KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE OF ECONOMIC RESEARCH

Discussion Paper No.948

"Nonparametric tests for the effect of treatment on conditional variance"

Yanchun Jin

August 2016

KYOTO UNIVERSITY

KYOTO, JAPAN

Nonparametric tests for the effect of treatment on conditional variance*

Yanchun Jin†

August, 2016

Abstract

This paper proposes nonparametric tests for the null hypothesis that a treatment has a zero effect on conditional variance for all subpopulations defined by covariates. Rather than the mean of outcome, which measures to what extent treatment changes the level of outcome, researchers are also interested in how the treatment affects the dispersion of outcome. We use variance to measure the dispersion and estimate the conditional variances by series method. We give a test rule comparing a Wald-type test statistic with the critical value from chi-squared distribution. We also construct a normalized test statistic that is asymptotically standard normal under the null hypothesis. We illustrate the usefulness of the proposed test by Monte Carlo simulations and an empirical example that investigates the effect of unionism on wage dispersion.

Keywords: treatment effect, conditional variance, series estimation.

JEL Classification: C12, C14, C21.

*Graduate School of Economics, Kyoto University, Yoshida Honmachi, Sakyo-ku, Kyoto City, Kyoto, 6068501, Japan; Research Fellow of Japan Society for Promotion of Science. E-mail: jin.yanchun.74v@st.kyoto-u.ac.jp

†I thank my supervisors, Yoshihiko Nishiyama and Ryo Okui, for their continuing guidance and support. I also appreciate insightful comments from seminar participants at Kyoto University and attendees at the Kansai Econometrics Meeting held in Kyoto. This work was supported by JSPS KAKENHI Grant Number 15J10229. All remaining errors are mine.

1 Introduction

Recently, treatment effect analysis has been an important tool in various fields of empirical research to evaluate the impacts of policies. In this paper, we consider the effect of treatment on variance. Most of the literature focuses on various treatment effects on the mean of the interested outcome, such as average treatment effect and local average treatment effect. These parameters measure to what extent the treatment changes the level of outcome. However, researchers are also interested in the treatment effect on the dispersion of outcome. For example, it is of substantive interest to investigate how unionism affects wage dispersion. A number of researches show that wages are flatter in union sectors compared to nonunion (Freeman, 1980; Gosling and Machin, 1995; DiNardo et al., 1996; Card, 2001). Freeman (1980), comparing the variances of the wages for union workers and nonunions, presented that the unionism reduces wage differential in the organized sector and this difference-reducing-effect within sectors is larger than the gap-increasing-effect across industries. Investigating the treatment effect on variance is important in understanding how unionism works.

In this paper, we provide nonparametric tests for the effect of the treatment on variance. In particular, we consider a test for the null hypothesis that the treatment has nonzero effect on the dispersion of the outcome for all subpopulations defined by covariates. It is useful, for example, when one wants to make it clear whether there is any evidence of heterogeneity in effect of unionism on dispersion of outcome. Card (2001) examined the gap of variance of wages for union and nonunion male workers defined by various skill groups and found that the role of unions in compressing wage dispersion for high-skilled workers is slightly stronger than low-skilled. Conducting the test proposed in this paper, we can study whether there is any subpopulations for which unions change their wage dispersion.

Although a large part of the recent literature on treatment effect focuses on the estimation, studies on hypothesis testing for treatment effect are limited. Abadie (2002) concerning the distributions of the outcome for the treatment group and control, tests for the null hypothesis of the equality and first-order stochastic dominance using bootstrap method. Crump et al. (2008) consider the test of the treatment effect heterogeneity and develop tests based on series estimation. They test for the null hypothesis that the average treatment effects conditional on the covariates are zero for all subpopulations defined by covariates. Also, they propose the test for the null hypothesis that the average effect conditional on the covariates is constant for all subpopulations.

Lee (2009) studies a nonparametric test of the null hypothesis of no distributional treatment effect for randomly censored outcomes. Hsu (2011) studies a Kolmogorov-Smirnov (KS) test of similar null hypothesis employing Andrews and Shi's (2013) instrumental variable approach. Chang et al. (2015) construct a test for the null hypothesis of conditional stochastic dominance treatment effect and positiveness of conditional average treatment for all covariates with test statistics based on kernel estimators.

None of the hypotheses considered in these papers provide tests for treatment effect on conditional variance. In this paper, we test for treatment effect on conditional variance. The null hypothesis considered in this paper is that the difference of conditional variance of outcome between the treatment group and control is zero for all subpopulations defined by covariates. To construct the test statistic, we need to estimate the conditional variance function. However, reasonable specifications for the conditional variance are limited, which makes it difficult to apply parametric method to test. For this reason, it is necessary to use a nonparametric method.

We provide tests based on a two-step series approach to estimating conditional variances. In the first step, we estimate the conditional mean function using series and then compute residuals. Then we estimate conditional variance in the second step by regressing squared residuals on a power series. We conduct the test using a Wald-type statistic. We compare the test statistic to the critical value of chi-squared distribution with a degree equal to the number of series terms. In addition, we give a normalized test statistic that is of F-statistic form and compare it to the critical value from a standard Gaussian distribution.

Doing regression using squared residuals, we take into account two kinds of biases: bias of the squared residuals in the first step and bias from conditional variance function estimation which arises in the usual nonparametric regression analysis. Given particular series terms, the test can be viewed as a test of whether coefficients for the treatment group and control estimated in the second step are identical. Thus, we can conduct the test as if it were set by a parametric model, which is easy to implement. We give some conditions under which the normalized test statistic converges to a standard Gaussian distribution when the null hypothesis holds. A key result leading to this asymptotic property is the theorem (Bentkus, 2005) that ensures the convergence to multivariate normality is fast enough even with the dimension of the vector increasing. Our tests extend the method in Crump et al. (2008) which considers a one-step test for the conditional mean.

Our tests are close to Hong and White's (1995) nonparametric tests in that they also estimate the nonparametric model by series regression and provide test statistics that converge in distribution to a unit normal under correct specification. Their test can be viewed as a test of the joint hypothesis that the true parameters of a series regression model are zero. They provide the conditions for the number of series terms to ensure the validity of their tests. In our tests, the increasing rates of the series terms are also important to make the test statistic converge to a standard normal. This paper is also related to the literature on the estimation of conditional variance. This problem was first studied when the explanatory variable is univariate. For example, Fan and Yao (1998) apply the local linear regression model to the squared residuals to estimate conditional variance; Song and Yang (2009) apply the polynomial spline regression model; and Yu and Jones (2004) apply the kernel-weighted local polynomial regression model. For a multivariate model, Zhu et al. (2013) consider a single-index structure to estimate the conditional variance function and provide an estimation that remains consistent even when the structure of the variance function is misspecified. In this paper, we test the hypothesis using a power series estimator of the coefficients of conditional variance functions, which is easy to compute.

The rest of this paper is organized as follows. In Section 2, we give the framework for the program evaluation analysis and give the null hypothesis and the alternative we consider in this paper. Section 3 illustrates the test in a parametric model. Then, in Section 4, we extend it to the nonparametric model with series estimation and provide the test statistics. In Section 5, we give the asymptotic theorem for our test statistic under some assumptions. In Section 6, we conduct a simulation and demonstrate the result of the test property in a finite sample. In Section 7, we consider an empirical application regarding the effects of unionism on the dispersion of wages using National Longitudinal Survey data. Section 8 concludes the paper.

2 Framework

Our basic framework is standard in the treatment effect literature. We have a random sample of size N. For each unit $i = 1, \ldots, N$ in the sample, let W_i indicate whether the treatment of interest is received, with $W_i = 1$ if unit i receives the treatment, and $W_i = 0$ if unit i receives the control treatment. Let $Y_i(1)$ and $Y_i(0)$ denote potential outcomes for each unit i under treatment and control, respectively. For each unit i, we observe W_i and Y_i, where $Y_i \equiv Y_i(W_i) = W_i \cdot Y_i(1) + (1 - W_i) \cdot Y_i(0)$. In addition, we observe a vector of pretreatment variables, denoted by X_i, the support of which is $\mathcal{X} \subset \mathbb{R}^d$. The treatment effect on conditional variance is $Var[Y(1)|X =$

$x] - Var[Y(0)|X = x]$.

Our test is concerned with the null hypothesis

$$H_0 : \forall x \in \mathcal{X}, Var[Y(1)|X = x] - Var[Y(0)|X = x] = 0, \tag{2.1}$$

against the alternative

$$H_1 : \exists x \in \mathcal{X}, Var[Y(1)|X = x] - Var[Y(0)|X = x] \neq 0. \tag{2.2}$$

Under the null hypothesis, for all values of the covariates, the treatment has no effect on the conditional variance; whereas under the alternative, there are some values of covariates where there is some effect on the conditional variance.

For $w = 0, 1$, let $\mu_w(x) = E[Y(w)|X = x]$ and $\epsilon_{w,i} = Y_i(w) - \mu_w(X_i)$ and assume $E[\epsilon_{w,i}|X] = 0$, then $\sigma_w^2(x) \equiv Var[Y(w)|X = x] = E[\epsilon_{w,i}^2|X = x]$. So the hypotheses can be stated as

$$H_0 : \forall x \in \mathcal{X}, \sigma_1^2(x) - \sigma_0^2(x) = 0, \tag{2.3}$$

$$H_1 : \exists x \in \mathcal{X}, \sigma_1^2(x) - \sigma_0^2(x) \neq 0. \tag{2.4}$$

Note that regardless of whether the mean functions are identical, we are only interested in the equality of variances.

Now, we make assumptions standard in the program evaluation literature.

First, we assume the sample is an i.i.d random sample.

Assumption 2.1. *(Independent and Identically Distributed Random Sample):*

Random variables $(Y_i, W_i, X_i), i = 1, \ldots, N$ are independent and identically distributed.

The central problem of treatment effect literature is that for unit i, we observe either $Y_i(1)$ or $Y_i(0)$, but never both. To achieve identification, we assume the unconfoundedness (Rosenbaum and Rubin, 1983), which can be described as

Assumption 2.2. *(Unconfoundedness):*

$$W \perp (Y(0), Y(1))|X,$$

where $\perp$ denotes the independence.

In addition, we assume that in the population for all values of covariates, there are both treatment and control units.

Assumption 2.3. *(Overlap):*

$$0 < \Pr(W = 1|X = x) < 1$$

Under these assumptions, the $\sigma_1^2(x)$ and $\sigma_0^2(x)$ can be identified as computing the conditional variances for the treatment group and control, respectively. Then we can implement our test with difference between the two conditional variances.

Without loss of generality, we arrange the data such that the first N_1 observations have $W_i = 1$, and the last N_0 observations have $W_i = 0$. Define the covariates for this rearranged data as an $N \times d$ matrix $\mathbf{X} = (X_1', \ldots, X_{N_1}', X_{N_1+1}', \ldots, X_N')$. Also let $N_1 \times d$ matrix $\mathbf{X}_1 = (X_1', X_2', \ldots, X_{N_1}')$, $N_0 \times d$ matrix $\mathbf{X}_0 = (X_{N_1+1}', X_{N_1+2}', \ldots, X_N')$, and let N_1 vector $\mathbf{Y}_1 = (Y_1, \ldots, Y_{N_1})'$, N_0 vector $\mathbf{Y}_0 = (Y_{N_1+1}, \ldots, Y_N)'$.

