A Consistent Nonparametric Test for Endogeneity*

Seolah Kim†

October 20, 2019

Abstract

I construct 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. Also, a wild bootstrap test is proposed to improve the finite sample performance. Using the data from Autor, Dorn, and Hanson (2013), I test if the Chinese import exposure is endogenous with the US local employment. I find a contradicting result between the Hausman test and the nonparametric tests.

Keywords: Nonparametric test, nonparametric estimation, wild bootstrap.

JEL: C12, C14, C15.

*I deeply appreciate Aman Ullah for his insightful advice and help. I also thank Michael Bates, Jianghao Chu, Sungjun Huh, Tae-Hwy Lee, Yoonseok Lee, Ruoyao Shi, and seminar participants at UC Riverside, Midwest Econometric Conference, Joint Statistical Meetings for helpful discussions, comments, and suggestions. All errors are my own.

†Department of Economics, 3129 Sproul Hall, University of California, Riverside, CA 92507, USA. Email: skim238@ucr.edu

I 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 Killian 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 (2001), Su et al. (2013), Lee et al. (2013), 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 (2001) 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), Killian 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 II introduces a model, hypotheses, and the test statistic for endogeneity. In Section III, 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 IV, I then apply the test to the empirical data in Section V. I will test the endogeneity of Chinese import shock to the US with the US local unemployment share using Autor et al. (2013). Section VI concludes the paper.

II 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.

II.A 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

$$\begin{cases} y_i = m\left(x_i\right) + u_i \\ x_i = g\left(z_i\right) + v_i \end{cases} \tag{1}$$

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 $\mathrm{E}[u_i \mid x_i, z_i] = 0$ and $\mathrm{E}[v_i \mid z_i] = 0$ are satisfied. From equation (1), the moment condition of interest in the structural equation is $\mathrm{E}[u_i \mid x_i,\, z_i] = 0$. 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):

$$\begin{aligned} y_i &= m(x_i) + u_i \\ &= m(x_i) + E[u_i \mid v_i] + u_i - E[u_i \mid v_i] \qquad (2) \\ &= m(x_i) + h(v_i) + \epsilon_i, \text{ where } \epsilon_i = u_i - E[u_i \mid v_i] \\ &= m_1(x_i, v_i) + \epsilon_i \end{aligned}$$

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 $\mathrm{E}[u_i \mid x_i,\, z_i]$ can be also expressed from the equation (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:

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

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 literatures 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 1. 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 1 in constructing a test for endogeneity.

Theorem 1 *$E[u_i|v_i] = 0$ 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 1 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.

For any $f\left(v_i\right) > 0$, I have

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

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

$f(v) > 0$. ■

II.B Test Statistic and Asymptotic Properties

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

$$
\begin{aligned}
I_n &= \widehat{E}[\widehat{f}\left(\widehat{v}_i\right) \widehat{u}_i \widehat{E}[\widehat{u}_i | \widehat{v}_i]] \\
&= \frac{1}{n} \sum_{i=1}^{n} \widehat{u}_i \widehat{f}\left(\widehat{v}_i\right) \left\{ \frac{1}{(n-1)\left|H_v\right| \widehat{f}\left(\widehat{v}_i\right)} \sum_{j \neq i}^{n} \widehat{u}_j K\left(H_v^{-1}\left(\widehat{v}_j - \widehat{v}_i\right)\right) \right\} \\
&= \frac{1}{n} \sum_{i=1}^{n} \widehat{u}_i \left\{ \frac{1}{(n-1)\left|H_v\right|} \sum_{j \neq i}^{n} \widehat{u}_j K\left(H_v^{-1}\left(\widehat{v}_j - \widehat{v}_i\right)\right) \right\} \\
&= \frac{1}{n(n-1)\left|H_v\right|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} \widehat{u}_i \widehat{u}_j K\left(H_v^{-1}\left(\widehat{v}_j - \widehat{v}_i\right)\right),
\end{aligned}
\tag{3}
$$

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 with Li-Wang type test is that I use generated regressors inside the kernel function[1]. Note that

$$
\begin{aligned}
\widehat{u}_i &= y_i - \widehat{m}\left(x_i\right), \\
\widehat{v}_i &= x_i - \widehat{g}\left(z_i\right)
\end{aligned}
$$

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).

[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.

(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.
(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 $|K^j(u) - K^j(v)| \leq C_1 |u - v|$, 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 Taylor expansion and especially 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 the assumptions, I standardize the estimator,

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

Theorem 2 *Under H_0, as $\hat{\Omega}$ is a consistent estimator of $\Omega = 2[\int K^2(\psi)d\psi]E[\sigma^4(v)f(v)]$,*

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

This 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.

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

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

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

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

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

[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.

hypotheses.

Theorem 4 *Assuming (A1)-(A6) and under 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 4 suggests the consistency of the test statistic. Under H_1, the probability of rejecting the null will converge to 1.

II.C 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.

Step 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.

Step 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$[u_i^*] = 0$, $E\left[u_i^{*2}\right] = \hat{u}_i^2$, and E$[u_i^{*3}] = \hat{u}_i^3$.

