K(H^{-1}(X_{is} - X_{it})}$$

Define $\hat{\rho} = \frac{T\hat{\sigma}_\alpha^2}{T\hat{\sigma}_\alpha^2+\hat{\sigma}_v^2}$, and $\bar{\hat{u}}_i = \sum_{s=1}^T \sum_{t=1}^T \hat{u}_{it}(H^{-1}(X_{is}-X_{it})/\sum_{s=1}^T \sum_{t=1}^T K(H^{-1}(X_{is}-X_{it})$.

$$\begin{aligned}
\hat{\alpha}_i &= \rho\bar{\hat{u}}_i + (\hat{\rho} - \rho)\bar{\hat{u}}_i \\
&= \rho(\bar{\hat{u}}_i - \bar{u}_i + \bar{u}_i) + (\hat{\rho} - \rho)(\bar{\hat{u}}_i - \bar{u}_i + \bar{u}_i) \\
&= \rho\bar{u}_i - \rho(\bar{\hat{m}}_i(x) - \bar{m}_i(x)) + O_p(nT^{-1}) \\
&= \rho\bar{u}_i - \rho\hat{\mu}_{n,i}(x) + O_p((nT)^{-1})
\end{aligned}$$

The test statistic is given as below.

$$
\begin{aligned}
I_n &= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{\alpha}_i\hat{\alpha}_j K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right) \\
&= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\rho\bar{u}_i-\rho\hat{\mu}_{n,i}(x))(\rho\bar{u}_j-\rho\hat{\mu}_{n,j}(x))K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right) \\
&= \frac{\rho^2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_i\bar{u}_j K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right) \\
&- \frac{2\rho^2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_i\hat{\mu}_{n,j}(x) K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right) \\
&+ \frac{\rho^2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{\mu}_{n,i}(x)\hat{\mu}_{n,j}(x) K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right) \\
&\equiv I_{1n}+I_{2n}+I_{3n}
\end{aligned}
$$

Following Lemma 1 from Yao and Ullah (2013), define second-order U-statistic $U_n = \frac{1}{n(n-1)}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\phi_n(X_i,X_j)$, where $\phi_n(X_i,X_j)$ is symmetric function of X_j and X_i, where $\{X_i\}_{i=1}^n$ is a sequence of IID random variables. $E[\phi_n(X_i,X_j)\mid X_j]=0$. I can easily verify that I_{1n} is a degenerated second order U-statistic.

$$
\begin{aligned}
I_{1n} &= \frac{\rho^2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_i\bar{u}_j K(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)) \\
&= \frac{\rho^2}{n(n-1)\left|\bar{H}\right|}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}[\bar{u}_i\bar{u}_j K(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i))+\bar{u}_j\bar{u}_i K(\bar{H}^{-1}(\bar{X}_i-\bar{X}_j))] \\
&= \frac{\rho^2}{n(n-1)\left|\bar{H}\right|}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}[\psi_n(W_i,W_j)+\psi_n(W_j,W_i)], \text{ where } W_i=(\bar{u}_i,\bar{X}_i) \\
&= \frac{\rho^2}{n(n-1)\left|\bar{H}\right|}\sum_{\substack{i=1\\i<j}}^{n}\sum_{j=1}^{n}\phi_n(W_i,W_j)
\end{aligned}
$$

As $E[\phi_n^2(W_i,W_j)] = E[\psi_n^2(W_i,W_j)] + E[\psi_n^2(W_j,W_i)] + 2E[\psi_n(W_i,W_j)\psi_n(W_j,W_i)]$,

$$\begin{aligned}
\frac{1}{|\bar{H}|}E\left[\psi_n^2(W_j,W_i)\right] &= \frac{1}{|\bar{H}|}E\left[K^2(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i))\bar{u}_i^2\bar{u}_j^2\right] \\
&= \frac{1}{|\bar{H}|}E\left[K^2(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i))\sigma^4\right] \\
&= \frac{1}{|\bar{H}|}\sigma^4\int K^2(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i))f(\bar{X}_i)f(\bar{X}_j)d\bar{X}_id\bar{X}_j \\
&= \sigma^4\int K^2(\varphi)f(\bar{X}_i)f(\bar{X}_i+\bar{H}\varphi)d\varphi d\bar{X}_i \\
&\to \sigma^4\int K^2(\varphi)d\varphi E[f(\bar{X}_i)] < \infty
\end{aligned}$$

$$\begin{aligned}
\therefore\ \frac{1}{|\bar{H}|}E\left[\psi_n(W_i,W_j)\psi_n(W_j,W_i)\right] &= \frac{1}{|\bar{H}|}E\left[K^2(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i))\sigma^4\right] \\
&\to \sigma^4\int K^2(\varphi)d\varphi E[f(\bar{X}_i)] < \infty,
\end{aligned}$$

which results in $(E[\phi_n^2(W_i,W_j)])^2 = O(|\bar{H}|^2)$.

By Cr inequality,

$$E[\phi_n^4(W_i,W_j)] \leq C\left[E[\psi_n^4(W_i,W_j)] + E[\psi_n^4(W_j,W_i)]\right].$$

$$\begin{aligned}
\frac{1}{|\bar{H}|}E[\psi_n^4(W_i,W_j)] &= \frac{1}{|\bar{H}|}E\left[E\left[K^4\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)\bar{u}_i^4\bar{u}_j^4 \mid \bar{X}_i,\bar{X}_j\right]\right] \\
&= \frac{1}{|\bar{H}|}E\left[\sigma^4(\bar{X}_i)\sigma^4(v_j)K^4\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)\right] \\
&= \int K^4(\varphi)\sigma^4(\bar{X}_i)\sigma^4(\bar{X}_i+\bar{H}\varphi)f(\bar{X}_i)f(\bar{X}_i+\bar{H}\varphi)d\bar{X}_id\varphi
\end{aligned}$$

$$= \int K^4(\varphi)\sigma^8(\bar{X}_i)f^2(\bar{X}_i)d\bar{X}_i d\varphi$$
$$= \left(\int K^4(\varphi)d\varphi\right)\left(\int \sigma^8(\bar{X}_i)f^2(\bar{X}_i)d\bar{X}_i\right)$$

Here, I have $\frac{1}{n}\mathrm{E}[\phi_n^4(W_i, W_j)] = O(n^{-1}\left|\bar{H}\right|)$.