3 Test Statistic in Parametric Models

We first give a test in a standard parametric model, which helps to explain the procedure in nonparametric settings. To construct the test statistic, it is necessary to estimate $\sigma_1^2(x) - \sigma_0^2(x)$. We specify $\sigma_w^2(x)$ as linear function

$$\sigma_w^2(x) = \beta_w' x. \tag{3.1}$$

In this parametric setting, the null and alternative hypotheses are

$$H_0 : \beta_1 = \beta_0, \tag{3.2}$$

$$H_1 : \beta_1 \neq \beta_0, \tag{3.3}$$

where β_1 is in the K-dimension. This can be tested using Wald-type test statistic

$$T_{para} = (\hat{\beta}_1 - \hat{\beta}_0)'(\hat{\Omega}_1/N_1 + \hat{\Omega}_0/N_0)^{-1}(\hat{\beta}_1 - \hat{\beta}_0),$$

where $\hat{\beta}_w$ is an estimator of β, Ω_w is the estimator for asymptotic variance matrix of $\hat{\beta}_w$, and N_1 and N_0 are the sample sizes for the treated and control groups, respectively.

Now we consider the least square estimator. Note that we conduct regressions in the treatment and control group, respectively to compute $\hat{\beta}_1$ and $\hat{\beta}_0$. Here we also assume mean functions as standard linear models, $\mu_w(x) = \xi_w' x$, and estimate the coefficients by their least square estimators $\hat{\gamma}_w$. Then the residuals are $\hat{\boldsymbol{\epsilon}}_w = \mathbf{Y}_w - \hat{\mu}_w(\mathbf{X}_w) = \mathbf{Y}_w - \mathbf{X}_w \hat{\xi}_w$. Let $\hat{\epsilon}_{1,i}$ be the i-th element of $\hat{\boldsymbol{\epsilon}}_1$

and $\hat{\epsilon}_{0,i}$ be the $(i-N_1)$-th element of $\hat{\boldsymbol{\epsilon}}_0$. This leads to the residual-based estimator as $\hat{\sigma}_w^2(x) = \hat{\beta}_w' x$ by solving the following problem,

$$\hat{\beta}_w = \arg\min_{\beta} \sum_{i|W_i=w}^{n} (\hat{\epsilon}_{w,i}^2 - \beta_w' X_i)^2.$$

Then under (3.2), T_{para} converges to a *chi*-squared distribution with K degrees of freedom.

However, specification of the variance function as model (3.1) is not standard, and there is no widely acceptable model for variance. To get rid of misspecification, we estimate the conditional variance function using the nonparametric method. In the next section, we extend the parametric test described above to the nonparametric procedure and provide a valid test without parametric specification.

4 Nonparametric Estimation of Conditional Variances

We estimate two conditional variances by running a "second step" model for the squared regression residuals obtained in the first step. Instead of specifying the function by standard linear model, here we use series estimators in both steps. In the first step, we estimate $\mu_w(x)$ by $\hat{\mu}_{w,K_1}(x)$ developed by Imbens et al. (2005), and then compute the residuals $\hat{\epsilon}_{w,i}$. Then in the second step, we estimate $\sigma_w^2(x)$ by $\hat{\sigma}_{w,K_2}^2(x)$ using $\hat{\epsilon}_{w,i}^2$ where K_1 and K_2 denote the number of series terms in two steps, respectively. Let $\lambda(d) = (\lambda_1, \ldots, \lambda_d)$ be a d-dimensional vector of non-negative integers, with $|\lambda(d)| = \sum_{m=1}^{d} \lambda_m$, and let $x^{\lambda(d)} = x_1^{\lambda_1} x_2^{\lambda_2} \ldots x_d^{\lambda_d}$. Consider a series $\{\lambda(l)\}_{l=1}^{\infty}$ containing all distinct vectors such that $|\lambda(l)|$ is nondecreasing. Let $p_l(x) = x^{\lambda(l)}$, $P_l(x) = (p_1(x), \ldots, p_l(x))'$. Let $P_{K_1}(X_i)$ denote K_1 series terms for the mean function and $P_{K_2}(X_i)$ denote K_2 series terms for the variance function. Define the $N_1 \times K_1$ matrix $P_{1,K_1} = (P_{K_1}'(X_1), \ldots, P_{K_1}'(X_{N_1}))$ and $N_0 \times K_1$ matrix $P_{0,K_1} = (P_{K_1}'(X_{N_1+1}), \ldots, P_{K_1}'(X_N))$. Also, define the $N_1 \times K_2$ matrix $P_{1,K_2} = (P_{K_2}'(X_1), \ldots, P_{K_2}'(X_{N_1}))$ and $N_0 \times K_2$ matrix $P_{0,K_2} = (P_{K_2}'(X_{N_1+1}), \ldots, P_{K_2}'(X_N))$.

Then the nonparametric series estimator of the regression function $\mu_w(x)$, given series terms P_{w,K_1}, is given by

$$\hat{\mu}_{w,K_1}(x) = P_{K_1}(x)'(P_{w,K_1}' P_{w,K_1})^{-} P_{w,K_1}' \mathbf{Y}_w, \tag{4.1}$$

where A^- denotes a generalized inverse of A. Then we compute the residuals by $\hat{\boldsymbol{\epsilon}}_w = \mathbf{Y}_w - \hat{\mu}_w(\mathbf{X}_w)$. The estimator of $\sigma_w^2(x)$ regressed by P_{w,K_2} is given by

$$\hat{\sigma}_{w,K_2}^2(x) = P_{K_2}(x)'(P_{w,K_2}' P_{w,K_2})^{-} P_{w,K_2}' \hat{\boldsymbol{\epsilon}}_w^2, \tag{4.2}$$

where $\hat{\boldsymbol{\epsilon}}_w^2$ denotes the vector whose elements are square of elements of $\hat{\boldsymbol{\epsilon}}_w$. In addition, for fixed K_2, we estimate the approximated limiting variance of $\sqrt{N_w}\hat{\eta}_{w,K_2}$ by $\hat{\Omega}_{P,w,K_2}^{-1}\hat{\Psi}_{P,w,K_2}\hat{\Omega}_{P,w,K_2}^{-1}$, where

$$\hat{\Omega}_{P,w,K_r} = \frac{P'_{w,K_r}P_{w,K_r}}{N_w}, \quad \hat{\Psi}_{P,w,K_2} = \frac{P'_{w,K_2}\hat{M}_{w,K_2}P_{w,K_2}}{N_w},$$

where $\hat{M}_{w,K_2} = diag\{(\hat{\epsilon}_{w,i}^2 - \hat{\sigma}_{w,K_2}^2(X_i))^2$, for i with $W_i = w\}$.

We give a test statistic as

$$Q = (\hat{\eta}_{1,K_2} - \hat{\eta}_{0,K_2})'\hat{V}_{P,K_2}(\hat{\eta}_{1,K_2} - \hat{\eta}_{0,K_2}),$$

where

$$\hat{V}_{P,K_2} = \frac{\hat{\Omega}_{P,0,K_2}^{-1}\hat{\Psi}_{P,0,K_2}\hat{\Omega}^{-1}}{N_0} + \frac{\hat{\Omega}_{P,1,K_2}^{-1}\hat{\Psi}_{P,1,K_2}\hat{\Omega}_{P,1,K_2}^{-1}}{N_1}$$

is the estimate for V_{P,K_2}, the variance of $\hat{\eta}_{1,K_2} - \hat{\eta}_{0,K_2}$. Test with Q is a Wald-type test to detect whether the coefficients are identical,which makes the test similar to the parametric test discussed in Section 3 when the parametric model is

$$\sigma_w^2(x) = P_{K_2}(x)'\eta_{w,K_2} \tag{4.3}$$

In addition, we give a normalized test statistic as

$$T = \frac{(\hat{\eta}_{1,K_2} - \hat{\eta}_{0,K_2})'\hat{V}_{P,K_2}^{-1}(\hat{\eta}_{1,K_2} - \hat{\eta}_{0,K_2}) - K_2}{\sqrt{2K_2}}, \tag{4.4}$$

$$\tag{4.5}$$

for the test of the null hypothesis. We will see that in large samples, T has a standard normal distribution under the null.

5 Asymptotic Theory

This section provides asymptotic theory for our test statistic T. We first state the conditions in addition to assumptions 2.1-2.3, we make the following assumptions to develop asymptotic theory.

Assumption 5.1. *(Distribution of Covariates):*

$X \in \mathcal{X} \subset \mathbb{R}^d$, where X is the Cartesian product of intervals $[x_{jL}, x_{jU}]$, $j = 1, \ldots, d$, with $x_{jL} < x_{jU}$. The density of X is bounded away from zero on $\mathcal{X}$.

Assumption 5.2. *(Conditional Variance Distributions):*

1. *The mean regression functions $\mu_w(x)$ are s_1 times continuously differentiable, and variance regression functions $\sigma_w^2(x)$ are s_2 times continuously differentiable, with $s_1/d > 2$ and $s_2/d > 7$.*

2. *For $\epsilon_i(w) = Y_i(w) - \mu_w(X_i)$, $u_i(w) = (\epsilon_i(w))^2 - \sigma_w^2(X_i)$,*

   (a) $\forall x \in \mathcal{X}$, $0 < \underline{\theta}^2 \leq \theta_w^2(x) \leq \bar{\theta}^2 < \infty$, $0 < \underline{\sigma}^2 \leq \sigma_w^2(x) \leq \bar{\sigma}^2 < \infty$.

   (b) $E[(u_i(w))^4] < \infty$.

Assumption 5.3. *(Rates for Series Estimators):*

The numbers of terms in the series, K_r, $r = 1, 2$ increase with the sample size N as $K_r = CN^{v_r}$, for an arbitrary positive constant C and some v_r such that $2d/(d + 2s_1) < v_1 < 1/3$ and $2d/(4s_2 - d) < v_2 < \min(2/13, (s_1 - d)v_1/(2d))$.

Because $\Pr(W = 1|X = x)$ is assumed to be bounded from 0, assumption 5.1 implies that the density of X conditioned on $W = w$ is also bounded away from 0 on its support. Assumption 5.2 imposes some smoothness and moment conditions on the data to ensure the asymptotic convergence of estimators. Assumption 5.2.2-(a) ensures the nonsingularity of $\hat{V}_{P,K_2}$ with probability approaching 1. Assumption 5.3 defines the increasing rates of the number of series terms. It guarantees that the test statistic converges to the standard normal under the null hypothesis. $v_1 < 1/3$ ensures that the eigenvalues of $\hat{\Omega}_{w,K_1}$ are bounded and bounded away from 0, which we should take into consideration in the second step. We find that the increasing rates of series terms used in the second step are also affected by the increasing speed v_1 used in the first step because we regressed the squared errors obtained from mean function estimation in the first step. In practice, we can use "top-down" or"bottom-up" method to select the covariates as in Section 7.

The following theorem shows that our normalized test statistic has a standard normal distribution asymptotically.

Theorem 1. *Suppose assumptions 2.1-2.3 and 5.1-5.3 hold. Then under H_0, we have,*

$$T \xrightarrow{d} N(0, 1).$$

.

Proof. See Appendix. □

The key is the fact that the *chi*-squared distribution converges to the normal distribution as the degree of freedom increases. In empirical application, we can use the test rule with Q for small-sample that does not affect large-sample properties of T. We can test the hypothesis comparing quadratic form Q with the critical values of a chi-squared distribution with K_2 degrees of freedom. Note that if Q has a chi-squared distribution with degrees of freedom equal to K_2, it leads to approximate asymptotic normality of Tin large samples. Hence, in large samples, tests with Q and T are approximately the same decision rules. The nonparametric test using T does not rely on the correct specification but relies on the order of power series.