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

Step 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, $\hat{v}_i^* = \hat{v}_i$.

Step 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.

Step 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 5.

Theorem 5 *Let* $J_n^* = n\left|H_v\right|^{1/2} I_n^* / \sqrt{\hat{\Omega}^*}$, *where* $\hat{\Omega}^* = \frac{2}{n(n-1)|H_v|}\sum_{i=1}^n \sum_{j\neq i}^n \hat{u}_i^{*2}\hat{u}_j^{*2} K\left(H_v^{-1}(\hat{v}_j^* - \hat{v}_i^*)\right)$. *Under* 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 5 will follow similarly to those of Theorem 2. Under the $H_1(\delta_n)$, $P(J_n > c^*) \to 1$ asymptotically, where c^* denotes the bootstrap critical valude based on the boostrap samples. This shows the consistency of the bootstrap test statistic.

III Simulations

III.A 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.

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

$$\text{where } i = 1, ..., n \text{ ,}$$

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$_2$.

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

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].

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

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

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 $(\hat{\alpha}, \hat{\beta}) = \underset{\alpha,\beta}{\arg\max} \sum_{t=1}^{n} (y_t - \alpha - \beta(x_i - x_t))^2$. 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 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.

III.B 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.

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

Table 1-4 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 5 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.

Table 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 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 3. 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. 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

[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.

[5]This underestimation of the size has been also noted in Li and Wang (1998) and Hsiao and Li (2002).

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 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 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.

Table 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 2: Power of Each Test of DGP_1

			$\theta = 0.2$			$\theta = 0.5$			$\theta = 0.8$		
		c	1%	5%	10%	1%	5%	10%	1%	5%	10%
$n = 100$	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
$n = 400$	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: Power of Each Test of DGP_2

			$\theta = 0.2$			$\theta = 0.5$			$\theta = 0.8$		
		c	1%	5%	10%	1%	5%	10%	1%	5%	10%
$n = 100$	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
$n = 400$	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 4: Power of Each Test of DGP_3

			$\theta = 0.2$			$\theta = 0.5$			$\theta = 0.8$		
		c	1%	5%	10%	1%	5%	10%	1%	5%	10%
$n = 100$	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
$n = 400$	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.

IV Extension: The Case with Other Exogenous Variables

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

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

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 $E[u_i \mid z_i, x_{i2}] = 0$ and $E[v_i \mid z_i, x_{i2}] = 0$ 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) $E\left[u\right| z, x_2] = 0.\sigma^2(v) = E[u^2 \mid v]$, $\sigma^2(v)$ 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.

(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 $\left|K^j(u) - K^j(v)\right| \leq C_1 \left|u - v\right|$, for $j = 0, 1, 2, 3$.

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

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

The test statistic will be written as follows:

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

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.

Note that the moment condition of interenst 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 standardized test statistic is

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

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.

V 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.

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

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 1 and Figure 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 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 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 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, 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

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

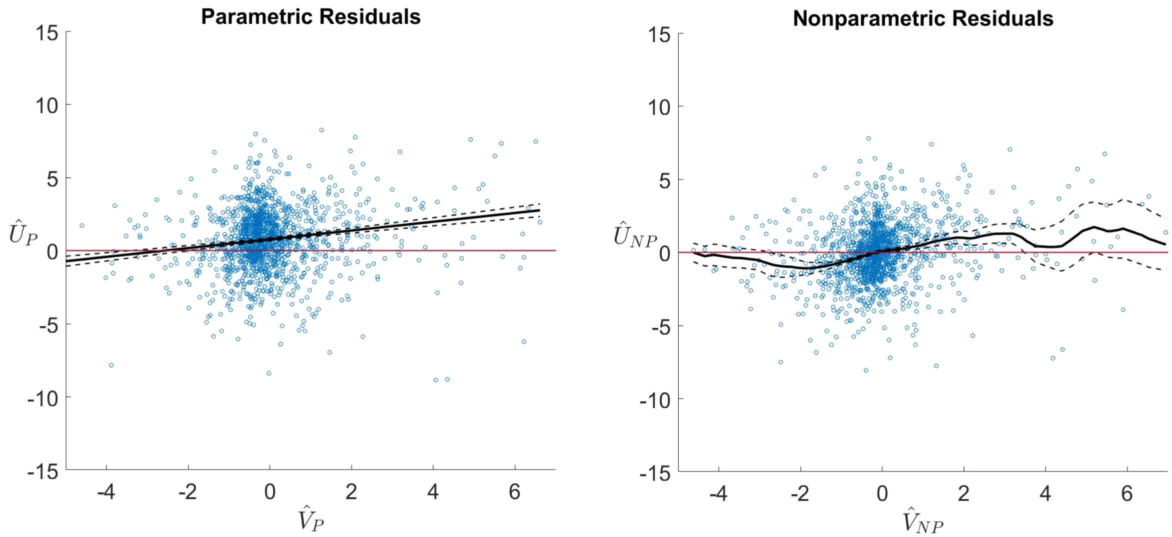
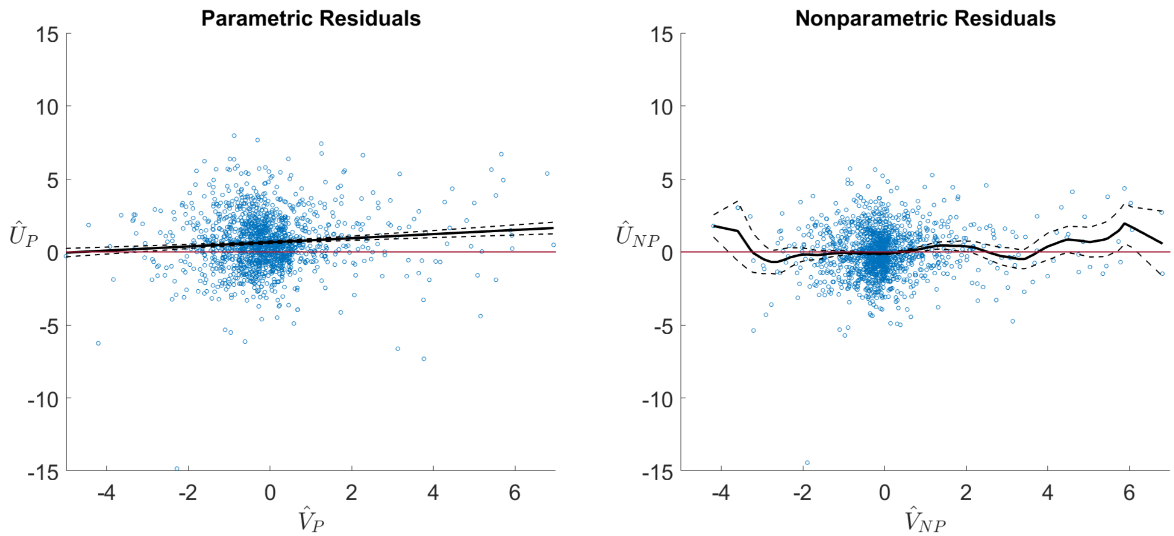