Lastly, define

$$\begin{aligned}
G_n(W_i, W_j) &= E\left[\phi_n(W_t, W_i)\phi_n(W_t, W_j) \mid W_i, W_j\right] \\
&= E\left[(\psi_n(W_t, W_i) + \psi_n(W_i, W_t))(\psi_n(W_t, W_j) + \psi_n(W_j, W_i)) \mid W_i, W_j\right] \\
&= E[\psi_n(W_t, W_i)\psi_n(W_t, W_j) \mid W_i, W_j] + E[\psi_n(W_t, W_i)\psi_n(W_j, W_t) \mid W_i, W_j] \\
&+ E[\psi_n(W_i, W_t)\psi_n(W_t, W_j) \mid W_i, W_j] + E[\psi_n(W_i, W_t)\psi_n(W_j, W_t) \mid W_i, W_j] \\
&= G_1(W_i, W_j) + G_2(W_i, W_j) + G_3(W_i, W_j) + G_4(W_i, W_j)
\end{aligned}$$

By Cr inequality,

$$\begin{aligned}
E\left[G_n^2(W_i, W_j)\right] &= E[(G_1(W_i, W_j) + G_2(W_i, W_j) + G_3(W_i, W_j) + G_4(W_i, W_j))^2] \\
&\leq C\left[G_1^2(W_i, W_j) + G_2^2(W_i, W_j) + G_3^2(W_i, W_j) + G_4^2(W_i, W_j)\right]
\end{aligned}$$

$$\begin{aligned}
&E[G_1^2(W_i, W_j)] \\
=&E\left[E\left[\psi_n(W_l, W_i)\psi_n(W_l, W_j) \mid W_i, W_j\right]^2\right] \\
=&E\left[E\left[K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_l)\right)\bar{u}_i\bar{u}_j\bar{u}_l^2 \mid \bar{X}_i, \bar{X}_j\right]^2\right] \\
=&E\left[\left(\bar{u}_i\bar{u}_j E\left[K\left(\bar{H}^{-1}(\bar{X}_i - \bar{X}_l)\right)K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_l)\right)\sigma^2(\bar{X}_l) \mid \bar{X}_i, \bar{X}_j\right]\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
=&E\left[\left(\bar{u}_i\bar{u}_j\int K\left(\bar{H}^{-1}(\bar{X}_i-\bar{X}_l)\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_l)\right)\sigma^2(\bar{X}_l)f(\bar{X}_l)d\bar{X}_l\right)^2\right]\\
=&E\Bigg[E\Bigg[\bar{u}_i^2\bar{u}_j^2\left(\int K\left(\varphi_1\right)K\left(\varphi_1+\bar{H}^{-1}(\bar{X}_i-\bar{X}_j)\right)\sigma^2(\bar{X}_i+\bar{H}\varphi_1)f(\bar{X}_i+\bar{H}\varphi_1)\left|\bar{H}\right|d\varphi_1\right)^2\\
&\times\Bigg|\bar{X}_i,\bar{X}_j\Bigg]\Bigg]\\
=&E\Bigg[\sigma^2(\bar{X}_i)\sigma^2(\bar{X}_j)\Bigg(\int K\left(\varphi_1\right)K\left(\varphi_1+\bar{H}^{-1}(\bar{X}_i-\bar{X}_j)\right)\sigma^2(\bar{X}_i+\bar{H}\varphi_1)f(\bar{X}_i+\bar{H}\varphi_1)\\
&\times\left|\bar{H}\right|d\varphi_1\Bigg)^2\Bigg]\\
=&\left|\bar{H}\right|^2\int\sigma^2(\bar{X}_i)\sigma^2(\bar{X}_j)\Bigg(\int K\left(\varphi_1\right)K\left(\varphi_1+\bar{H}^{-1}(\bar{X}_i-\bar{X}_j)\right)\sigma^2(\bar{X}_i+\bar{H}\varphi_1)\\
&\times f(\bar{X}_i+\bar{H}\varphi_1)d\varphi_1\Bigg)^2 f(\bar{X}_i)f(\bar{X}_j)d\bar{X}_id\bar{X}_j\\
=&\left|\bar{H}\right|^2\int\sigma^2(\bar{X}_i)\sigma^2(\bar{X}_i-\bar{H}\varphi_2)\left(\int K\left(\varphi_1\right)K\left(\varphi_1+\varphi_2\right)\sigma^2(\bar{X}_i+\bar{H}\varphi_1)f(\bar{X}_i+\bar{H}\varphi_1)d\varphi_1\right)^2\\
&\times f(\bar{X}_i)f(\bar{X}_i-\bar{H}\varphi_2)\left|\bar{H}\right|d\bar{X}_id\varphi_2\\
=&\left|\bar{H}\right|^3\int\sigma^4(\bar{X}_i)\left(\int K\left(\varphi_1\right)K\left(\varphi_1+\varphi_2\right)\sigma^2(\bar{X}_i)f(\bar{X}_i)d\varphi_1\right)^2 f^2(\bar{X}_i)d\bar{X}_id\varphi_2\\
=&\left|\bar{H}\right|^3\left(\int\sigma^8(\bar{X}_i)f^4(\bar{X}_i)d\bar{X}_i\right)\int\left(\int K\left(\varphi_1\right)K\left(\varphi_1+\varphi_2\right)d\varphi_1\right)^2d\varphi_2\\
=&O(\left|\bar{H}\right|^3)
\end{aligned}
$$

I can conclude that

$$
\begin{aligned}
\frac{E[G_n^2(W_i,W_j)]+n^{-1}E[\phi_n^4(W_i,W_j)]}{\left(E[\phi_n^2(W_i,W_j)]\right)^2}&=\frac{O(\left|\bar{H}\right|^2)+n^{-1}O(\left|\bar{H}\right|)}{O(\left|\bar{H}\right|^2)}\\
&=O(\left|\bar{H}\right|)+O((n\left|\bar{H}\right|)^{-1})\\
&\to 0 \text{ as } n\to\infty
\end{aligned}
$$

Also, $\frac{1}{|\bar{H}|}\mathrm{E}[\phi_n^2(W_i,W_j)] \to \Omega$, where $\Omega \equiv 2\int K^2(\varphi)d\varphi \mathrm{E}[\sigma^4(\bar{X}_i)f(\bar{X}_i)]$. By applying Hall's Central Limit Theorem,

$$\sqrt{n^2\left|\bar{H}\right|}I_{1n} \xrightarrow{d} N(0,\Omega)$$

(I)-(B)