Now we provide an informal description to understand the proof of the test. We define the pseudo-true values, η^*_{w,K_2} for $w = 0, 1,\ K_2 = 1, 2, \ldots,$ as

$$\begin{aligned}\eta^*_{w,K_2} &= \arg\min_{\eta} E[(\sigma^2_w(X) - P_{K_2}(X)'\eta)^2|W = w] \\ &= (E[R_{K_2}(x)P_{K_2}(X)'|W = w])^{-1}E[P_{K_2}(X)\epsilon^2_w|W = w]\end{aligned}$$

so that for fixed K_2, as $N \to \infty$, $\hat{\eta}_{w,K_2} \to \eta^*_{w,K_2}$.

$$\begin{aligned}&V^{-1/2}_{P,K_2} \cdot (\hat{\eta}_{1,K_2} - \hat{\eta}_{0,K_2}) \\ =&V^{-1/2}_{P,K_2} \cdot (\eta^*_{1,K_2} - \eta^*_{0,K_2}) + V^{-1/2}_{P,K_2} \cdot (\hat{\eta}_{1,K_2} - \eta^*_{1,K_2}) + V^{-1/2}_{P,K_2} \cdot (\hat{\eta}_{0,K_2} - \eta^*_{0,K_2}).\end{aligned}$$

Under assumptions 2.1-2.3 and 5.1-5.3, the last two terms are normally distributed with mean 0 for given K_2 in large samples. We can have the asymptotic distribution of T based on this approximate normality. The first term can be ignored for large K_2 because $\sigma^2_w(x)$ is close to $P_{K_2}(x)'\eta^*_{w,K_2}$ for all x. And, under the null hypothesis, the difference between $P_{K_2}(x)'V^{-1/2}_{P,K_2}\eta^*_{1,K_2}$ and $P_{K_2}(x)'V^{-1/2}_{P,K_2}\eta^*_{0,K_2}$ is close to 0. We can maintain these properties by controlling the increasing speed of K_2. We increase K_2 fast enough to make deviation of the first term from 0 small, while at the same time slowly enough to make the normal distribution of the last two terms hold. Note that by Bentkus (2005, Theorem 1.1), we can appropriate the distribution of a vector with a multivariate standard Gaussian distribution fast enough while the dimension of the vector increases. This contributes to the normalized quadratic form of the test statistic converging to a normal standard.

Our test considers the conditional variances while Crump et al. (2008) consider a similar issue in the context with two conditional means. Their test can be considered as the test with the first step estimators in this paper. In the second step, conducting regression of the residuals obtained from the first step estimation, we should take into consideration the error from the mean estimation. This can be intuitively explained as follows. Observe that

$$\hat{\epsilon}_{w,i}^2 - \epsilon_{w,i}^2 = (\epsilon_{w,i} + \mu_w(X_i) - \hat{\mu}_{w,K}(X_i))^2 - \epsilon_{w,i}^2 \tag{5.1}$$

$$= 2\epsilon_{w,i}(\mu_w(X_i) - \hat{\mu}_{w,K}(X_i)) + (\mu_w(X_i) - \hat{\mu}_{w,K}(X_i))^2. \tag{5.2}$$

We find that the bias of squared residuals is of the same order as the bias of $\hat{\mu}_w$ itself.

We also consider the properties of the test statistic under the local alternative.

Theorem 2. *Consistency of Test under Local Alternative: Suppose assumptions 2.1-2.3 and 5.1-5.3 hold. Then, under the local alternative hypothesis,*

$$\sigma_1^2(x) - \sigma_0^2(x) = \rho_N \cdot \Delta(x)$$

with $\Delta(x)$ s_2 times continuously differentiable, $|\Delta(x_0)| = C_0 > 0$ for some x_0, and $\rho_N^{-1} = O(N^{1/2-3v/2_1-3v_2/2-\varepsilon})$ for some $\varepsilon > 0$. Then, as $N \to \infty$, for all M,

$$\Pr(T \geq M) \to 1.$$

Proof. See Appendix. □

This theorem shows that we can test the alternatives when the two conditional variance functions are arbitrarily close to $N^{-1/2}$ under sufficient smoothness conditions. We can see when the true model is as (4.3) for a fixed K_2, the nonparametric test will decrease the power of the test. The nonparametric test checks for additional parameters larger than the necessary k, whereas under parametric model, this would be zero. Therefore, this additional procedure reduces the power.

6 Monte Carlo Experiment

In this section, we conduct finite-sample Monte Carlo simulations to illustrate the finite sample performance of our test.

Our data-generating process is as follows.

$$X_i \sim U[0,1]$$

$$\mu_0(X_i) = X_i,$$

$$
\begin{aligned}
&\mu_1(X_i) = 5X_i^2 + 2X_i,\\
&\epsilon_{0,i} = \tau_0\sqrt{\sigma_0^2(X_i)},\\
&\epsilon_{1,i} = \tau_1\sqrt{\sigma_1^2(X_i)/2},\\
&Y_{0,i} = \mu_0(X_i) + \epsilon_{0,i} = x + \tau_0\sqrt{e^{X_i}}\\
&Y_{1,i} = \mu_1(X_i) + \epsilon_{1,i} = 5X_i^2 + 2X_i + \tau_1\sqrt{e^{X_i}/2},
\end{aligned}
$$

where τ_0 and τ_1 are parameters in constructing the variance function, with $\tau_0 \sim N(0,1)$, $\tau_1 \sim t$ distribution with freedom 4. $\sigma_1^2(x)$ and $\sigma_0^2(x)$ denote the conditional variances for treatment group and control, respectively, which are of interest. In the following, we consider different specifications for $\sigma_1^2(x)$ and $\sigma_0^2(x)$ in the cases where the null hypothesis is correct and where the alternative is correct.

First, we simulate the asymptotic property of the test statistic under the null hypothesis. We specify the two identical variance functions as follows:

$$
\sigma_0^2(X_i) = \sigma_1^2(X_i) = e^{X_i}
$$

In the experiment, we consider three different sample sizes, $N_0 = 45, N_1 = 55; N_0 = 200, N_1 = 300; and N_0 = 450, N_1 = 550$ and 10000 repetitions. We use power series with order $K = 2, 3, 4$. We use $1, x$ for $K = 2$; $1, x, x^2$ for $K = 3$; and $1, x, x^2, x^3$ for $K = 4$. In addition, we conduct the test using the series terms selected by "bottom-up" method mentioned in Section 7.

Table 1 summarizes the results of the experiments. It shows the empirical rejection probabilities when testing the null hypothesis that the two conditional variance functions are identical. The left panel of Table 1 shows the rejection rates of the nonparametric test with statistic T under significance level 10%, 5%, and 1%, respectively, and the right panel shows the probabilities of the nonparametric test with statistic Q under the same significance level. The number in the bracket shows the average number of selected series terms. From the table, we see that empirical coverage probabilities are sensitive to the choice of the order of power series terms. However, we find in these sample settings, tests with two series terms perform well, which is consistent with the result using "bottom-up" method selection. With the number of series terms increasing, the tests expose distortions, which might be attributed to testing for additional coefficients for two functions. In addition, the table shows that the chi-squared test with Q performs better than the test with T.

Table 1: Probabilities of rejecting H_0

H_0 is true							
		Test with T			Test with Q		
Sample Size	Order of Power Series	Nominal Size			Nominal Size		
		0.1	0.05	0.01	0.1	0.05	0.01
100	2	0.1102	0.0769	0.0373	0.1084	0.0541	0.0116
	3	0.1728	0.1242	0.0664	0.1667	0.092	0.0248
	4	0.2072	0.1506	0.0832	0.1984	0.1159	0.0361
	bottom-up (2.0121)	0.1242	0.0891	0.0462	0.1216	0.0655	0.0152
500	2	0.0884	0.0587	0.0285	0.0859	0.0398	0.0068
	3	0.1454	0.1046	0.0531	0.1405	0.0764	0.0187
	4	0.1541	0.1095	0.0544	0.1468	0.0809	0.02
	bottom-up (2.0123)	0.1021	0.0679	0.0353	0.0998	0.0485	0.0098
1000	2	0.0902	0.0626	0.0315	0.0884	0.0428	0.0071
	3	0.1683	0.1236	0.0739	0.1634	0.0969	0.032
	4	0.1843	0.1349	0.0758	0.1756	0.1038	0.0318
	bottom-up (2.0132)	0.1055	0.0776	0.0398	0.1037	0.0536	0.0114

Table 2: Probabilities of rejecting H_0

H_1 is true							
Sample Size	Order of Power Series	Test with T Nominal Size			Test with Q Nominal Size		
		0.1	0.05	0.01	0.1	0.05	0.01
100	2	0.983	0.9759	0.951	0.9826	0.9655	0.8652
	3	0.9977	0.9952	0.9882	0.9975	0.9928	0.9577
	4	0.9968	0.9947	0.9866	0.9967	0.9921	0.95
500	2	0.9995	0.9994	0.9994	0.9994	0.9994	0.9992
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1000	2	0.9999	0.9998	0.9998	0.9999	0.9998	0.9997
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

We also conduct the experiment under the alternative hypothesis. In this experiment, we generate the data with

$$\sigma_0^2(x) = e^x, \quad \sigma_1^2(x) = x^2$$

and other variables similar to the preceding processes. Table 2 presents the results of this experiment, which demonstrates the power of our tests of the null hypothesis $\sigma_1^2(x) - \sigma_0^2(x) = 0$ for all $x \in X$. In this case, it is also reported that our test is sensitive to the choice of the order of power series terms, especially in small sample sizes. On the other hand, our test is quite powerful, especially in large sample sizes.

7 Empirical Application

We apply the test approach proposed in this paper to the wage data with union workers and nonunion workers to study the effect of unionism on wage dispersion. We first discuss the literature of the relationship between unionism and wage and then describe the data in this application. Then we present the results of tests concerning the effect of unions on the dispersion of wages.

Social scientists have long been struggling to give an answer to how unions affect the distribution of wages. In the past, most researchers accepted the view that unions tended to increase wage inequality, for example, a survey written by Johnson (1975). Freeman's important study led a renewed viewpoint to investigate the relationship between unionism and inequality. Analyzing cross-sectional micro data on workers in union and nonunion sectors, Freeman (1980) found that unionism reduces white-collar/blue-collar wage differentials in the organized sector, which overwhelms the increase in dispersion of wages across industries. Then studies were conducted using variants of a framework to illustrate the effect of unions more completely by considering the union coverage rates (DiNardo et al., 1996), union effect across different types of workers (Card, 2001), and unobserved skill differences (Lemieux, 1993). Card (2001) divides the workers into 10 equally sized skill groups in the nonunion sector to estimate the overall effect of unions on the variance of wage relative to the situation that would be observed if all workers were paid according to the existing wage structure in the nonunion. He argued that the role of unions in flattening the wage dispersion for high-skilled workers is just slightly stronger than that for low-skilled workers.

Researchers may be interested in whether the wage dispersion gap between union and nonunion workers exists in some subpopulations with covariates beyond particular skill characteristics. In the following, we consider an empirical application regarding the effects of unionism on dispersion of wages using US data. We attempt to analyze whether there is significant evidence that unionism in the United States differs the dispersion of wages between union workers and nonunion workers with some characteristics.