Figure 2: Correlation between $\hat{u}_i$ and $\hat{v}_i$ in Model 2



level.

Table 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.

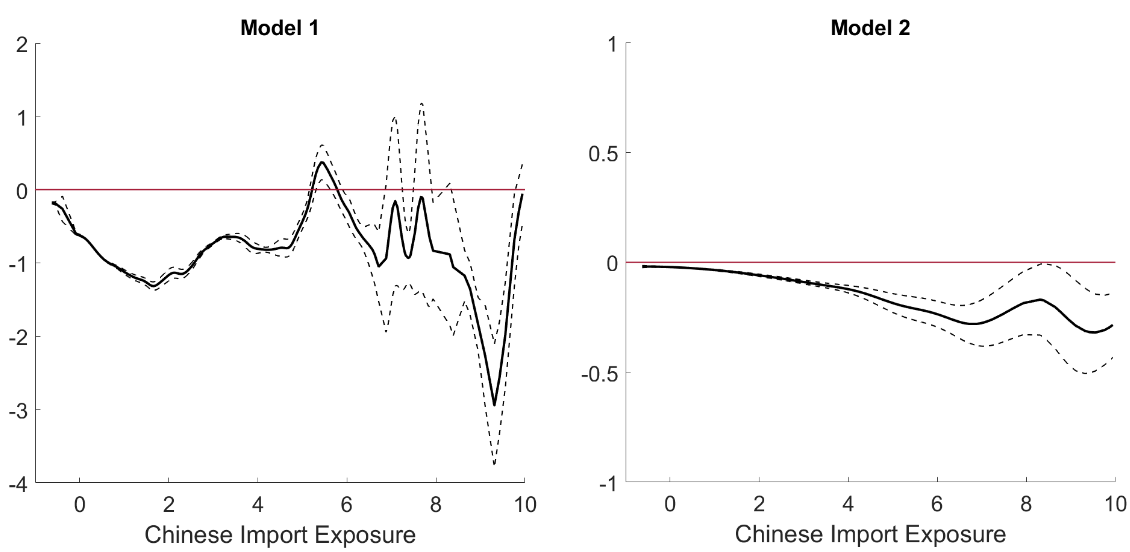
There can be two possible explanations why I have such a contradicting test results for endogeneity. Considering that Model Specification 2 is estimated by adding other exogenous variables, some factors which cause endogeneity of Chinese import variable might have been filtered out by those variables. 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. 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.

Figure 3: Marginal Effect of Chinese Import Exposure



VI 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 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.

References

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

Autor, H., D. Dorn, and G. H. Hanson. American economic review, forthcoming the china syndrome: Local labor market effects of import competition in the united states.

Blundell, R. and J. L. Horowitz (2007). A non-parametric test of exogeneity. *The Review of Economic Studies 74*(4), 1035–1058.

Blundell, R. and J. L. Powell (2003). Endogeneity in nonparametric and semiparametric regression models. *Econometric society monographs 36*, 312–357.

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

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

Coglianese, J., L. W. Davis, L. Kilian, and J. H. Stock (2017). Anticipation, tax avoidance, and the price elasticity of gasoline demand. *Journal of Applied Econometrics 32*(1), 1–15.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011). Nonparametric instrumental regression. *Econometrica 79*(5), 1541–1565.