As $\hat{m}(x_{it})$ is a local linear estimator and $z_{ls}' = (1, \frac{x_{ls}-x_{it}}{h})$,

$$\begin{aligned}
&\hat{m}(x_{it}) - m(x_{it})\\
&= e_1'\frac{1}{2}\left(\sum_{l=1}^{n}\sum_{s=1}^{T} z_{ls}'K\left(H^{-1}(x_{ls}-x_{it})\right)z_{ls}\right)^{-1}\sum_{l=1}^{n}\sum_{s=1}^{T} z_{ls}'K\left(H^{-1}(x_{ls}-x_{it})\right)\\
&\quad\times(x_{ls}-x_{it})m^{(2)}(x_{ls,it})(x_{ls}-x_{it})'\\
&\quad+e_1'\left(\sum_{l=1}^{n}\sum_{s=1}^{T} z_{ls}'K\left(H^{-1}(x_{ls}-x_{it})\right)z_{ls}\right)^{-1}\sum_{l=1}^{n}\sum_{s=1}^{T} z_{it}'K\left(H^{-1}(x_{ls}-x_{it})\right)u_{ls}\\
&= e_1'\left(\sum_{l=1}^{n}\sum_{s=1}^{T} z_{ls}'K\left(H^{-1}(x_{ls}-x_{it})\right)z_{ls}\right)^{-1}\sum_{l=1}^{n}\sum_{s=1}^{T} z_{ls}'K\left(H^{-1}(x_{ls}-x_{it})\right)\\
&\quad\times\left(\frac{1}{2}m^*(x_{ls,it})+u_{ls}\right),
\end{aligned}$$

where $m^*(x_{ls,it}) = (x_{ls}-x_{it})m^{(2)}(x_{ls,it})(x_{ls}-x_{it})'$ and $x_{ls,it} = \lambda x_{ls} + (1-\lambda)x_{it}$.

$$\begin{aligned}
\bar{\hat{m}}_i - \bar{m}_i &= \sum_{t=1}^{T}\sum_{r=1}^{T}(\hat{m}(x_{it}) - m(x_{it}))w_{ir}(x_{it})\\
&= \sum_{t=1}^{T}\sum_{r=1}^{T}\left(e_1'\left(\sum_{l=1}^{n}\sum_{s=1}^{T} z_{ls}'K\left(H^{-1}(x_{ls}-x_{it})\right)z_{ls}\right)^{-1}\right.\\
&\quad\left.\times\sum_{l=1}^{n}\sum_{s=1}^{T} z_{ls}'K\left(H^{-1}(x_{ls}-x_{it})\right)\left(\frac{1}{2}m^*(x_{ls,it})+u_{ls}\right)\right)w_{ir}(x_{it}),
\end{aligned}$$

where $w_{ir}(x_{it}) = K(H^{-1}(x_{ir}-x_{it}))/\sum_{t=1}^{T}\sum_{r=1}^{T}K(H^{-1}(x_{ir}-x_{it}))$.

Define

$$\mu_{n,i} = \sum_{t=1}^{T}\sum_{r=1}^{T} \frac{1}{nT^2 \left|H\right|^2 f_i f(x_{it})} \left(\sum_{l=1}^{n}\sum_{s=1}^{T} K\left(H^{-1}(x_{ls}-x_{it})\right)\left(\frac{1}{2}m^*(x_{ls,it}) + u_{ls}\right)\right)$$

$$\times K(H^{-1}(x_{ir}-x_{it}))$$

$$\hat{\mu}_{n,i} = \bar{\bar{m}}_i - \bar{m}_i = \mu_{n,i} + \hat{\mu}_{n,i} - \mu_{n,i}.$$

$$\begin{aligned}
I_{2n} &= -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n}\sum_{j\neq i}^{n} \bar{u}_j(\bar{\bar{m}}_i(x_{it}) - \bar{m}_i(x_{it}))K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right) \\
&= -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n}\sum_{j\neq i}^{n} \bar{u}_j \left[\sum_{t=1}^{T}\sum_{r=1}^{T} e_1' \left(\sum_{l=1}^{n}\sum_{s=1}^{T} z_{ls}' K\left(H^{-1}(x_{ls}-x_{it})\right) z_{ls} \right)^{-1} \right. \\
&\left. \times \sum_{l=1}^{n}\sum_{s=1}^{T} z_{ls}' K\left(H^{-1}(x_{ls}-x_{it})\right)\left(\frac{1}{2}m^*(x_{ls,it}) + u_{ls}\right) w_{ir}(x_{it}) \right] K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right) \\
&= -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T} \bar{u}_j \ \hat{\mu}_{n,i} K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right) \\
&= -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n}\sum_{j\neq i}^{n}\sum_{t=1}^{T} \bar{u}_j \left(\mu_{n,i} + \hat{\mu}_{n,i} - \mu_{n,i}\right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right) \\
&= -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n}\sum_{j\neq i}^{n} \bar{u}_j \left(\sum_{t=1}^{T}\sum_{r=1}^{T} \frac{1}{2nT^2\left|H\right|^2 f_i f(x_{it})} \left(\sum_{l=1}^{n}\sum_{s=1}^{T} K\left(H^{-1}(x_{ls}-x_{it})\right) \right.\right. \\
&\left.\left. \times m^*(x_{ls,it}) \right) K(H^{-1}(x_{ir}-x_{it})) \right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right) \\
&- \frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n}\sum_{j\neq i}^{n} \bar{u}_j \left(\sum_{t=1}^{T}\sum_{r=1}^{T} \frac{1}{nT^2\left|H\right|^2 f_i f(x_{it})} \left(\sum_{l=1}^{n}\sum_{s=1}^{T} K\left(H^{-1}(x_{ls}-x_{it})\right) u_{ls} \right) \right. \\
&\left. \times K(H^{-1}(x_{ir}-x_{it})) \right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right) \\
&- \frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n}\sum_{j\neq i}^{n} \bar{u}_j \left(\hat{\mu}_{n,i} - \mu_{n,i}\right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right) \\
&\equiv S_{1n} + S_{2n} + S_{3n}
\end{aligned}$$

(I)-(B)-(i)

1) $(l=i, i \neq j, r=s=t)$, $(l=i, i \neq j, s=t, r \neq t)$

$$S_{1n}=0 \qquad \because m^{*}(x_{it,it})=(x_{it}-x_{it})m^{(2)}(x_{it})(x_{it}-x_{it})=0.$$