We analyze data from the National Longitudinal Survey (Youth Sample) containing full-time working males who have completed their schooling by 1980, were then followed over the period 1981 to 1987, and provided sufficient information. The data set is an excerpt from Vella et al. (1998) and the data are obtained from supplemental content for Wooldridge (2010). The sample consists of 411 union workers and 134 nonunion workers. We test for zero change of conditional wage variances, where we condition on measures of workers' background characteristics, including years of schooling, years of working after school and its square, annual hours worked, married status, ethnicity (Hispanic, black), health and living region (rural area, Northeast, Northern Central, South), as well as industry dummies (agricultural, mining, construction, trading, transportation, finance, business and repair Service, manufacturing, professional and related service, public administration) and occupational dummies (professional, technical and kindred; managers, officials and proprietors; sales workers; clerical and kindred; craftsmen, foremen and kindred; operatives and kindred; service workers).

Like Crump et al. (2008), we conduct two specifications of the test with covariates selected by "top-down" and "bottom-up" methods. For "top down" method, we start with the full covariates and drop the covariates one by one with the smallest t-statistic until the t-statistic of all remaining covariates are not smaller than 2 (in absolute value). For the "bottom up" method, we regress with one intercept and one covariate and then select the covariate with the highest t-statistics. With one intercept, this selected covariate and one of all remaining covariates, we select again the one that has the highest t-statistics. Repeat this procedure until all remaining covariates have t-statistics smaller than 2 in absolute value. In both of the two specifications, we specify both mean and variance functions by selecting the covariates using the nonunion group and applying the same specification to the union group. In the first step, we specify the mean function by either the "top-down" or "bottom-up" method. The number of selected covariates in the first step corresponds to K_1. With these K_1 covariates, we calculate residuals for both union and nonunion groups and then in the second step, we select the covariates again, by employing the same "top-down" or "bottom-up" method for the variance function regression. It results in K_2 covariates in this step. .

The results for these four versions of the tests are reported in Table 3.[1] We also provide the zero conditional average effect of unionism on wages which is proposed by Crump et al. (2008) as well as our test for zero effect of union on conditional variance. Both Q statistics for the chi-squared test and T statistic for the normal distribution test and their p-values are recorded in the table. Comparing the results in the top and bottom panels of this table, we see that the results are robust to the variables selection procedure. The null hypothesis of the zero conditional average treatment effect is rejected at 1% level for each year, while for conditional variance, there

[1]In the test of zero average treatment effect for 1983 data using "bottom-up" procedure, we have an occupational indicator for managers, officials, and proprietors remaining in the final regression because in other years, these variables are always significant to the regression.

Table 3: Tests for zero conditional average treatment effects and zero conditional variance effects

	Test for conditional mean					Test for conditional variance				
year	Qmean	(K_1)	Qmepval	Tmean	Tmepval	Qvar	(K_2)	Qvarpval	Tvar	Tvarpval
					top-down					
1980	28.04	(5)	0.00	7.29	0.00	15.43	(4)	0.00	4.04	0.00
1981	23.87	(10)	0.01	3.10	0.00	4.01	(4)	0.40	0.00	0.50
1982	99.61	(15)	0.00	15.45	0.00	8.86	(2)	0.01	3.43	0.00
1983	42.86	(13)	0.00	5.86	0.00	8.15	(3)	0.04	2.10	0.02
1984	27.76	(10)	0.00	3.97	0.00	2.59	(3)	0.46	-0.17	0.57
1985	35.33	(12)	0.00	4.76	0.00	13.08	(3)	0.00	4.11	0.00
1986	74.91	(13)	0.00	12.14	0.00	6.51	(2)	0.04	2.26	0.01
1987	65.62	(20)	0.00	7.21	0.00	71.91	(8)	0.00	15.98	0.00
					bottom-up					
1980	36.37	(7)	0.00	7.85	0.00	3.07	(2)	0.20	0.54	0.30
1981	20.33	(7)	0.00	3.56	0.00	3.86	(4)	0.43	-0.05	0.52
1982	61.83	(10)	0.00	11.59	0.00	5.29	(2)	0.07	1.64	0.05
1983	42.66	(10)	0.00	7.30	0.00	8.20	(3)	0.04	2.12	0.02
1984	27.76	(10)	0.00	3.97	0.00	2.13	(2)	0.34	0.07	0.47
1985	42.81	(12)	0.00	6.29	0.00	10.70	(3)	0.01	3.14	0.00
1986	29.36	(9)	0.00	4.80	0.00	3.90	(2)	0.14	0.95	0.17
1987	43.26	(13)	0.00	5.93	0.00	45.42	(5)	0.00	12.78	0.00

is not always significant evidence against the null hypothesis. That is, in some years, conditioned on some subpopulation, unionism might affect the inequality of workers' wages, but in the other years (1981 and 1984), there is no statistical evidence that unionism has changed the inequality of their wage for any subpopulation. The results for year s 198-1983 and 1985-1987 are consistent with the conclusion argued by Card et al. (2004) that unions reduce the variance of wages for men.

8 Conclusion

In this paper, we developed nonparametric tests for the null hypothesis that the conditional variances of the outcome for the treatment group and control group are identical for all subpopulations defined by covariates. We gave test statistic T, which has a standard normal distribution in large samples under the null. In practice, we tested with Q, which performs better. Applying the working males wage data during the period from 1980 to 1987, we find unionism has no treatment effect heterogeneity on wages between union and nonunion groups. However, in some years, unionism tends to lead to differences of inequality in some groups of workers.

Appendix

We work with a normalized version of the parameters for convenience. Given assumption 5.1, $\Omega_{P,1,K} = E[P_K(X)P_K(X)'|W = 1]$ is nonsingular for all K (Newey, 1994). We can construct a sequence of basis functions $R_{K_1}(x) = \Omega_{P,1,K_1}^{-1/2} P_{K_1}(x)$ with $E[R_{K_1}(X)R_{K_1}(X)'|W = 1] = I_{K_1}$ and $R_{K_2}(x) = \Omega_{P,1,K_2}^{-1/2} P_{K_2}(x)$ with $E[R_{K_2}(X)R_{K_2}(X)'|W = 1] = I_{K_2}$. Below we prove the theorem by the sequence of basis functions, $R_{K_r}(x)$, instead of $P_{K_r}(x)$, where $r = 1, 2$. This replacement will not affect the estimators.

Now, we give some notations described in Section 4 when the basis functions $R_{K_r}(x)$ are used. Define $N_w \times K_r$ matrix $R_{1,K_r} = (R_{K_r}(X_1)', \ldots, R_{K_r}(X_{N_1})')$ and $R_{0,K_r} = (R_{K_r}(X_{N_1+1})', \ldots, R_{K_r}(X_N)')$. Then, nonparametric series estimator of $\mu_w(x)$, given K_1 terms in the series, is given by

$$\begin{aligned}
\hat{\mu}_{w,K_1}(x) &= R_{K_1}(x)'(R'_{w,K_1} R_{w,K_1})^{-}(R'_{w,K_1}\mathbf{Y}_w) \\
&= R_{K_1}(x)'\hat{\gamma}_{w,K_1},
\end{aligned}$$

where $\hat{\gamma}_{w,K_1} = (R'_{w,K_1} R_{w,K_1})^{-}(R'_{w,K_1}\mathbf{Y}_w)$. Then nonparametric series estimator of $\sigma^2_w(x)$, given K_2 terms in the series, is given by

$$\begin{aligned}
\hat{\sigma}^2_{w,K_2}(x) &= R_{K_2}(x)'(R'_{w,K_2} R_{w,K_2})^{-}(R'_{w,K_2}\hat{\boldsymbol{\epsilon}}^2_w) \\
&= R_{K_2}(x)'\hat{\alpha}_{w,K_2},
\end{aligned}$$

where $\hat{\alpha}_{w,K_2} = (R'_{w,K_2} R_{w,K_2})^{-}(R'_{w,K_2}\hat{\boldsymbol{\epsilon}}^2_w)$.

In addition,

$$\begin{aligned}
&\Omega_{w,K_r} \equiv E[R_{K_r}(X)R_{K_r}(X)'|W = w], \\
&\Psi_{w,K_1} \equiv E[\sigma^2_w(X)R_{K_1}(X)R_{K_1}(X)'|W = w] \text{ and} \\
&\Psi_{w,K_2} \equiv E[\theta^2_w(X)R_{K_2}(X)R_{K_2}(X)'|W = w],
\end{aligned}$$

and we estimate them by

$$\hat{\Omega}_{w,K_r} = \frac{R'_{w,K_r}R_{w,K_r}}{N_w}, \quad \hat{\Psi}_{w,K_1} = \frac{R'_{w,K_1}\hat{D}_{w,K_1}R_{w,K_1}}{N_w}, \quad \hat{\Psi}_{w,K_2} = \frac{R'_{w,K_2}\hat{M}_{w,K_2}R_{w,K_2}}{N_w}.$$

Finally, we have

$$N \cdot V = \frac{1}{\pi_0}\Omega^{-1}_{0,K_2}\Psi_{0,K_2}\Omega^{-1}_{0,K_2} + \frac{1}{\pi_1}\Omega^{-1}_{1,K_2}\Psi_{1,K_2}\Omega^{-1}_{1,K_2}$$

and estimator

$$\hat{V} = \frac{\hat{\Omega}^{-1}_{0,K_2}\hat{\Psi}_{0,K_2}\hat{\Omega}^{-1}_{0,K_2}}{N_0} + \frac{\Omega^{-1}_{1,K_2}\hat{\Psi}_{1,K_2}\hat{\Omega}^{-1}_{1,K_2}}{N_1}$$

where $\pi_w = \Pr(W = w)$. Note that $\Omega_{1,K_2} = I_{K_2}$. Moreover, we define $\zeta(K) = \sup_x \|R_K(x)\|$, where here and in the following, $\|\cdot\|$ denotes the Euclidean matrix norm, that is, for a matrix A, $\|A\| = tr(A'A)$. In this paper, we use orthonormal polynomials, and then $\zeta(K) = O(K)$ by Newey (1997). In addition, define C as a generic positive constant.

Let $\boldsymbol{\epsilon}_1^2 = ((Y_1 - \mu_1(X_1))^2, (Y_2 - \mu_1(X_2))^2, \ldots, (Y_{N_1} - \mu_1(X_{N_1}))^2)'$ and $\boldsymbol{\epsilon}_2^2 = ((Y_{N_1+1} - \mu_2(X_{N_1+1}))^2, (Y_{N_1+2} - \mu_2(X_{N_1+2}))^2, \ldots, (Y_N - \mu_2(X_N))^2)'$. Let $\boldsymbol{\epsilon}_w = \mathbf{Y}_w - \mu_w(\mathbf{X_w})$ and $\mathbf{u}_w = \boldsymbol{\epsilon}_w^2 - \sigma_w^2(\mathbf{X}_w)$.