Das, M., W. K. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies 70*(1), 33–58.

Davidson, R. and E. Flachaire (2008). The wild bootstrap, tamed at last. *Journal of Econometrics 146*(1), 162–169.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, pp. 569–593. Springer.

Hahn, J., Z. Liao, and G. Ridder (2018). Nonparametric two-step sieve m estimation and inference. *Econometric Theory*, 1–44.

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

Hardle, W., E. Mammen, et al. (1993). Comparing nonparametric versus parametric regression fits. *The Annals of Statistics 21*(4), 1926–1947.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251–1271.

Henderson, D. J., R. J. Carroll, and Q. Li (2008). Nonparametric estimation and testing of fixed effects panel data models. *Journal of Econometrics 144*(1), 257–275.

Henderson, D. J. and C. F. Parmeter (2012). Normal reference bandwidths for the general order, multivariate kernel density derivative estimator. *Statistics & Probability Letters 82*(12), 2198–2205.

Horowitz, J. L. (2011). Applied nonparametric instrumental variables estimation. *Econometrica 79*(2), 347–394.

Hsiao, C. and Q. Li (2001). A consistent test for conditional heteroskedasticity in time-series regression models. *Econometric Theory 17*(1), 188–221.

Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*(5), 1481–1512.

Kasy, M. (2011). Identification in triangular systems using control functions. *Econometric Theory 27*(3), 663–671.

Kocherlakota, N. R. and K.-M. Yi (1996). A simple time series test of endogenous vs. exogenous growth models: An application to the united states. *The Review of Economics and Statistics*, 126–134.

Lee, J. (1990). U-statistics: Theory and practice.

Lee, T.-H., Y. Tu, and A. Ullah (2015). Forecasting equity premium: Global historical average versus local historical average and constraints. *Journal of Business & Economic Statis-*

tics 33(3), 393–402.

Li, Q. and S. Wang (1998). A simple consistent bootstrap test for a parametric regression function. *Journal of Econometrics 87*(1), 145–165.

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

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

Martins-Filho, C. and F. Yao (2007). Nonparametric frontier estimation via local linear regression. *Journal of Econometrics 141*(1), 283–319.

Matzkin, R. L. (2008). Identification in nonparametric simultaneous equations models. *Econometrica 76*(5), 945–978.

Miguel, E., S. Satyanath, and E. Sergenti (2004). Economic shocks and civil conflict: An instrumental variables approach. *Journal of political Economy 112*(4), 725–753.

Murtazashvili, I. and J. M. Wooldridge (2016). A control function approach to estimating switching regression models with endogenous explanatory variables and endogenous switching. *Journal of econometrics 190*(2), 252–266.

Newey, W. K., J. L. Powell, and F. Vella (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica 67*(3), 565–603.

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

Su, L., I. Murtazashvili, and A. Ullah (2013). 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.

Su, L. and A. Ullah (2008). Local polynomial estimation of nonparametric simultaneous equations models. *Journal of Econometrics 144*(1), 193–218.

Sun, Y. and Q. Li (2006). An alternative series based consistent model specification test. *Economics Letters 93*(1), 37–44.

Wang, X. and Y. Hong (2018). Characteristic function based testing for conditional independence: a nonparametric regression approach. *Econometric Theory 34*(4), 815–849.

Wu, C.-F. J. et al. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics 14*(4), 1261–1295.

Wu, D.-M. (1973). Alternative tests of independence between stochastic regressors and disturbances. *Econometrica (pre-1986) 41*(4), 733.

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

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

Appendix

Proof of Theorem 2 I let

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

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

$$\begin{aligned} \widehat{\Delta}_{ij} - \Delta_{ij} &= \left(\widehat{g}\left(z_i\right) - g\left(z_i\right)\right) - \left(\widehat{g}\left(z_j\right) - g\left(z_j\right)\right), \\ \widehat{u}_i &= y_i - \widehat{m}\left(x_i\right) = u_i - \left(\widehat{m}\left(x_i\right) - m\left(x_i\right)\right). \end{aligned}$$

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

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

from which I have

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

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

(III) $I_{3n} = O\left(\frac{\ln n}{n^2\sqrt{\left|H_v\right|^{1/2}\left|H_z\right|}}\right) + O\left(\frac{\left|H_z\right|^4}{\sqrt{n^2\left|H_v\right|}}\right) + O\left(\frac{(\ln n)^{1/2}\left|H_z\right|^{3/2}}{n^{3/2}\left|H_v\right|^{1/2}}\right) = o_p((\sqrt{n^2\left|H_v\right|})^{-1})$ by (A5)

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

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

$$\begin{aligned} I_{1n} &= \frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\left(u_i - \mu_n(x_i)\right)\left(u_j - \mu_n(x_j)\right)K\left(H^{-1}\Delta_{ij}\right) \\ &= \frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}u_i u_j K\left(H_v^{-1}\Delta_{ij}\right) \\ &\quad + \frac{2}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}u_i\mu_n(x_j)K\left(H_v^{-1}\Delta_{ij}\right) \\ &\quad + \frac{1}{n(n-1)\left|H_v\right|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\mu_n(x_i)\mu_n(x_j)K\left(H_v^{-1}\Delta_{ij}\right) \\ &\equiv I_{11n} + I_{12n} + I_{13n}. \end{aligned}$$

(I)-(A) Following Lemma 1 from Yao and Ullah (2013), define second-order U-statistic $U_n = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i<j}}^{n} \phi_n(X_i, X_j)$, 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.