2) $(l=i, i \neq j, r=s, r \neq t)$

$$\begin{aligned}
S_{1n} &= -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_{j}\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{2T\left|H\right|^{2}f_{i}f(x_{it})}K\left(H^{-1}(x_{ir}-x_{it})\right)\right. \\
&\quad \left.\times\, m^{*}(x_{ir,it})K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_{j}-\bar{X}_{i})\right) \\
&= O((nT|\bar{H}|^{1/2})^{-1}|H|^{1/2})
\end{aligned}$$

3) $(l=i, i \neq j, r=t, s \neq r)$

$$\begin{aligned}
S_{1n} &= -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_{j}\left(\sum_{t=1}^{T}\frac{1}{2T\left|H\right|^{2}f_{i}f(x_{it})}\sum_{s=1}^{T}K\left(H^{-1}(x_{is}-x_{it})\right)\right. \\
&\quad \left.\left.\times\, m^{*}(x_{is,it})\right)K(0)\right)K\left(\bar{H}^{-1}(\bar{X}_{j}-\bar{X}_{i})\right) \\
&= O((nT|\bar{H}|^{1/2})^{-1}|H|^{1/2})
\end{aligned}$$

4) $(l = i, i \neq j, r \neq s \neq t)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{2T^2\left|H\right|^2 f_i f(x_{it})}\left(\sum_{s=1}^{T}K\left(H^{-1}(x_{is}-x_{it})\right)\right.\right.$$
$$\left.\left.\times\, m^*(x_{is,it})\right)K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$
$$= O((nT^2|\bar{H}|^2)^{-1}|H|)$$

5) $(l = j, l \neq i, r = s = t)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\frac{1}{2\left|H\right|^2 f_i f(x_{it})}K\left(H^{-1}(x_{jt}-x_{it})\right)m^*(x_{jt,it})\right.$$
$$\left.\times\, K(0)\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$
$$= O((n|\bar{H}|^2)^{-1}|H|^{1/2})$$

6) $(l = j, l \neq i, r = s, r \neq t)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{2T\left|H\right|^2 f_i f(x_{it})}K\left(H^{-1}(x_{jr}-x_{it})\right)\right.$$
$$\left.\times\, m^*(x_{jr,it})K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$
$$= O((nT|\bar{H}|^{1/2})^{-1}|H|)$$

7) $(l = j, l \neq i, s = t, r \neq s)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{2T\left|H\right|^2 f_i f(x_{it})}K\left(H^{-1}(x_{jt}-x_{it})\right)\right.$$
$$\left.\times\, m^*(x_{jt,it})K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$
$$= O((nT|\bar{H}|^{1/2})^{-1}|H|)$$

8) $(l = j, l \neq i, r = t, s \neq r)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\frac{1}{2T\left|H\right|^2 f_i f(x_{it})}\left(\sum_{s=1}^{T}K\left(H^{-1}(x_{js}-x_{it})\right)\right.\right.$$
$$\left.\left.\times\, m^*(x_{js,it})\right)K(0)\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$
$$= O((nT|\bar{H}|^{1/2})^{-1}|H|)$$

9) $(l = j, l \neq i, r \neq s \neq t)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{2T^2\left|H\right|^2 f_i f(x_{it})}\left(\sum_{s=1}^{T}K\left(H^{-1}(x_{js}-x_{it})\right)\right.\right.$$
$$\left.\left.\times\, m^*(x_{js,it})\right)K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$
$$= O((nT^2|\bar{H}|^{1/2})^{-1}|H|)$$

10) $(i \neq j \neq l, r = s = t)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\frac{1}{2n\left|H\right|^2 f_i f(x_{it})}\left(\sum_{l=1}^{n}K\left(H^{-1}(x_{lt}-x_{it})\right)\right.\right.$$

$$\left.\left.\times\, m^*(x_{lt,it})\right)K(0)\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= O((T^2|\bar{H}|^{1/2})^{-1}|H|^{1/2})$$

11) $(i \neq j \neq l, r = s, r \neq t)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{2nT\left|H\right|^2 f_i f(x_{it})}\left(\sum_{l=1}^{n}K\left(H^{-1}(x_{lr}-x_{it})\right)\right.\right.$$

$$\left.\left.\times\, m^*(x_{lr,it})\right)K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= O((nT|\bar{H}|^{1/4})^{-2}|H|)$$

12) $(i \neq j \neq l, s = t, r \neq s)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{2nT\left|H\right|^2 f_i f(x_{it})}\left(\sum_{l=1}^{n}K\left(H^{-1}(x_{lt}-x_{it})\right)\right.\right.$$

$$\left.\left.\times\, m^*(x_{lt,it})\right)K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= O((n^2T|\bar{H}|^{1/2})^{-1}|H|)$$

13) $(i \neq j \neq l, r = t, s \neq r)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \bar{u}_j \left(\sum_{t=1}^{T} \frac{1}{2nT \left|H\right|^2 f_i f(x_{it})} \left(\sum_{l=1}^{n} \sum_{s=1}^{T} K\left(H^{-1}(x_{ls} - x_{it})\right) \times m^*(x_{ls,it}) \right) K(0) \right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)$$

$$= O((n^2 T |\bar{H}|^{1/2})^{-1} |H|^{1/2})$$

14) $(i \neq j \neq l, r \neq s \neq t)$

$$S_{1n} = -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \bar{u}_j \left(\sum_{t=1}^{T} \sum_{r=1}^{T} \frac{1}{2nT^2 \left|H\right|^2 f_i f(x_{it})} \left(\sum_{l=1}^{n} \sum_{s=1}^{T} K\left(H^{-1}(x_{ls} - x_{it})\right) \times m^*(x_{ls,it}) \right) K(H^{-1}(x_{ir} - x_{it})) \right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)$$

$$= O((n^2 T |\bar{H}|^{1/2})^{-1} |H|)$$

(I)-(B)-(ii)

1) $(l = i, i \neq j, r = s = t)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \bar{u}_j \left(\sum_{t=1}^{T} \frac{1}{\left|H\right|^2 f_i f(x_{it})} u_{it} (K(0))^2 \right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)$$

$$= 0 \quad \because \; E[u_{it} \mid x_{it}] = 0$$

2) $(l = i, i \neq j, s = t, r \neq t)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{T\left|H\right|^2 f_i f(x_{it})}u_{it}K\left(0\right)K(H^{-1}(x_{ir}-x_{it}))\right)$$

$$\times K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= 0 \quad \because \; E[u_{it} \mid x_{it}] = 0$$