First, we establish asymptotic normality for pseudo statistics. We define,

$$\begin{aligned}\alpha^*_{w,K_2} &\equiv (E[R_{K_2}(X)R_{K_2}(X)'|W=w])^{-1}E[R_{K_2}(X)\epsilon^2_{w,i}|W=w] \\ &= \Omega^{-1}_{w,K_2}E[R_{K_2}(X)\epsilon^2_{w,i}|W=w]\end{aligned}$$

Now we consider asymptotic normality of $\sqrt{N_w} \cdot \frac{1}{N_w}\Omega^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w$.

$$\sqrt{N_w} \cdot \frac{1}{N_w}\Omega^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w = \frac{1}{\sqrt{N_w}}\sum_{i=1}^{N}\Omega^{-1}_{w,K_2}\mathbf{1}(W_i = w)R_{K_2}(X_i)\mathbf{u}_{w,i}.$$

with

$$\begin{aligned}E[\Omega^{-1}_{w,K_2}R'_{w,K_2}u_w] &= \Omega^{-1}_{w,K_2}E[\mathbf{1}(W_i = w)R_{K_2}(X_i)E[u_{w,i}|X_i, W_i = w]] \\ &= \Omega^{-1}_{w,K_2}E[\mathbf{1}(W_i = w)R_{K_2}(X_i)E[u_{w,i}|X_i]] = 0,\end{aligned}$$

and

$$\begin{aligned}Var[\Omega^{-1}_{w,K_2}R'_{w,K}\mathbf{u}_w] &= \Omega^{-1}_{w,K_2}E[\mathbf{1}(W_i = w)u^2_{w,i}R_{K_2}(X_i)R_{K_2}(X_i)']\Omega^{-1}_{w,K_2} \\ &= \Omega^{-1}_{w,K_2}E[\mathbf{1}(W_i = w)^2R_{K_2}(X_i)R_{K_2}(X_i)'E[u^2_{w,i}|X_i, W_i = w]]\Omega^{-1}_{w,K_2} \\ &= \Omega^{-1}_{w,K_2}E[\mathbf{1}(W_i = w)\theta^2_w(X_i)R_K(X_i)R_K(X_i)']\Omega^{-1}_{w,K_2} \\ &= \Omega^{-1}_{w,K_2}E[\theta^2_w(X_i)R_K(X_i)R_K(X_i)'|W_i = w]\Omega^{-1}_{w,K_2} \cdot \Pr(W_i = w)\end{aligned}$$

$$= \Omega^{-1}_{w,K_2} E[\theta^2_w(X_i) R_K(X_i) R_K(X_i)'] \Omega^{-1}_{w,K_2} \cdot \pi_w$$
$$= \Omega^{-1}_{w,K_2} \Psi_{w,K_2} \Omega^{-1}_{w,K_2} \cdot \pi_w.$$

Therefore,

$$Var[\sqrt{N_w} \cdot \frac{1}{N_w} \Omega^{-1}_{w,K_2} R'_{w,K_2} \mathbf{u}_w] = \frac{1}{N_w} N \cdot \Omega^{-1}_{w,K_2} \Psi_{w,K_2} \Omega^{-1}_{w,K_2} \cdot \pi_w \to \Omega^{-1}_{w,K_2} \Psi_{w,K_2} \Omega^{-1}_{w,K_2}.$$

Se define

$$\begin{aligned} S_{w,K_2} &= \frac{1}{\sqrt{N_w}} \sum_{i=1}^{N} \Omega^{-1}_{w,K} \mathbf{1}(W_i = w) R_K(X_i) u_{w,i} \\ &= \frac{1}{\sqrt{N_w}} \sum_{i=1}^{N} [\Omega^{-1}_{w,K_2} \Psi_{w,K_2} \Omega^{-1}_{w,K_2} \cdot \pi_w]^{-1/2} \Omega^{-1}_{w,K_2} \mathbf{1}(W_i = w) R_{K_2}(X_i) u_{w,i} \\ &\equiv \frac{1}{\sqrt{N_w}} \sum_{i=1}^{N} Z_i. \end{aligned}$$

Then S_{w,K_2} is a normalized summation of N_w independent random vectors distributed with expectation 0 and variance-covariance matrix I_{K_2}. By theorem 1.1 in Bentkus (2005), we can appropriate the distribution of S_{w,K_2}, denoted by Q_{N_w}, with a multivariate standard Gaussian distribution.

Lemma 1. *: Suppose assumptions 2.1-2.3 and 5.1-5.3 hold. In particular, let $K_2(N) = N^{v_2}$ where $v_2 < 2/13$. Then,*

$$\sup_{A \in A_K} |Q_{N_w}(A) - \Phi(A)| = o(1),$$

where A_{K_2} is the class of all measurable convex sets in K_2-dimensional Euclidean space and Φ is a multivariate standard Gaussian distribution.

Proof. Theorem 1.1 in Bentkus (2005) shows

$$\sup_{A \in A_{K_2}} |Q_{N_w}(A) - \Phi(A)| \leq C \beta_3 K_2^{1/4}.$$

Consider,

$$\begin{aligned} \beta_3 &\equiv N_w^{-3/2} \sum_{i=1}^{N} E\|Z_i\|^3 \\ &= N_w^{-3/2} \sum_{i=1}^{N} E\|[\Omega^{-1}_{w,K} \Psi_{w,K} \Omega^{-1}_{w,K_2} \cdot \pi]^{-1/2} \mathbf{1}(W_i = w) R_K(X_i) u_{w,i}\|^3 \\ &\leq \underline{\theta}^2 \cdot N_w^{-3/2} \sum_{i=1}^{N} E\|[\Omega^{-1}_{w,KK_2} \cdot \pi]^{-1/2} \cdot \mathbf{1}(W_i = w) R_{K_2}(X_i) u_{w,i}\|^3 \end{aligned}$$

$$
\begin{aligned}
&\leq CN^{-3/2}\sum_{i=1}^{N} E\|\Omega_{w,K_2}^{-1/2}R_K(X_i)u_{w,i}\|^3 \\
&\leq CN^{-3/2}\sum_{i=1}^{N} E[\lambda_{max}(\Omega_{w,K_2}^{-1/2})\|R_{K_2}(X_i)u_{w,i}\|]^3 \\
&\leq CN^{-3/2}\sum_{i=1}^{N} \lambda_{max}(\Omega_{w,K_2}^{-1/2})^3\zeta(K_2)^3 E|u_{w,i}|^3 \\
&\leq C\zeta(K_2)^3 N^{-1/2}.
\end{aligned}
$$

Thus,

$$
C\beta_3 K_2^{1/4} = O(N^{-1/2}\zeta(K_2)^3 K_2^{1/4}) = O(N^{-1/2}K_2^{13/4}).
$$

Under assumption 5.3, $v_2 < 2/13$ holds, which leads to $O(N^{-1/2}K_2^{13/4})$. So, $\sup_{A\in A_{K_2}} |Q_{N_w}(A) - \Phi(A)| = o(1)$. $\square$

Now, we have a multivariate asymptotic normality, under which we may further proceed a univariate standard Gaussian distribution. We consider the quadratic form $S'_{w,K_2}S_{w,K_2}$,

$$
S'_{w,K_2}S_{w,K_2} = \sum_{j=1}^{K_2}\left(\frac{1}{\sqrt{N_w}}\sum_{i}^{N} Z_{ij}\right)^2,
$$

where Z_{ij} is the jth element of the vector Z_i. Thus, $S'_{w,K_2}S_{w,K_2}$ is a sum of K_2 uncorrelated, squared random variables with each random variable converging to a standard Gaussian distribution. Next lemma shows that this sum converges to a chi-squared random variable with K_2 degrees of freedom.

Lemma 2. *:Suppose assumptions 2.1-2.3 and 5.1-5.3 hold. Then,*

$$
\sup_c |\Pr(S'_{w,K_2}S_{w,K_2} \leq c) - \chi^2_{K_2}(c)| \to 0.
$$

Proof. Define the set $A(c) \equiv \{S \in \mathbb{R}^{K_2} | S'S \leq c\}$, and it is a measurable, convex set in $\mathbb{R}^{K_2}$. For $Z \sim N(0, I_{K_2})$,

$$
\begin{aligned}
&\sup_c |\Pr(S'_{w,K_2}S_{w,K_2} \leq c) - \chi^2_{K_2}(c)| \\
=&\sup_c |\Pr(S'_{w,K_2}S_{w,K_2} \leq c) - \Pr(Z'Z \leq c)| \\
=&\sup_c |\Pr(S_{w,K_2} \in A(c)) - \Pr(Z'Z \leq c)| \\
\leq& \sup_{A\in A_{K_2}} |Q_{N_w}(A) - \Phi(A)|
\end{aligned}
$$

$$=o(1).$$

□

The normalized version of $S'_{w,K_2}S_{w,K_2}$ converges to a standard Gaussian distribution by the following lemma.

Lemma 3. *Suppose assumptions 2.1-2.3 and 5.1-5.3 hold. Then,*

$$\sup_c |\Pr(\frac{S'_{w,K_2}S_{w,K_2}-K_2}{\sqrt{2K_2}} \leq c) - \Phi(c)| \to 0.$$

Proof.

$$\begin{aligned}
&\sup_c |\Pr(\frac{S'_{w,K_2}S_{w,K_2}-K_2}{\sqrt{2K_2}} \leq c) - \Phi(c)| \\
=&\sup_c |\Pr(S'_{w,K_2}S_{w,K_2} \leq K + c\sqrt{2K_2}) - \Phi(c)| \\
\leq&\sup_c |\Pr(S'_{w,K_2}S_{w,K} \leq K_2 + c\sqrt{2K_2}) - \chi^2_{K_2}(K_2 + c\sqrt{2K})| \\
&+\sup_c |\chi^2(K_2 + c\sqrt{2K_2}) - \Phi(c)| \\
=&\sup_c |\Pr(S'_{w,K_2}S_{w,K_2} \leq K_2 + c\sqrt{2K_2}) - \chi^2_{K_2}(K_2 + c\sqrt{2K_2}) \\
&+\sup_c |\Pr(\frac{Z'Z-K_2}{\sqrt{2K_2}} \leq c) - \Phi(c)|,
\end{aligned}$$

where $Z \sim N(0, I_{K_2})$. The first term goes to zero by lemma 2. The second term is of order $O(K_2^{-1/2})$ by the Berry-Esseen Theorem, and for $v_2 < 0$, it also converges to zero. So the result holds. □

Before we go ahead with test statistic T, we show a couple of preliminary lemmas.

Lemma 4. *Suppose assumptions 2.1-2.3 and 5.1-5.3 hold. Then, for $r = 1, 2$,*

(i) $\|\hat{\Omega}_{w,K_r} - \Omega_{w,K_r}\| = O_p(\zeta(K_r)K_r^{1/2}N^{-1/2})$,

(ii) the eigenvalues of Ω_{w,K_r} are bounded and bounded away from zero,

(iii) the eigenvalues of $\hat{\Omega}_{w,K_r}$ are bounded and bounded away from zero in probability.

Proof. When $r = 1$, the proofs can be found in Crump et al. (2008) Lemma A.1. Similarly, we can prove the statements for $r = 2$. Note that under assumption 5.3, $O_p(\zeta(K_2)K_2^{1/2}N^{-1/2}) = o_p(1)$. □

Lemma 5. *Suppose assumptions 2.1-2.3 and 5.1-5.3 hold. Then,*

(i) *there is a sequence of vector* γ^0_{w,K_1}*, such that*

$\sup_x |\mu_w(x) - R_{K_1}(x)'\gamma^0_{w,K_1}| \equiv \sup_x |\mu_w(x) - \mu^0_{w,K_1}(x)| = O(K_1^{-s_1/d})$,

(ii) $\|\gamma^*_{w,K_1} - \gamma^0_{w,K_1}\| = O(K_1^{1/2}K_1^{-s_1/d})$,

(iii) $\sup_x |R_{K_1}(x)'\gamma^*_{w,K_1} - R_{K_1}(x)'\gamma^0_{w,K_1}| = O(\zeta(K_1)K_1^{1/2}K_1^{-s_1/d})$,

(iv) $\|\hat{\gamma}_{w,K_1} - \gamma^0_{w,K_1}\| = O(K_1^{-s_1/d}) + O_p(K_1^{1/2}N^{-1/2})$,

(v) $\sup_x |\mu_w(x) - R_{K_1}(x)'\hat{\gamma}_{w,K_1}| \equiv \sup_x |\mu_w(x) - \hat{\mu}_{w,K_1}(x)| = O(\zeta(K_1)K_1^{-s_1/d}) + O_p(\zeta(K_1)K_1^{1/2}N^{-1})$.