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

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

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

By (A2),

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

Then, I have $\left(\mathrm{E}[\phi_n^2(W_i,W_j)]\right)^2=O(|H_v|^2)$.

By Cr inequality, $\mathrm{E}[\phi_n^4(W_i,W_j)]\le C\left[\mathrm{E}[\psi_n^4(W_i,W_j)]+\mathrm{E}[\psi_n^4(W_j,W_i)]\right]$

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

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

Lastly, define

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

By Cr inequality,

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

Then,

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

$$
\begin{aligned}
&= |H_v|^2 \int \sigma^2(v_i)\sigma^2(v_i - H_v\psi_2) \left(\int K(\psi_1) K(\psi_1+\psi_2) \sigma^2(v_i + H_v\psi_1) f(v_i + H_v\psi_1) d\psi_1 \right)^2 \\
&\quad \times f(v_i) f(v_i - H_v\psi_2) |H_v| \, dv_i d\psi_2 \\
&= |H_v|^3 \int \sigma^4(v_i) \left(\int K(\psi_1) K(\psi_1+\psi_2) \sigma^2(v_i) f(v_i) d\psi_1 \right)^2 f^2(v_i) dv_i d\psi_2 \\
&= |H_v|^3 \left(\int \sigma^8(v_i) f^4(v_i) dv_i \right) \int \left(\int K(\psi_1) K(\psi_1+\psi_2) d\psi_1 \right)^2 d\psi_2 \\
&= O(|H_v|^3)
\end{aligned}
$$

I can conclude that

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

Also, I have $\frac{1}{|H|}\mathrm{E}[\phi_n^2(W_i, W_j)] \to \Omega$, 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 $z_t' = (1, \frac{x_t - x_i}{h_x})$,

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

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

$$\begin{aligned}
I_{12n} &= -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j (\hat{m}(x_i) - m(x_i)) K\left(H_v^{-1}(v_j - v_i)\right) \\
&= -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j \left[e_1' \left(\sum_{t=1}^{n} z_{tt} K\left(H_x^{-1}(x_t - x_i)\right) z_t \right)^{-1} \sum_{t=1}^{n} z_t K\left(H_x^{-1}(x_t - x_i)\right) \right. \\
&\quad \left. \times \left(\frac{1}{2} m^*(x_{it}) + u_t \right) \right] K\left(H_v^{-1}(v_j - v_i)\right) \\
&= -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j \; \hat{\mu}_n(x_i) K\left(H_v^{-1}(v_j - v_i)\right) \\
&= -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j \left(\mu_n(x_i) + \hat{\mu}_n(x_i) - \mu_n(x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right) \\
&= -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j \left(\frac{1}{nh_x f(x_i)} \sum_{t=1}^{n} K\left(H_x^{-1}(x_t - x_i)\right) \frac{1}{2} m^*(x_{it}) \right) K\left(H_v^{-1}(v_j - v_i)\right) \\
&\quad -\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j \left(\frac{1}{nh_x f(x_i)} \sum_{t=1}^{n} K\left(H_x^{-1}(x_t - x_i)\right) u_t \right) K\left(H_v^{-1}(v_j - v_i)\right) \\
&\quad +\frac{2}{n(n-1)\,|H_v|} \sum_{i=1}^{n} \sum_{j \neq i}^{n} u_j \left(\hat{\mu}_n(x_i) - \mu_n(x_i)\right) K\left(H_v^{-1}(v_j - v_i)\right) \\
&\equiv S_{1n} + S_{2n} + S_{3n}
\end{aligned}$$

(I)-(B)-(i)

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

1) $i = t$

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

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 $\mathbb{E}[u \mid x, v] = 0$,