3) $(l = i, i \neq j, r = s, r \neq t)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{T\left|H\right|^2 f_i f(x_{it})}u_{ir}K^2\left(H^{-1}(x_{ir}-x_{it})\right)\right)$$

$$\times K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= 0 \quad \because \; E[u_{ir} \mid x_{it}] = 0$$

4) $(l = i, i \neq j, r = t, s \neq r)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\frac{1}{T\left|H\right|^2 f_i f(x_{it})}\left(\sum_{s=1}^{T}K\left(H^{-1}(x_{is}-x_{it})\right)u_{is}\right)\right.$$

$$\left.\times K(0)\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= O((nT|\bar{H}|^{1/2}|H|^{3/2})^{-1})$$

5) $(l=i, i\neq j, r\neq s\neq t)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{T^2\left|H\right|^2 f_i f(x_{it})}\left(\sum_{s=1}^{T}K\left(H^{-1}(x_{is}-x_{it})\right)\right.\right.$$

$$\left.\left.\times\, u_{is}\right)K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= O((nT^2|\bar{H}|^{1/2}|H|)^{-1})$$

6) $(l=j, l\neq i, r=s=t)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\frac{1}{\left|H\right|^2 f_i f(x_{it})}K\left(H^{-1}(x_{jt}-x_{it})\right)u_{jt}K(0)\right)$$

$$\times\, K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= O((n|\bar{H}|^{1/2}|H|^{3/2})^{-1})$$

7) $(l=j, l\neq i, r=s, r\neq t)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{T\left|H\right|^2 f_i f(x_{it})}K\left(H^{-1}(x_{jr}-x_{it})\right)u_{jr}\right.$$

$$\left.\times\, K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= O((nT|\bar{H}|^{1/2}|H|)^{-1})$$

8) $(l = j, l \neq i, s = t, r \neq s)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{T\left|H\right|^2 f_i f(x_{it})}K\left(H^{-1}(x_{jt}-x_{it})\right)u_{jt}\right.$$

$$\left.\times K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= O((nT|\bar{H}|^{1/2}|H|)^{-1})$$

9) $(l = j, l \neq i, r = t, s \neq r)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\frac{1}{T\left|H\right|^2 f_i f(x_{it})}\left(\sum_{s=1}^{T}K\left(H^{-1}(x_{js}-x_{it})\right)u_{js}\right)\right.$$

$$\left.\times K(0)\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= O((nT|\bar{H}|^{1/2}|H|^{3/2})^{-1})$$

10) $(l = j, l \neq i, r \neq s \neq t)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{T^2\left|H\right|^2 f_i f(x_{it})}\left(\sum_{s=1}^{T}K\left(H^{-1}(x_{js}-x_{it})\right)u_{js}\right)\right.$$

$$\left.\times K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= O((nT^2|\bar{H}|^{1/2}|H|)^{-1})$$

11) $(i \neq j \neq l, r = s = t)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \bar{u}_j \left(\sum_{t=1}^{T} \frac{1}{n\left|H\right|^2 f_i f(x_{it})} \left(\sum_{l=1}^{n} K\left(H^{-1}(x_{lt} - x_{it})\right) u_{lt} \right.\right.$$

$$\left.\left. \times K(0) \right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)$$

$$= 0 \quad \because \; E[u_{lt} \mid x_{it}] = 0$$

12) $(i \neq j \neq l, r = s, r \neq t)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \bar{u}_j \left(\sum_{t=1}^{T} \sum_{r=1}^{T} \frac{1}{nT\left|H\right|^2 f_i f(x_{it})} \left(\sum_{l=1}^{n} K\left(H^{-1}(x_{lr} - x_{it})\right) u_{lr} \right.\right.$$

$$\left.\left. \times K(H^{-1}(x_{ir} - x_{it})) \right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)$$

$$= O((n^2 T|\bar{H}|^{1/2}|H|)^{-1})$$

13) $(i \neq j \neq l, s = t, r \neq s)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \bar{u}_j \left(\sum_{t=1}^{T} \sum_{r=1}^{T} \frac{1}{nT\left|H\right|^2 f_i f(x_{it})} \left(\sum_{l=1}^{n} K\left(H^{-1}(x_{lt} - x_{it})\right) u_{lt} \right.\right.$$

$$\left.\left. \times K(H^{-1}(x_{ir} - x_{it})) \right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)$$

$$= O((n^2 T|\bar{H}|^{1/2}|H|)^{-1})$$

14) $(i \neq j \neq l, r = t, s \neq r)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\frac{1}{nT\left|H\right|^2 f_i f(x_{it})}\left(\sum_{l=1}^{n}\sum_{s=1}^{T}K\left(H^{-1}(x_{ls}-x_{it})\right)\right.\right.$$

$$\left.\left.\times\, u_{ls}\right)K(0)\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= 0 \quad \because\ E[u_{lt} \mid x_{it}] = 0$$

15) $(i \neq j \neq l, r \neq s \neq t)$

$$S_{2n} = -\frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j\left(\sum_{t=1}^{T}\sum_{r=1}^{T}\frac{1}{nT^2\left|H\right|^2 f_i f(x_{it})}\left(\sum_{l=1}^{n}\sum_{s=1}^{T}K\left(H^{-1}(x_{ls}-x_{it})\right)\right.\right.$$

$$\left.\left.\times\, u_{ls}\right)K(H^{-1}(x_{ir}-x_{it}))\right)K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right)$$

$$= 0 \quad \because\ E[u_{lt} \mid x_{it}] = 0$$

(I)-(B)-(iii)

Note that $\frac{1}{\hat{f}(x_{it})} - \frac{1}{f(x_{it})} = O_p\left(\left(\frac{\ln nT}{nT|H|}\right)^{1/2}\right) + O_p(|H|)$.

$$
\begin{aligned}
S_{3n} &= -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \sum_{t=1}^{T} \bar{u}_j \left(\hat{\mu}_{n,i} - \mu_{n,i}\right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right) \\
&= -\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{l=1}^{n} \bar{u}_j \left(\frac{1}{\hat{f}(x_{it})} - \frac{1}{f(x_{it})}\right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right) \\
&\times \left(\frac{1}{2} m^*(x_{ls,it}) + u_{ls}\right) K\left(H^{-1}(x_{ls} - x_{it})\right) \\
&\leq \left(O_p\left(\left(\frac{\ln nT}{nT\left|H\right|}\right)^{1/2}\right) + O_p(|H|)\right) \left[-\frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{l=1}^{n} \bar{u}_j \right. \\
&\left. \times \left(\frac{1}{2} m^*(x_{ls,it}) + u_{ls}\right) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right) K\left(H^{-1}(x_{ls} - x_{it})\right) \right] \\
&= O((nT|\bar{H}|^{-1/2}|H|^{3/2})^{-1}(\ln nT)^{1/2}) + O(nT|\bar{H}|)^{-1})
\end{aligned}
$$