Proof. The proofs can be found in Crump et al. (2008) Lemma A.6. Because $O_p(\zeta(K_r)K_r^{1/2}N^{-1/2}) = o_p(1)$ under assumptions 5.1-5.3, we can have (iv) and (v) from Crump et al. (2007). □

Lemma 6. *Suppose assumptions 2.1-2.3 and 5.1-5.3 hold. Then,*

(i) $\sup_x |\sigma^2_w(x) - R_{K_2}(x)'\alpha^0_{w,K_2}| \equiv \sup_x |\sigma^2_w(x) - \sigma^{2^0}_{w,K_2}(x)| = O(K_2^{-s_2/d})$,

(ii) $\|\alpha^*_{w,K_2} - \alpha^0_{w,K_2}\| = O(K_2^{1/2}K_2^{-s_2/d})$,

(iii) $\sup_x |R_{K_2}(x)'\alpha^*_{w,K_2} - R_{K_2}(x)'\alpha^0_{w,K_2}| = O(\zeta(K_2)K_2^{1/2}K_2^{-s_2/d})$,

(iv) $\|\hat{\alpha}_{w,K_2} - \alpha^0_{w,K_2}\| = O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2}N^{-1}) + O_p(K_2^{1/2}N^{-1/2}) + O(K_2^{-s_2/d})$,

(v) $\sup_x |\sigma^2_w(x) - R_{K_2}(x)'\hat{\alpha}_{w,K_2}| \equiv \sup_x |\sigma^2_w(x) - \hat{\sigma}^2_{w,K_2}(x)|$
$= O_p(\zeta(K_1)\zeta(K_2)K_1^{-s_1/d} + \zeta(K_1)\zeta(K_2)K_1^{1/2}N^{-1} + \zeta(K_2)K_2^{1/2}N^{-1/2} + \zeta(K_2)K_2^{-s_2/d})$,

Proof. We can prove (i), (ii) and (iii) as Lemma A.2 in Crump et al. (2008). For (iv), for $M \in (0, \infty)$

$$\begin{aligned}
&Pr(|\hat{\epsilon}^2_{w,i} - \epsilon^2_{w,i}| > M) \\
\leq&\frac{E|\hat{\epsilon}^2_{w,i} - \epsilon^2_{w,i}|}{M} \\
=&\frac{E|(\epsilon_{w,i} + \mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i))^2 - \epsilon^2_{w,i}|}{M} \\
\leq&\frac{E|2\epsilon_{w,i}(\mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i))|}{M} + \frac{E|\mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i)|^2}{M} \\
\leq&2 \cdot \sup_i |\mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i)| \cdot \frac{1}{M}E|\epsilon_{w,i}| + (\sup_i |\mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i)|)^2 \\
\leq&C \cdot \sup_i |\mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i)| + (O(\zeta(K_1)K_1^{-s_1/d}) + O_p(\zeta(K_1)K_1^{1/2}N^{-1}))^2,
\end{aligned}$$

$$=O(\zeta(K_1)K_1^{-s_1/d}) + O_p(\zeta(K_1)K_1^{1/2}N^{-1}).$$

The last line holds because $E|\epsilon_{w,i}|$ is bounded and bounded away from 0 from assumption 5.2. So $|\hat{\epsilon}^2_{w,i} - \epsilon^2_{w,i}| = O(\zeta(K_1)K_1^{-s_1/d}) + O_p(\zeta(K_1)K_1^{1/2}N^{-1})$. In addition, we can know $\sup_i |\hat{\epsilon}^2_{w,i} - \epsilon^2_{w,i}| = O(\zeta(K_1)K_1^{-s_1/d}) + O_p(\zeta(K_1)K_1^{1/2}N^{-1})$.

$$\begin{aligned}
\|\hat{\alpha}_{w,K_2} - \alpha^0_{w,K_2}\| &= \|\frac{1}{N_w}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\hat{\boldsymbol{\epsilon}}^2_w - \frac{1}{N_w}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}R_{w,K_2}\alpha^0_{w,K_2})\| \\
&= \|\frac{1}{N_w}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}(\hat{\boldsymbol{\epsilon}}^2_w - R_{w,K_2}\alpha^0_{w,K_2})\| \\
&\leq \lambda_{\max}(\hat{\Omega}^{-1/2}_{w,K_2}) \cdot \|\frac{1}{N_w}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}(\hat{\boldsymbol{\epsilon}}^2_w - R_{w,K_2}\alpha^0_{w,K_2})\|
\end{aligned}$$

Note that $\lambda_{\max}(\hat{\Omega}^{-1/2}_{w,K_2}) = \lambda_{\max}(\Omega^{-1/2}_{w,K_2}) + O_p(\zeta(K_2)K_2^{1/2}N^{-1/2}) = O(1) + o_p(1)$ [2].

$$\begin{aligned}
&\|\frac{1}{N_w}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}(\hat{\boldsymbol{\epsilon}}^2_w - R_{w,K_2}\alpha^0_{w,K_2})\| \\
=&\|\frac{1}{N_w}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}(\hat{\boldsymbol{\epsilon}}^2_w - \boldsymbol{\epsilon}^2_w + \boldsymbol{\epsilon}^2_w - \sigma^2_w(\mathbf{X}_w) + \sigma^2_w(\mathbf{X}_w) - R_{w,K_2}\alpha^0_{w,K_2})\| \\
\leq&\|\frac{1}{N_w}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}(\hat{\boldsymbol{\epsilon}}^2_w - \boldsymbol{\epsilon}^2_w)\| && \text{(A.1)} \\
+&\|\frac{1}{N_w}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}\mathbf{u}_w\| && \text{(A.2)} \\
+&\|\frac{1}{N_w}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}(\sigma^2_w(\mathbf{X}_w) - R_{w,K_2}\alpha^0_{w,K_2})\|, && \text{(A.3)}
\end{aligned}$$

For (A.1), we have,

$$\begin{aligned}
&E\|\frac{1}{N_w}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}(\hat{\boldsymbol{\epsilon}}^2_w - \boldsymbol{\epsilon}^2_w)\|^2 \\
=&\frac{1}{N_w}E[(\hat{\boldsymbol{\epsilon}}^2_w - \boldsymbol{\epsilon}^2_w)'R_{w,K_2}(R'_{w,K_2}R_{w,K_2})^{-1}R'_{w,K_2}(\hat{\boldsymbol{\epsilon}}^2_w - \boldsymbol{\epsilon}^2_w)] \\
\leq&\frac{1}{N_w}E[(\hat{\boldsymbol{\epsilon}}^2_w - \boldsymbol{\epsilon}^2_w)'(\hat{\boldsymbol{\epsilon}}^2_w - \boldsymbol{\epsilon}^2_w)] \\
\leq&E|\hat{\epsilon}^2_{w,i} - \epsilon^2_{w,i}|^2 \\
=&O(\zeta(K_1)^2K_1^{-2s_1/d}) + O_p(\zeta(K_1)^2K_1N^{-2}).
\end{aligned}$$

The third line follows by the fact that $I - R_{w,K_2}(R'_{w,K_2}R_{w,K_2})^{-1}R'_{w,K_2}$ is a positive semi-definite. So, $\|\frac{1}{N_w}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}(\hat{\boldsymbol{\epsilon}}^2_w - \boldsymbol{\epsilon}^2_w)\| = O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2}N^{-1})$.

For (A.2), we have,

$$E\|\frac{1}{N_w}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}\mathbf{u}_w\|^2$$

[2] $\lambda_{\max}(\hat{\Omega}^{-1}_{w,K}) = \lambda_{\max}(\Omega^{-1}_{w,K}) + O_p(\zeta(K)K^{1/2}N^{-1/2})$

$$
\begin{aligned}
&=E[tr(\frac{1}{N_w}u_w' R_{w,K_2}\hat{\Omega}_{w,K_2}^{-1}R_{w,K_2}\mathbf{u}_w)]\\
&=\frac{1}{N_w}E[tr(R_{w,K_2}(R_{w,K_2}'R_{w,K_2})^{-1}R_{w,K_2}'\mathbf{u}_w\mathbf{u}_w')]\\
&=\frac{1}{N_w}tr(E[R_{w,K_2}(R_{w,K_2}'R_{w,K_2})^{-1}R_{w,K_2}'E[\mathbf{u}_w\mathbf{u}_w'|X]])\\
&\leq\bar{\theta}^2\cdot\frac{1}{N_w}E[tr(R_{w,K_2}(R_{w,K_2}'R_{w,K_2})^{-1}R_{w,K_2}')]\\
&=\bar{\theta}^2\cdot\frac{1}{N_w}K_2\\
&\leq CK_2N^{-1},
\end{aligned}
$$

and so $\|\frac{1}{N_w}\hat{\Omega}_{w,K_2}^{-1/2}R_{w,K_2}'\mathbf{u}_w\| = O_p(K_2^{1/2}N^{-1/2})$.

For (A.3), we have,

$$
\begin{aligned}
&\|\frac{1}{N_w}\hat{\Omega}_{w,K_2}^{-1/2}R_{w,K_2}'(\sigma_w^2(\mathbf{X}_w)-R_{w,K_2}\alpha_{w,K_2}^0)\|^2\\
=&\frac{1}{N_w}(\sigma_w^2(\mathbf{X}_w)-R_{w,K_2}\alpha_{w,K_2}^0)'R_{w,K_2}(R_{w,K_2}'R_{w,K_2})^{-1}R_{w,K_2}'(\sigma_w^2(\mathbf{X}_w)-R_{w,K_2}\alpha_{w,K_2}^0)\\
\leq&\frac{1}{N_w}(\sigma_w^2(\mathbf{X}_w)-R_{w,K_2}\alpha_{w,K_2}^0)'(\sigma_w^2(\mathbf{X}_w)-R_{w,K_2}\alpha_{w,K_2}^0)\\
\leq&(\sup_x|\sigma_w^2(x)-R_{w,K_2}\alpha_{w,K_2}^0|)^2\\
\leq&CK_2^{-2s_2/d}
\end{aligned}
$$

by (i), and so $\|\frac{1}{N_w}\hat{\Omega}_{w,K_2}^{-1/2}R_{w,K_2}'(\sigma_w^2(\mathbf{X}_w)-R_{w,K_2}\alpha_{w,K_2}^0)\| = O(K_2^{-s_2/d})$. Combining these, we have,

$$
\begin{aligned}
&\|\hat{\alpha}_{w,K}-\alpha_{w,K}\|\\
=&[O(1)+o_p(0)]\cdot[O_p(\zeta(K_1)K_1^{-s_1/d}+\zeta(K_1)K_1^{1/2}N^{-1})+O_p(K_2^{1/2}N^{-1/2})+O(K_2^{-s_2/d})]\\
=&O_p(\zeta(K_1)K_1^{-s_1/d}+\zeta(K_1)K_1^{1/2}N^{-1})+O_p(K_2^{1/2}N^{-1/2})+O(K_2^{-s_2/d}).
\end{aligned}
$$