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

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),

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

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,

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

Then,

$$\begin{aligned}
& \mathrm{E}[\psi^2(W_i, W_j)] \\
= \; & \mathrm{E}\left[\frac{1}{4f(x_i)^2 |H_v|^2} u_j^2 (m^*(x_{ij}))^2 K^2\left(H_x^{-1}(x_t - x_i), H_v^{-1}(v_j - v_i)\right)\right] \\
= \; & \mathrm{E}\left[\frac{|H_x|^4}{4f(x_i)^2 |H_v|^2} u_j^2 \left(H_x^{-1}(x_j - x_i)\right)^2 (m^{(2)}(x_{ij}))^2 \left(H_x^{-1}(x_j - x_i)\right)^2\right. \\
& \left. K^2\left(H_x^{-1}(x_t - x_i), H_v^{-1}(v_j - v_i)\right)\right] \\
= \; & \int \frac{|H_x|^4}{4f(x_i)^2 |H_v|^2} \sigma^2(w_j) \left(H_x^{-1}(x_j - x_i)\right)^2 (m^{(2)}(x_{ij}))^2 \left(H_x^{-1}(x_j - x_i)\right)^2 \\
& \times K^2\left(H_x^{-1}(x_t - x_i), H_v^{-1}(v_j - v_i)\right) f(w_i) f(w_j) dw_i dw_j \\
& \text{Let } \psi_x = H_x^{-1}(x_j - x_i), \psi_v = H_v^{-1}(v_j - v_i), \text{ and } \psi = (\psi_x, \psi_v). \\
= \; & \int \frac{|H_x|^4}{4f(x_j + H_x\psi_x)^2 |H|^2} \sigma^2(w_j) \psi_x^2 (m^{(2)}(x_j + (1-\lambda) H_x \psi_x))^2 \psi_x^2 K^2(\psi_x, \psi_v) \\
& \times f(\psi) f(w_j + H_x H_v \psi) |H| \, d\psi dw_j \\
= \; & \int \frac{|H_x|^4}{4f(x_j)^2 |H_v|} \sigma^2(w_j) \psi_x^2 (m^{(2)}(x_j))^2 \psi_x^2 K^2(\psi_x, \psi_v) f(w_j)^2 d\psi dw_j + O() \\
= \; & \frac{|H_x|^4}{|H_v|} \left(\int K^2(\psi) \psi_x^4 d\psi\right) \left(\int \frac{1}{f(x_j)^2} \sigma^2(w_j) (m^{(2)}(x_j))^2 f(w_j)^2 dw_j\right)
\end{aligned}$$

Then, I have

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

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

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

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

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

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

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

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

By H-decomposition,

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

Then,

$$\begin{aligned}
& \mathrm{E}[\phi_{2n}^2(W_j,W_t)] \\
= \; & \mathrm{E}\left[\left(u_j\mathrm{E}\left[\frac{1}{2f(x_i)\,|H_v|}m^*(x_{it})K\left(H_x^{-1}(x_t-x_i)\right)K\left(H_v^{-1}(v_j-v_i)\right)\mid W_j,W_t\right]\right)^2\right] \\
= \; & \mathrm{E}\left[u_j\left(\mathrm{E}\left[\frac{|H_x|^4}{2f(x_i)\,|H|}\left(H_x^{-1}(x_t-x_i)\right)^2(m^{(2)}(x_{it}))^2\left(H_x^{-1}(x_t-x_i)\right)^2\right.\right.\right. \\
& \left.\left.\left.\times\; K\left(H_x^{-1}(x_t-x_i)\right)K\left(H_v^{-1}(v_j-v_i)\right)\mid W_j,W_t\right]\right)^2\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}\left[\sigma^2(v_j)\left(\int \frac{|H_x|^4}{2f(x_i)\,|H|}\left(H_x^{-1}(x_t-x_i)\right)^2(m^{(2)}(x_{it}))^2\left(H_x^{-1}(x_t-x_i)\right)^2 K\left(H_x^{-1}(x_t-x_i)\right)\right.\right.\\
&\quad \left.\left.\times K\left(H_v^{-1}(v_j-v_i)\right)f(w_i)dw_i\right)^2\right]\\
&= \mathrm{E}\left[\sigma^2(v_j)\left(\int \frac{|H_x|^4}{2f(x_t-H_x\psi_{x,1})\,|H|}\psi_{x,1}^2(m^{(2)}(x_t-\lambda H_x\psi_{x,1}))^2\psi_{x,1}^2 K\left(\psi_{x,1}\right)\right.\right.\\
&\quad \left.\left.\times K\left(\psi_{v,1}+H_v^{-1}(v_j-v_t)\right)f(w_t-HH_x\psi_1)\,|H|\,d\psi_1\right)^2\right]\\
&= \int \sigma^2(v_j)\left(\int \frac{|H_x|^4}{2f(x_t-h_x\psi_{x,1})}\psi_{x,1}^2(m^{(2)}(x_t-\lambda H_x\psi_{x,1}))^2\psi_{x,1}^2 K\left(\psi_{x,1}\right)\right.\\
&\quad \left. K\left(\psi_{v,1}+H_v^{-1}(v_j-v_t)\right)f(w_t-HH_x\psi_1)d\psi_1\right)^2 f(w_j)f(w_t)dw_jdw_t\\
&= \int \sigma^2(v_t+H_v\psi_{v,2})\left(\int \frac{|H_x|^4}{2f(x_t-H_x\psi_{x,1})}\psi_{x,1}^2(m^{(2)}(x_t-\lambda H_x\psi_{x,1}))^2\psi_{x,1}^2 K\left(\psi_{x,1}\right)K\left(\psi_{v,1}+\psi_{v,2}\right)\right.\\
&\quad \left. f(w_t-H\psi_1)d\psi_1\right)^2 f(w_j)f(w_t)dw_jdw_t\\
&= \frac{|H_x|^8\,|H|}{4}\int \sigma^2(v_t)\frac{f^4(w_t)}{f^2(x_t)}(m^{(2)}(x_t))^2dw_t\int\left(\int \psi_{x,1}^4 K\left(\psi_{x,1}\right)K\left(\psi_{v,1}+\psi_{v,2}\right)d\psi_1\right)^2 d\psi_2\\
&= O(|H_x|^8\,|H|)
\end{aligned}
$$