(I)-(C)

By letting $S_n(z_{ls}) = \left(\sum_{l=1}^{n} \sum_{s=1}^{T} z_{ls} K_{it,ls} z_{ls}\right)^{-1}$ and $S(z_{ls}) = \begin{pmatrix} f(x_{ls}) & 0 \\ 0 & f(x_{ls})\sigma_k^2 \end{pmatrix}$,

$$
\begin{aligned}
&\left|\hat{\mu}_{n,i} - \mu_{n,i}\right| \\
=& \frac{1}{nT^2\left|H\right|^2} \left| e_1' S_n(z_{ls})^{-1} \sum_{l=1}^{n} \sum_{s=1}^{T} \sum_{r=1}^{T} z_{ls} K\left(H^{-1}(x_{ls} - x_{it})\right) \left(\frac{1}{2} m^*(x_{ls,it}) + u_{ls}\right) \right. \\
& \times K(H^{-1}(x_{ir} - x_{it})) - \frac{1}{f_i f(x_i)} \sum_{l=1}^{n} \sum_{s=1}^{T} \sum_{r=1}^{T} z_{ls} K\left(H^{-1}(x_{ls} - x_{it})\right) \left(\frac{1}{2} m^*(x_{ls,it}) + u_{ls}\right) \\
& \left. \times K(H^{-1}(x_{ir} - x_{it})) \right|
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{1}{nT^2 |H|^2}\left| e_1'(S_n(z_{ls})^{-1} - S(z_{ls})^{-1}) \sum_{l=1}^{n}\sum_{s=1}^{T}\sum_{r=1}^{T} z_{ls} K\left(H^{-1}(x_{ls}-x_{it})\right)\left(\frac{1}{2}m^*(x_{ls,it}) + u_{ls}\right)\right.\\
&\left.\times K(H^{-1}(x_{ir}-x_{it}))\right|\\
&\le \frac{1}{|H|^2}((1,0)(S_n(z_{ls})^{-1} - S(z_{ls})^{-1})^2(1,0)')^{1/2}\\
&\times \frac{1}{nT^2}\left(\left|\sum_{l=1}^{n}\sum_{s=1}^{T}\sum_{r=1}^{T} z_{ls} K\left(H^{-1}(x_{ls}-x_{it})\right)\left(\frac{1}{2}m^*(x_{ls,it}) + u_{ls}\right) K(H^{-1}(x_{ir}-x_{it}))\right|\right.\\
&+\left|\sum_{l=1}^{n}\sum_{s=1}^{T}\sum_{r=1}^{T}\left(H^{-1}(x_{ls}-x_{it})\right)' K\left(H^{-1}(x_{ls}-x_{it})\right)\left(\frac{1}{2}m^*(x_{ls,it}) + u_{ls}\right)\right.\\
&\left.\left.(H^{-1}(x_{ir}-x_{it}))'K(H^{-1}(x_{ir}-x_{it}))\right|\right)
\end{aligned}
$$

I follow Lemma 2 of Martins-Filho and Yao (2007) to obtain

$$
\begin{aligned}
|\hat{\mu}_{n,i} - \mu_{n,i}| &= O\left(|H|^2\left(\frac{\ln nT}{nT|H|}\right)^{\frac{1}{2}}\right) + O(|H|^4)\\
\sup\left|\bar{\hat{m}}_i - m_i\right| &= O\left(\left(\frac{\ln nT}{nT|H|}\right)^{\frac{1}{2}}\right) + O(|H|^2)
\end{aligned}
$$

Using $\hat{m}(x_i) = m^{(1)}(x_{ij})(x_i - x_j)$, where $x_{ij} = \lambda x_i + (1-\lambda)x_j$,

$$
\begin{aligned}
I_{13n} &= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{\mu}_{n,i}'\hat{\mu}_{n,j} K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)\\
&= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\mu_{n,i} + \hat{\mu}_{n,i} - \mu_{n,i})'\\
&\times (\mu_{n,j} + \hat{\mu}_{n,j} - \mu_{n,j}) K\left(\bar{H}^{-1}(\bar{X}_j - \bar{X}_i)\right)
\end{aligned}
$$

$$
\begin{aligned}
\sup|\mu_{n,i}+\hat{\mu}_{n,i}-\mu_{n,i})| &\leq \sup|\mu_{n,i}|+\sup|\hat{\mu}_{n,i}-\mu_{n,i}| \\
&= O_p\left(\left(\frac{\ln nT}{nT\,|H|}\right)^{1/2}\right)+O_p(|H|^2)
\end{aligned}
$$

$$
\begin{aligned}
\sup\left|(\mu_{n,i}+\hat{\mu}_{n,i}-\mu_{n,i}))'(\mu_{n,j}+\hat{\mu}_{n,j}-\mu_{n,j})\right| &\leq (\sup|\mu_{n,i}+\hat{\mu}_{n,i}-\mu_{n,i}|)^2 \\
&= O_p\left(\frac{\ln nT}{nT\,|H|}\right)+O_p(|H|^4)
\end{aligned}
$$

$$
\begin{aligned}
I_{13n} &\leq (\sup|\mu_{n,i}+\hat{\mu}_{n,i}-\mu_{n,i}|)^2\frac{1}{n(n-1)\,|H_v|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}K\left(\bar{H}^{-1}(\bar{X}_j-\bar{X}_i)\right) \\
&= O_p\left(n^{-2}\left|\bar{H}\right|\left(\frac{\ln nT}{nT\,|H|}\right)\right)+O_p\left(n^{-2}\left|\bar{H}\right|^{-1}|H|^4\right)
\end{aligned}
$$

(IV)

$$
\begin{aligned}
\hat{\Omega} &= \frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{\alpha}_i^2\hat{\alpha}_j^2K^2\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right) \\
&= \frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_i^2\bar{u}_j^2K^2\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right)+o_p(1) \\
&= \frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i<j}^{n}\sum^{n}\left[\bar{u}_i^2\bar{u}_j^2K^2\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right)+\bar{u}_j^2\bar{u}_i^2K^2\left(\bar{H}^{-1}\left(\bar{X}_i-\bar{X}_j\right)\right)\right]
\end{aligned}
$$

By using the properties of U-statistics, $\Omega = 2\int K^2(\psi)d\psi\mathrm{E}[\sigma^4(\bar{x})f(\bar{x})]$.