Finally, for (v), we have,

$$
\sup_x|\sigma_w^2(x)-\hat{\sigma}_{w,K_2}^2(x)|\leq\sup_x|\sigma_{w,K_2}^2(x)-\sigma_w^{2^0}(x)|+\sup_x|\sigma_{w,K_2}^{2^0}(x)-\hat{\sigma}_{w,K_2}^2(x)|.
$$

The first term is $O(K_2^{-s_2/d})$ by (i). For the second term, we have

$$
\begin{aligned}
&\sup_x|\sigma_w^{2^0}(x)-\hat{\sigma}_w^2(x)|\\
=&\sup_x|R_{K_2}(x)'(\alpha_{w,K_2}^0-\hat{\alpha}_{w,K_2})|\\
\leq&\sup_x\|R_{K_2}(x)\|\cdot\|\alpha_{w,K_2}^0-\hat{\alpha}_{w,K_2}\|
\end{aligned}
$$

$$=\zeta(K_2)[O_p(\zeta(K_1)K_1^{-s_1/d}+\zeta(K_1)K_1^{1/2}N^{-1})+O_p(K_2^{1/2}N^{-1/2})+O(K_2^{-s_2/d})],$$

where we use the result from (iv). Thus,

$$\begin{aligned}&\sup_x|\sigma_w^2(x)-\hat{\sigma}_{w,K_2}^2(x)|\\=&\zeta(K_2)[O_p(\zeta(K_1)K_1^{-s_1/d}+\zeta(K_1)K_1^{1/2}N^{-1})+O_p(K_2^{1/2}N^{-1/2})+O(K_2^{-s_2/d})]+O(K_2^{-s_2/d})\\=&O_p(\zeta(K_1)\zeta(K_2)K_1^{-s_1/d}+\zeta(K_1)\zeta(K_2)K_1^{1/2}N^{-1}+\zeta(K_2)K_2^{1/2}N^{-1/2}+\zeta(K_2)K_2^{-s_2/d}).\end{aligned}$$

□

Lemma 7. *Suppose assumptions 2.1-2.3 and 5.1-5.3 hold. Then,*

(i) $\|\hat{\Psi}_{w,K_2}-\Psi_{w,K_2}\|=O_p(\zeta(K_1)^2\zeta(K_2)^2K_1^{-2s_1/d}K_2+\zeta(K_2)^2K_2K_2^{-s_2/d}$
$+\zeta(K_1)^2\zeta(K_2)^2K_1^{1/2}K_1^{-s_1/d}K_2N^{-1}+\zeta(K_1)\zeta(K_2)^2K_1^{-s_1/d}K_2^{3/2}N^{-1/2}$
$+\zeta(K_1)\zeta(K_2)^2K_1^{1/2}K_2^{3/2}N^{-3/2}+\zeta(K_2)^2K_2^{1/2}N^{-1/2})$,

(ii) the eigenvalues of Ψ_{w,K_2} are bounded and bounded away from zero,

(iii) the eigenvalues of $\hat{\Psi}_{w,K_2}$ are bounded and bounded away from zero in probability,

(iv) the eigenvalues of $N\cdot V$ are bounded and bounded away from zero,

(v) the eigenvalues of $N\cdot\hat{V}$ are abounded and bounded away from zero in probability.

Proof. Let us first define,

$$\tilde{\Psi}_{w,K_2}=\frac{R'_{w,K_2}\tilde{M}_{w,K_2}R_{w,K_2}}{N_w},\quad\text{where }\tilde{M}_{w,K}=diag\{u_{w,i}^2,\text{ for }i\text{ with }W_i=w\}.$$

Then,

$$\begin{aligned}&E\|\hat{\Psi}_{w,K_2}-\Psi_{w,K_2}\|^2\\=&E\|\hat{\Psi}_{w,K_2}-\tilde{\Psi}_{w,K_2}+\tilde{\Psi}_{w,K_2}-\Psi_{w,K_2}\|^2\\\le&2E\|\hat{\Psi}_{w,K_2}-\tilde{\Psi}_{w,K_2}\|^2+2E\|\tilde{\Psi}_{w,K_2}-\Psi_{w,K_2}\|^2\\=&2E\|\frac{R'_{w,K_2}(\hat{M}_{w,K_2}-\tilde{M}_{w,K_2})R_{w,K_2}}{N_w}\|^2 &\text{(A.4)}\\&+2E\|\tilde{\Psi}_{w,K_2}-\Psi_{w,K_2}\|^2. &\text{(A.5)}\end{aligned}$$

Consider (A.4),

$$E\|\frac{R'_{w,K_2}(\hat{M}_{w,K_2}-\tilde{M}_{w,K_2})R_{w,K_2}}{N_w}\|^2$$

$$=\frac{1}{N_w^2}\sum_{k=1}^{K_2}\sum_{l=1}^{K_2}\sum_{i=1}^{N}\sum_{j=1}^{N}E[\mathbf{1}(W_i=w)\mathbf{1}(W_j=w)(\hat{u}_{w,i}^2-u_{w,i}^2)R_{kK_2}(X_i)R_{lK_2}(X_i)R_{kK_2}(X_j)R_{lK_2}(X_j)].$$

Since

$$\begin{aligned}\sup_i|\hat{u}_{w,i}-u_{w,i}| =&|\hat{\epsilon}_{w,i}^2(X_i)-\hat{\sigma}_w^2(X_i)-(\epsilon_w^2(X_i)-\sigma_w^2(X_i))|\\ =&\sup_i|\hat{\epsilon}_{w,i}^2-\epsilon_{w,i}^2|+\sup_i|\hat{\sigma}_{w,K_2}^2(X_i)-\sigma_w^2(X_i)|,\end{aligned}$$

$$\begin{aligned}E|\hat{u}_{w,i}^2-u_{w,i}^2| =&E|(\hat{u}_{w,i}-u_{w,i})^2+2u_{w,i}(\hat{u}_{w,i}-u_{w,i})|\\ \leq&E|u_{w,i}|\cdot\sup_i|\hat{u}_{w,i}-u_{w,i}|+o_p(\hat{u}_{w,i}-u_{w,i})\\ =&C(\sup_i|\hat{\epsilon}_{w,i}^2-\epsilon_{w,i}^2|+\sup_i|\hat{\sigma}_{w,K_2}^2(X_i)-\sigma_w^2(X_i)|)\\ =&O_p(\zeta(K_1)\zeta(K_2)K_1^{-s_1/d}+\zeta(K_1)\zeta(K_2)K_1^{1/2}N^{-1}+\zeta(K_2)K_2^{1/2}N^{-1/2}+\zeta(K_2)K_2^{-s_2/d}).\end{aligned}$$

We have

$$\begin{aligned}&E\|\frac{R_{w,K_2}'(\hat{M}_{w,K_2}-\tilde{M}_{w,K_2})R_{w,K_2}}{N_w}\|^2\\ =&[O_p(\zeta(K_1)\zeta(K_2)K_1^{-s_1/d}+\zeta(K_1)\zeta(K_2)K_1^{1/2}N^{-1}+\zeta(K_2)K_2^{1/2}N^{-1/2}+\zeta(K_2)K_2^{-s_2/d})]^2\\ &\cdot\frac{1}{N_w^2}\sum_{k=1}^{K_2}\sum_{l=1}^{K_2}\sum_{i=1}^{N}\sum_{j=1}^{N}E[R_{kK_2}(X_i)R_{lK_2}(X_i)R_{kK_2}(X_j)R_{lK_2}(X_j)].\end{aligned}$$

From Crump et al. (2007), we can know the last term is

$$\frac{1}{N_w^2}\sum_{k=1}^{K_2}\sum_{l=1}^{K_2}\sum_{i=1}^{N}\sum_{j=1}^{N}E[R_{kK_2}(X_i)R_{lK_2}(X_i)R_{kK_2}(X_j)R_{lK_2}(X_j)]=tr(E[\hat{\Omega}_{w,K_2}^2])=O(K_2)$$

So for (A.4), we have

$$\begin{aligned}&E\|\frac{R_{w,K_2}'(\hat{M}_{w,K_2}-\tilde{M}_{w,K_2})R_{w,K_2}}{N_w}\|^2\\ =&[O_p(\zeta(K_1)\zeta(K_2)K_1^{-s_1/d}+\zeta(K_1)\zeta(K_2)K_1^{1/2}N^{-1}+\zeta(K_2)K_2^{1/2}N^{-1/2}+\zeta(K_2)K_2^{-s_2/d})]^2O(K_2).\end{aligned} \tag{A.6}$$

Following the steps in the proof of Lemma A.1(iv) in Crump et al. (2007), we can show (A.5) is,

$$E\|\tilde{\Psi}_{w,K_2}-\Psi_{w,K_2}\|^2=O_p(\zeta(K_2)^2K_2N^{-1}). \tag{A.7}$$

Combining (A.6) and (A.16) yields

$$\|\hat{\Psi}_{w,K_2}-\Psi_{w,K_2}\|$$

$$
\begin{aligned}
=&[O_p(\zeta(K_1)\zeta(K_2)K_1^{-s_1/d}+\zeta(K_1)\zeta(K_2)K_1^{1/2}N^{-1}+\zeta(K_2)K_2^{1/2}N^{-1/2}+\zeta(K_2)K_2^{-s_2/d})]^2O(K_2)\\
&+O_p(\zeta(K_2)^2K_2^{1/2}N^{-1/2})\\
=&O_p(\zeta(K_1)^2\zeta(K_2)^2K_1^{-2s_1/d}K_2+\zeta(K_2)^2K_2K_2^{-s_2/d}+\zeta(K_1)^2\zeta(K_2)^2K_1^{1/2}K_1^{-s_1/d}K_2N^{-1}\\
&+\zeta(K_1)\zeta(K_2)^2K_1^{-s_1/d}K_2^{3/2}N^{-1/2}+\zeta(K_1)\zeta(K_2)^2K_1^{1/2}K_2^{3/2}N^{-3/2}+\zeta(K_2)^2K_2^{1/2}N^{-1/2}).
\end{aligned}
$$

We can prove (ii) as Crump et al. (2008) Lemma A.1 (v).