Then, I have

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

(I)-(B)-(ii)

$$
\begin{aligned}
S_{2n} &= -\frac{2}{n(n-1)\,|H_v|}\sum_{i=1}^n\sum_{j\neq i}^n u_j\left(\frac{1}{n\,|H_x|\,f(x_i)}\sum_{t=1}^n K\left(H_x^{-1}(x_t-x_i)\right)u_t\right)K\left(H_v^{-1}(v_j-v_i)\right)\\
&= -\frac{2}{n(n-1)\,|H_v|\,|H_x|}\sum_{i=1}^n\sum_{j\neq i}^n\sum_{t=1}^n\frac{1}{f(x_i)}u_ju_tK\left(H_x^{-1}(x_t-x_i)\right)K\left(H_v^{-1}(v_j-v_i)\right)\\
&= -\frac{2}{n(n-1)\,|H_v|\,|H_x|}\sum_{i=1}^n\frac{1}{f(x_i)}\left[\sum_{j\neq i}^n\sum_{t=1}^n u_ju_tK\left(H_x^{-1}(x_t-x_i)\right)K\left(H_v^{-1}(v_j-v_i)\right)\right.\\
&\quad \left.+\sum_{j\neq i}^n\sum_{t=1}^n u_tu_jK\left(H_x^{-1}(x_t-x_i)\right)K\left(H_v^{-1}(v_j-v_i)\right)\right]
\end{aligned}
$$

1) $i = t$

$$\begin{aligned}
S_{2n} &= -\frac{2}{n(n-1)\,|H_v|\,|H_x|}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{f(x_i)}u_j u_i K\left(H_x^{-1}(x_j-x_i), H_v^{-1}(v_j-v_i)\right) \\
&= -\frac{2}{n(n-1)\,|H_v|\,|H_x|}\sum_{i<j}^{n}\sum^{n}\left[\frac{1}{f(x_i)}u_j u_i K\left(H_x^{-1}(x_j-x_i), H_v^{-1}(v_j-v_i)\right)\right. \\
&\quad \left. +\frac{1}{f(x_j)}u_i u_j K\left(H_x^{-1}(x_j-x_i), H_v^{-1}(v_j-v_i)\right)\right] \\
\text{By letting } W_i &= (u_i, x_i, v_i), \\
&= -\frac{2}{n(n-1)}\sum_{i<j}^{n}\sum^{n}\left[\psi_n(W_i,W_j)+\psi_n(W_j,W_i)\right] \\
&= -\frac{2}{n(n-1)}\sum_{i<j}^{n}\sum^{n}\phi_n(W_i,W_j)
\end{aligned}$$

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

$$\begin{aligned}
\frac{1}{|H_v|\,|H_x|}\mathrm{E}[\phi_n^2(W_i,W_j)] &= \frac{1}{|H_v|\,|H_x|}\mathrm{E}[\phi_n^2(Z_j,Z_i)] \\
&= \frac{1}{|H_v|\,|H_x|}\mathrm{E}\left[\frac{1}{f(x_i)^2}u_j^2 u_i^2 K^2\left(H_x^{-1}(x_j-x_i), H_v^{-1}(v_j-v_i)\right)\right] \\
&= \frac{1}{|H_v|\,|H_x|}\int\frac{1}{f(x_i)^2}\sigma^2(w_i)\sigma^2(w_j)K^2\left(H_x^{-1}(x_j-x_i), H_v^{-1}(v_j-v_i)\right) \\
&\quad \times f(w_i)f(w_j)dw_i dw_j
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{|H_v|\,|H_x|}\int\frac{1}{f(x_i)^2}\sigma^2(w_i)\sigma^2(w_i+H\psi)K^2(\psi)\,f(w_i)f(w_i+H\psi)\,|H|\,|H_x|\,dw_i d\psi \\
&= \int\frac{1}{f(x_i)^2}\sigma^4(w_i)K^2(\psi)\,f(w_i)^2 dw_i d\psi \\
&= \left(\int K^2(\psi)d\psi\right)\left(\int\frac{1}{f(x_i)^2}\sigma^4(w_i)f(w_i)^2 dw_i\right)
\end{aligned}$$

Therefore, I have

$$n\left(|H_v|\,|H_x|\right)^{1/2}S_{2n}\xrightarrow{d}N(0,\Sigma),\ \text{where } \Sigma = \left(\int K^2(\psi)d\psi\right)\left(\int\frac{1}{f(x_i)^2}\sigma^4(w_i)f(w_i)^2 dw_i\right)$$

Then,

$$
\begin{aligned}
n\,|H_v|^{1/2}\,S_{2n} &= (n^{-1}\,|H_x|^{-1/2})(n\,(|H_v|\,|H_x|)^{1/2}\,S_{2n}) \\
&\rightarrow 0 \text{ as } n \rightarrow \infty
\end{aligned}
$$