C.2 Proof of Theorem 9

Under the alternative, $m_1(X_{it}) = m(X_{it}) + \delta_n l(\bar{X}_i)$. Then, $u_{it} = \eta_{it} + \delta_n l(\bar{X}_i)$, where $\eta_{it} = y_{it} - m(X_{it})$ and $\delta_n = n^{-1/2}\left(\left|\bar{H}\right|\right)^{-1/4}$.

$$\begin{aligned}
I_n &= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\hat{\alpha}_i\hat{\alpha}_j K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right) \\
&= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_i\bar{u}_j K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right) \\
&- \frac{2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_j(\bar{\hat{m}}_i-\bar{m}_i) K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right) \\
&+ \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\bar{\hat{m}}_i-\bar{m}_i)(\bar{\hat{m}}_j-\bar{m}_j) K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right) \\
&= I_{1nG}+I_{2nG}+I_{3nG}
\end{aligned}$$

For the following sections, I will show

$$n\left|\bar{H}\right|^{1/2} I_{1nG} \xrightarrow{d} N(0,\Omega)$$

$$\begin{aligned}
I_{1nG} &= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{u}_i\bar{u}_j K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right) \\
&= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}(\bar{\eta}_i+\delta_n l(\bar{X}_i))(\bar{\eta}_j+\delta_n l(\bar{X}_j)) K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right) \\
&= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{\eta}_i\bar{\eta}_j K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right) \\
&+ \frac{2\delta_n}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{\eta}_i l(\bar{X}_j) K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right)
\end{aligned}$$

$$+\frac{\delta_n^2}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n} l(\bar{X}_i)l(\bar{X}_j)K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right)$$

$$= Q_{1n} + 2\delta_n Q_{2n} + \delta_n^2 Q_{3n}$$

(I)-(A)

$$\begin{aligned}
Q_{1n} &= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\bar{\eta}_i\bar{\eta}_j K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right) \\
&= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n}\left[\bar{\eta}_i\bar{\eta}_j K\left(\bar{H}^{-1}\left(\bar{X}_j-\bar{X}_i\right)\right)+\bar{\eta}_j\bar{\eta}_i K\left(\bar{H}^{-1}\left(\bar{X}_i-\bar{X}_j\right)\right)\right]
\end{aligned}$$

By letting $W_i = (\bar{\eta}_i, \bar{X}_i)$,

$$\begin{aligned}
Q_{1n} &= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n}\left[\psi_n(W_i,W_j)+\psi_n(W_j,W_i)\right] \\
&= \frac{1}{n(n-1)\left|\bar{H}\right|}\sum_{\substack{i=1 \\ i<j}}^{n}\sum_{j=1}^{n}\phi_n(W_i,W_j)
\end{aligned}$$

As $E[\phi_n^2(W_i,W_j)] = E[\psi_n^2(W_i,W_j)] + E[\psi_n^2(W_j,W_i)] + 2E[\psi_n(W_i,W_j)\psi_n(W_j,W_i)]$,

$$\frac{1}{\left|\bar{H}\right|}E[\psi_n^2(W_i,W_j)] \to \left(\int K^2\left(\varphi\right)d\varphi\right)E\left[\sigma^4(\bar{X}_i)f(\bar{X}_i)\right] < \infty$$

$$\frac{2}{\left|\bar{H}\right|}E[\psi_n(W_i,W_j)\psi_n(W_j,W_i)] \to 2\left(\int K^2\left(\varphi\right)d\varphi\right)E\left[\sigma^4(\bar{X}_i)f(\bar{X}_i)\right]$$

$$\therefore n\left|\bar{H}\right|^{1/2}Q_{1n} \xrightarrow{d} N(0,\Omega),\ \text{ where } \Omega = 2\left(\int K^2\left(\varphi\right)d\varphi\right)E\left[\sigma^4(\bar{X}_i)f(\bar{X}_i)\right]$$

(I)-(B)

$$
\begin{aligned}
Q_{2n} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \frac{1}{|\bar{H}|} \bar{\eta}_i l(\bar{X}_j)) K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right) \\
&= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left[\frac{1}{|\bar{H}|} \bar{\eta}_i l(\bar{X}_j) K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right)\right. \\
&+ \left.\frac{1}{|\bar{H}|} \bar{\eta}_j l(\bar{X}_i)) K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right)\right] \\
&= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \left[\psi_n(W_i, W_j) + \psi_n(W_j, W_i)\right], \text{ where } W_i = (\bar{\eta}_i, \bar{X}_i).
\end{aligned}
$$

$E[\psi_n(W_i, W_j)] = 0$. By applying Lipschitz condition,

$$
\begin{aligned}
E[\psi_n^2(W_i, W_j)] &= E\left[\frac{1}{|H_v|^2} \bar{\eta}_i^2 l^2(\bar{X}_j) K^2\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right)\right] \\
&= \frac{1}{|\bar{H}|^2} \int \sigma^2(\bar{X}_i) l^2(\bar{X}_j) K^2\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right) f(\bar{X}_i) f(\bar{X}_j) d\bar{X}_i d\bar{X}_j \\
&= \frac{1}{|\bar{H}|} \int \sigma^2(\bar{X}_i) l^2(\bar{X}_i + \bar{H}\varphi)) K^2(\varphi) f(\bar{X}_i) f(\bar{X}_i + \bar{H}\varphi) d\bar{X}_i d\varphi \\
&= \frac{1}{|\bar{H}|} \left(\int K^2(\varphi) d\varphi\right) \left(\int \sigma^2(\bar{X}_i) l^2(\bar{X}_i) f(\bar{X}_i)^2 d\bar{X}_i\right)
\end{aligned}
$$

$$
Q_{2n} = n^{-1}\left(E[\phi_n^2(W_i, W_j)]\right)^{\frac{1}{2}} = O(n^{-1}|\bar{H}|^{-1/2})
$$

$$
\therefore\ n|\bar{H}|^{1/2} \delta_n Q_{2n} = n^{-1/2}|\bar{H}|^{-1/4} \xrightarrow{p} 0
$$