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

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

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

Then, I have

$$
\begin{aligned}
S_{2n} &= n^{-1}\left(\mathrm{E}[\psi_n^2(W_i, W_j)]\right)^{1/2} \\
&= O(n^{-1}\,(|H_v|\,|H_x|)^{-1/2})
\end{aligned}
$$

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

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

By H-decomposition,

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

Then, I can write U-statistic

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

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

$$\begin{aligned}
&= \int \sigma^2(v_j)\sigma^2(v_t)\left(\int \frac{1}{f(x_t - H_x\psi_{x,1})}K\left(\psi_{x,1}\right)K\left(\psi_{v,1} + H_v^{-1}(v_j - v_t)\right)f(w_t - HH_x\psi_1)d\psi_1\right)^2 \\
&\quad \times f(w_j)f(w_t)dw_j dw_t \\
&= \int \sigma^2(v_t)\sigma^2(v_t + H\psi_{v,2})\left(\int \frac{1}{f(x_t - H_x\psi_{x,1})}K\left(\psi_{x,1}\right)K\left(\psi_{v,1} + \psi_{v,2}\right)f(w_t - HH_x\psi_1)d\psi_1\right)^2 \\
&\quad \times f(w_t)f(w_t + HH_x\psi_2)\left|H_v\right|\left|H_x\right|d\psi_2 dw_t \\
&= \left|H_v\right|\left|H_x\right|\left(\int \sigma^4(v_t)\frac{f^4(w_t)}{f^2(x_t)}dw_t\right)\int\left(\int K\left(\psi_{x,1}\right)K\left(\psi_{v,1} + \psi_{v,2}\right)d\psi_1\right)^2 d\psi_2 \\
&= O(\left|H_v\right|\left|H_x\right|)
\end{aligned}$$

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

In summary,

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

(I)-(B)-(iii)

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

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

(I)-(C)

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

$$
\begin{aligned}
|\hat{\mu}_n(x_i) - \mu_n(x_i)| &= \frac{1}{n\,|H_v|}\left| e_1' S_n(z_t)^{-1} \sum_{t=1}^n z_t K\left(H_x^{-1}(x_t - x_i)\right)\left(\frac{1}{2}m^*(x_{it}) + u_t\right)\right. \\
&\quad \left. - \frac{1}{f(x_i)} \sum_{t=1}^n z_t K\left(H_x^{-1}(x_t - x_i)\right)\left(\frac{1}{2}m^*(x_{it}) + u_t\right)\right| \\
&= \frac{1}{n\,|H_v|}\left| e_1'(S_n(z_t)^{-1} - S(z_t)^{-1}) \sum_{t=1}^n z_t K\left(H_x^{-1}(x_t - x_i)\right)\left(\frac{1}{2}m^*(x_{it}) + u_t\right)\right| \\
&\leq \frac{1}{|H_v|}((1,0)(S_n(z_t)^{-1} - S(z_t)^{-1})^2(1,0)')^{1/2} \\
&\quad * \frac{1}{n}\left(\left|\sum_{t=1}^n K\left(H_x^{-1}(x_t - x_i)\right)\left(\frac{1}{2}m^*(x_{it}) + u_t\right)\right|\right. \\
&\quad \left. + \left|\sum_{t=1}^n \left(H_x^{-1}(x_t - x_i)\right)' K\left(H_x^{-1}(x_t - x_i)\right)\left(\frac{1}{2}m^*(x_{it}) + u_t\right)\right|\right)
\end{aligned}
$$

I follow Lemma 2 of Martins-Filho and Yao (2007) to obtain $|\hat{\mu}_n(x_i) - \mu_n(x_i)| = O\left(|H|\left(\frac{\ln n}{n|H_x|}\right)^{\frac{1}{2}}\right) + O(|H_x|^3)$.

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 $x_{ij} = \lambda x_i + (1-\lambda)x_j$,

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

Note that

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

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

Therefore,

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

(II)

By using

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

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

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

(III)

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

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

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

(IV)

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

By using the properties of U-statistics, I can easily derive that $\Omega = 2\int K^2(\psi)d\psi\mathrm{E}[\sigma^4(v)f(v)]$.

Proof of Theorem 3

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 $\delta_n = n^{-1/2} \left(|H_v| \right)^{-1/4}$.

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

Here, note that

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

For the following sections, I will show

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

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

(I)-(A)

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

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

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

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

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

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

Therefore, I have

$$n\left|H_v\right|^{1/2}Q_{1n}\xrightarrow{d}N(0,\Omega),\ \text{where } \Omega=2\left(\int K^2\left(\psi_v\right)d\psi_v\right)\mathrm{E}\left[\sigma^4(v_i)f(v_i)\right]$$

(I)-(B)

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

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

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

Then, I have

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

Therefore,

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

(I)-(C)

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

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

Therefore, I have

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

In summary,

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

Proof of Theorem 4

Under H_1,

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

Then, note that

$$\begin{aligned}
\mathrm{E}\left[u_i^2 \mid v_i\right] &= \mathrm{E}\left[(\varepsilon_i + h(v_i))^2 \mid v_i\right] \\
&= \sigma^2(v_i) + (h(v_i))^2
\end{aligned}$$

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

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

Now,

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

Then,

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

Therefore, I have

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

In summary,

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