(I)-(C)

$$Q_{3n} = \frac{1}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} l(\bar{X}_i) l(\bar{X}_j) K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right)$$

$$\begin{aligned}
&\frac{1}{\left|\bar{H}\right|} E\left[l(\bar{X}_i) l(\bar{X}_j) K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right)\right] \\
=&\frac{1}{\left|\bar{H}\right|} \int l(\bar{X}_i) l(\bar{X}_i + \bar{H}\varphi) K\left(\varphi\right) f(\bar{X}_i) f(\bar{X}_i + \bar{H}\varphi) \left|\bar{H}\right| d\bar{X}_i d\varphi \\
=&\int l^2(\bar{X}_i) K\left(\varphi\right) f(\bar{X}_i)^2 d\bar{X}_i d\varphi \\
=&\left(\int K\left(\varphi\right) d\varphi\right) \left(\int l^2(\bar{X}_i) f(\bar{X}_i)^2 d\bar{X}_i\right) \\
=&E[l^2(\bar{X}_i) f(\bar{X}_i)] \\
=&O_p(1)
\end{aligned}$$

$$n\left|\bar{H}\right|^{1/2} \delta_n^2 Q_{3n} = n\left|\bar{H}\right|^{1/2} (n^{-1}\left|\bar{H}\right|^{-1/2}) Q_{3n} = Q_{3n} \xrightarrow{p} E[l^2(\bar{X}_i) f(\bar{X}_i)]$$

$$n\left|\bar{H}\right|^{1/2} I_{11nG} \xrightarrow{d} N(E[l^2(\bar{X}_i) f(\bar{X}_i)], \Omega)$$

C.3 Proof of Theorem 10

Under $\mathbb{H}_1$,

$$
\begin{aligned}
\bar{\hat{u}}_i &= \bar{y}_i - \bar{\hat{m}}_i(X_{it}) \\
&= \bar{y}_i - \bar{m}_1(X_{it}) + \bar{m}_1(X_{it}) - \bar{\hat{m}}_i(X_{it}) \\
&= \bar{\eta}_i + \bar{m}_1(X_{it}) - \bar{\hat{m}}(X_{it}) \\
&= \bar{\eta}_i + (\bar{m}_1(X_{it}) - \bar{m}(X_{it})) - (\bar{\hat{m}}(X_{it}) - \bar{m}(X_{it})) \\
&= \bar{u}_i - (\bar{\hat{m}}(X_{it}) - \bar{m}(X_{it}))
\end{aligned}
$$

The test statistic is then written as follows:

$$
\begin{aligned}
I_n &= \frac{1}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{\alpha}_i \hat{\alpha}_j K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right) \\
&= \frac{1}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \bar{u}_i \bar{u}_j K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right) + o((n\left|\bar{H}\right|^{1/2})^{-1}) \\
&= \frac{1}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} (\bar{\eta}_i + h(\bar{X}_i))(\bar{\eta}_j + h(\bar{X}_j)) K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right)
\end{aligned}
$$

Define $\phi_n(W_i, W_j) = \frac{1}{\left|\bar{H}\right|} h(\bar{X}_i) h(\bar{X}_j) K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right)$.

$$
\begin{aligned}
E[\phi_n(W_i, W_j)] &= E[E[\phi_n(W_i, W_j) \mid W_i, W_j]] \\
&= E\left[\frac{1}{\left|\bar{H}\right|} K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right) h(\bar{X}_i) h(\bar{X}_j)\right] \\
&= \frac{1}{\left|\bar{H}\right|} \int K\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right) h(\bar{X}_i) h(\bar{X}_j) f(\bar{X}_i) f(\bar{X}_j) d\bar{X}_i d\bar{X}_j \\
&= \int K\left(\varphi\right) h(\bar{X}_i) h(\bar{X}_i + \bar{H}\varphi) f(\bar{X}_i) f(\bar{X}_i + \bar{H}\varphi) d\bar{X}_i d\varphi
\end{aligned}
$$

$$= \int K\left(\varphi\right) d\varphi \int (h(\bar{X}_i))^2 f(\bar{X}_i)^2 d\bar{X}_i$$

$$= E[(h(\bar{X}_i))^2 f(\bar{X}_i)]$$

$$\therefore\ \hat{\Omega} = \frac{2}{n(n-1)\left|\bar{H}\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \hat{\alpha}_i^2 \hat{\alpha}_j^2 K^2\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right)$$

$$E\left[\phi_n(W_i, W_j)\right] = E\left[E\left[\phi_n(W_i, W_j) \mid \bar{X}_i\right]\right]$$

$$= E\left[\frac{1}{\left|\bar{H}\right|} K^2\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right)\left(\sigma^2(\bar{X}_i) + (h(\bar{X}_i))^2\right)\left(\sigma^2(\bar{X}_j) + (h(\bar{X}_j))^2\right)\right]$$

$$= \frac{1}{\left|\bar{H}\right|} \int K^2\left(\bar{H}^{-1}\left(\bar{X}_j - \bar{X}_i\right)\right)\left(\sigma^2(\bar{X}_i) + (h(\bar{X}_i))^2\right)\left(\sigma^2(\bar{X}_j) + (h(\bar{X}_j))^2\right)$$

$$\times f(\bar{X}_i) f(\bar{X}_j) d\bar{X}_i d\bar{X}_j$$

$$= \int K^2\left(\psi\right)\left(\sigma^2(\bar{X}_i) + (h(\bar{X}_i))^2\right)\left(\sigma^2(\bar{X}_i + \bar{H}\psi) + (h(\bar{X}_i + \bar{H}\psi))^2\right)$$

$$\times f(\bar{X}_i) f(\bar{X}_i + \bar{H}\psi) d\bar{X}_i d\psi$$

$$= \left(\int K^2(\psi) d\psi\right)\left(\int \left(\sigma^2(\bar{X}_i) + (h(\bar{X}_i))^2\right)^2 f(\bar{X}_i)^2 d\bar{X}_i\right)$$

$$= \left(\int K^2(\psi) d\psi\right)\left[E\left[\sigma^4(\bar{X}_i) f(\bar{X}_i)\right] + 2E\left[\sigma^2(\bar{X}_i)(h(\bar{X}_i))^2 f(\bar{X}_i)\right] + E\left[(h(\bar{X}_i))^2 f(\bar{X}_i)\right]\right]$$

$$= B_1$$

$$\hat{\Omega} \xrightarrow{p} 2B_1$$

$$J_n = \frac{n\left|\bar{H}\right|^{1/2} I_n}{\sqrt{\hat{\Omega}}} > c_n = o_p(n\left|\bar{H}\right|^{1/2})$$