For (iii),

$$
\begin{aligned}
&\lambda_{\min}(\hat{\Psi}_{w,K_2})=\min_{d'd=1}d'\hat{\Psi}_{w,K_2}d\\
=&\min_{d'd=1}[d'\Psi_{w,K_2}d+d'(\Psi_{w,K_2}-\hat{\Psi}_{w,K_2})d]\\
\geq&\min_{d_1'd_1=1}d_1'\Psi_{w,K_2}d_1+\min_{d_2'd_2=1}d_2'(\Psi_{w,K_2}-\hat{\Psi}_{w,K_2})d_2\\
=&\lambda_{\min}(\Psi_{w,K_2})+\lambda_{\min}(\Psi_{w,K_2}-\hat{\Psi}_{w,K_2})\\
\geq&\lambda_{\min}(\Psi_{w,K_2})-\|\hat{\Psi}_{w,K_2}-\Psi_{w,K_2}\|\\
=&C-O_p(\zeta(K_1)^2\zeta(K_2)^2K_1^{-2s_1/d}K_2+\zeta(K_2)^2K_2K_2^{-s_2/d}+\zeta(K_1)^2\zeta(K_2)^2K_1^{1/2}K_1^{-s_1/d}K_2N^{-1}\\
&+\zeta(K_1)\zeta(K_2)^2K_1^{-s_1/d}K_2^{3/2}N^{-1/2}+\zeta(K_1)\zeta(K_2)^2K_1^{1/2}K_2^{3/2}N^{-3/2}+\zeta(K_2)^2K_2^{1/2}N^{-1/2})\\
=&C+o_p(1).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\lambda_{\max}(\hat{\Psi}_{w,K_2})\\
=&O(1)+O_p(\zeta(K_1)^2\zeta(K_2)^2K_1^{-2s_1/d}K_2+\zeta(K_2)^2K_2K_2^{-s_2/d}+\zeta(K_1)^2\zeta(K_2)^2K_1^{1/2}K_1^{-s_1/d}K_2N^{-1}\\
&+\zeta(K_1)\zeta(K_2)^2K_1^{-s_1/d}K_2^{3/2}N^{-1/2}+\zeta(K_1)\zeta(K_2)^2K_1^{1/2}K_2^{3/2}N^{-3/2}+\zeta(K_2)^2K_2^{1/2}N^{-1/2})\\
=&O(1)+o_p(1).
\end{aligned}
$$

We can prove (iv) as Lemma A.2 in Crump et al. (2008). For (v),

$$
\begin{aligned}
&\lambda_{\min}(\hat{\Omega}_{w,K_2}^{-1}\hat{\Psi}_{w,K_2}\hat{\Omega}_{w,K_2}^{-1})\\
\geq&\lambda_{\min}(\Psi_{w,K_2})\cdot\lambda_{\min}(\Omega_{w,K_2}^{-1})^2\\
\geq&[\lambda_{\min}(\Psi_{w,K_2})+o_p(1)]\lambda_{\min}(\Omega_{w,K_2}^{-1})^2\\
\geq&[C+o_p(1)][C+o_p(1)]^2\\
=&C+o_p(1).
\end{aligned}
$$

Thus,

$$\begin{aligned}\lambda_{\min}(N\cdot\hat{V}) &= \min_{d'd=1} d'(\frac{N}{N_0}\hat{\Omega}^{-1}_{0,K_2}\hat{\Psi}_{0,K_2}\hat{\Omega}^{-1}_{0,K_2} + \frac{N}{N_1}\hat{\Omega}^{-1}_{1,K_2}\hat{\Psi}_{1,K_2}\hat{\Omega}^{-1}_{1,K_2})d \\ &\geq \frac{N}{N_0}\min_{d'd=1} d'(\hat{\Omega}^{-1}_{0,K_2}\hat{\Psi}_{0,K_2}\hat{\Omega}^{-1}_{0,K_2})d + \frac{N}{N_1}\min_{d'd=1} d'(\hat{\Omega}^{-1}_{1,K_2}\hat{\Psi}_{1,K_2}\hat{\Omega}^{-1}_{1,K_2})d \\ &= \lambda_{\min}(\hat{\Omega}^{-1}_{0,K_2}\hat{\Psi}_{0,K_2}\hat{\Omega}^{-1}_{0,K_2}) + \lambda_{\min}(\hat{\Omega}^{-1}_{1,K_2}\hat{\Psi}_{1,K_2}\hat{\Omega}^{-1}_{1,K_2})\end{aligned}$$

is bounded away from zero in probability. Similarly, we can prove $\lambda_{\max}(N\cdot\hat{V})$ is bounded in probability. □

Lemma 8. *Suppose assumptions 2.1-2.3 and 5.1-5.3 hold. Then,*

$$\frac{N_w(\hat{\alpha}_{w,K_2}-\alpha^*_{w,K_2})'[\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1}(\hat{\alpha}_{w,K_2}-\alpha^*_{w,K_2})-K_2}{\sqrt{2K_2}} \xrightarrow{d} N(0,1).$$

Proof. Because Lemma 3 holds, we need only show that

$$\|[\hat{\Omega}^{-1}_{0,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\sqrt{N_w}(\hat{\alpha}_{w,K_2}-\alpha^*_{w,K_2}) - S_{w,K_2}\| = o_p(1). \tag{A.8}$$

Let $\mathbf{u}^*_{w,K_2} \equiv \hat{\boldsymbol{\epsilon}}^2_w - R_{w,K_2}\alpha^*_{w,K_2}$, then

$$\begin{aligned}\hat{\alpha}_{w,K_2}-\alpha^*_{w,K_2} &= (R'_{w,K_2}R_{w,K_2})^{-1}(R'_{w,K_2}(R_{w,K_2}\alpha^*_{w,K_2}+\mathbf{u}^*_w)) - \alpha^*_{w,K_2} \\ &= \frac{1}{N_w}\cdot\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}^*_w\end{aligned}$$

The left side of (A.8) can be written as

$$\begin{aligned}&\|[\hat{\Omega}^{-1}_{0,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\sqrt{N_w}(\hat{\alpha}_{w,K_2}-\alpha^*_{w,K_2}) - S_{w,K_2}\| \\ =&\frac{1}{\sqrt{N_w}}\|[\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K}R'_{w,K_2}\mathbf{u}^*_w - [\Omega^{-1}_{w,K}\Psi_{w,K_2}\Omega^{-1}_{w,K_2}]^{-1/2}\Omega^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w\| \\ \leq&\frac{1}{\sqrt{N_w}}\|[\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w - [\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w\| \quad \text{(A.9)} \\ +&\frac{1}{\sqrt{N_w}}\|[\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w - [\hat{\Omega}^{-1}_{w,K_2}\Psi_{w,K_2}\Omega^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w\| \\ &\text{(A.10)} \\ +&\frac{1}{\sqrt{N_w}}\|[\hat{\Omega}^{-1}_{w,K_2}\Psi_{w,K_2}\Omega^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w - [\Omega^{-1}_{w,K_2}\Psi_{w,K_2}\Omega^{-1}_{w,K_2}]^{-1/2}\Omega^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w\|. \\ &\text{(A.11)}\end{aligned}$$

For (A.9),

$$\frac{1}{\sqrt{N_w}}\|[\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}^*_w - [\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w\|$$

$$\leq \|[\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\| \cdot \|\frac{\hat{\Omega}^{-1/2}_{w,K_2} R'_{w,K_2}(\mathbf{u}^*_w - \mathbf{u}_w)}{\sqrt{N_w}}\|.$$

The first factor is

$$\begin{aligned}
&\|[\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\| \\
=&\|[\hat{\Psi}^{1/2}_{w,K_2}\hat{\Omega}^{-1/2}_{w,K_2}\| \\
\leq&\lambda_{\max}(\hat{\Psi}^{-1/2}_{w,K_2})\lambda_{\max}(\hat{\Omega}^{-1/2}_{w,K_2})K_2^{1/2} \\
\leq&([C + o_p(1)])^2 K_2^{1/2} \\
=&O_p(K_2^{1/2}).
\end{aligned}$$

For the second factor we have,

$$\begin{aligned}
&E\left\|\frac{\hat{\Omega}^{-1/2}_{w,K_2} R'_{w,K_2}(\mathbf{u}^*_w - \mathbf{u}_w)}{\sqrt{N_w}}\right\|^2 \\
=&E[\frac{1}{N_w} tr((\mathbf{u}^*_w - \mathbf{u}_w)' R_{w,K}\hat{\Omega}^{-1}_{w,K_2} R'_{w,K_2}(\mathbf{u}^*_w - \mathbf{u}_w))] \\
=&E[(\mathbf{u}^*_w - \mathbf{u}_w)' R_{w,K_2}(R'_{w,K_2}R_{w,K_2})^{-1}R_{w,K_2}(\mathbf{u}^*_w - \mathbf{u}_w)'] \\
=&E[(\mathbf{u}^*_w - \mathbf{u}_w)'(\mathbf{u}^*_w - \mathbf{u}_w)] \\
=&E[(\sigma^2_w(\mathbf{X}_w) - R_{w,K_2}\alpha^*_{w,K_2})'(\sigma^2(\mathbf{X}_w) - R_{w,K_2}\alpha^*_{w,K_2})] \\
\leq&N_w \sup_x |\sigma^2_w(x) - R_{K_2}(x)'\alpha^*_{w,K_2}|^2 \\
\leq&N_w(\sup_x |\sigma^2_w(x) - R_{K_2}(x)'\alpha^0_{w,K_2}| + \sup_x |R_{K_2}(x)'\alpha^0_{w,K_2} - R_{K_2}(x)'\alpha^*_{w,K_2}|)^2 \\
=&N_w(O(K_2^{-s_2/d})O(\zeta(K_2)K_2^{1/2}K_2^{-s_2/d}))^2 \\
=&O(N)(O(\zeta(K_2)K_2^{1/2}K_2^{-s_2/d}))^2.
\end{aligned}$$

Then, equation (A.9) is

$$\begin{aligned}
&\frac{1}{\sqrt{N_w}}\|[\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}^*_w - [\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w\| \\
=&O(K_2^{1/2})O_p(\zeta(K_2)K_2^{1/2}K_2^{-s_2/d}N^{1/2}) \\
=&O_p(\zeta(K_2)K_2K_2^{-s_2/d}N^{1/2}),
\end{aligned}$$

which is $o_p(1)$ under assumption 5.2 and 5.3. For (A.10),

$$\frac{1}{\sqrt{N_w}}\|[\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}\mathbf{u}_w - [\hat{\Omega}^{-1}_{w,K_2}\Psi_{w,K_2}\Omega^{-1}_{w,K}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K}u_w\|$$

$$\leq \|[\hat{\Psi}_{w,K_2}\Omega^{-1}_{w,K_2}]^{-1/2} - [\hat{\Psi}_{w,K_2}\Omega^{-1}_{w,K_2}]^{-1/2}\| \|N_w^{-1/2}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}u_w\|. \tag{A.12}$$

For the first factor in (A.12), we consider

$$\begin{aligned}
&\|\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2} - \Psi_{w,K_2}\Omega^{-1}_{w,K_2}\| \\
=&\|\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2} - \Psi_{w,K_2}\hat{\Omega}^{-1}_{w,K_2} + \Psi_{w,K_2}\hat{\Omega}^{-1}_{w,K_2} - \Psi_{w,K_2}\Omega^{-1}_{w,K_2}\| \\
\leq&\|\hat{\Psi}_{w,K_2} - \Psi_{w,K_2}\|\|\hat{\Omega}^{-1}_{w,K_2}\| + \|\Psi_{w,K_2}\|\|\hat{\Omega}^{-1}_{w,K_2} - \Omega^{-1}_{w,K_2}\| \\
\leq&O_p(\zeta(K_1)^2\zeta(K_2)^2K_1^{-2s_1/d}K_2 + \zeta(K_2)^2K_2K_2^{-s_2/d} + \zeta(K_1)^2\zeta(K_2)^2K_1^{1/2}K_1^{-s_1/d}K_2N^{-1} \\
&+ \zeta(K_1)\zeta(K_2)^2K_1^{-s_1/d}K_2^{3/2}N^{-1/2})[K_2^{1/2}\lambda_{\max}(\Omega^{-1}_{w,K_2})] + K_2^{1/2}\lambda_{\max}(\Psi_{w,K_2})\cdot O_p(\zeta(K_2)K_2^{1/2}N^{-1/2}) \\
=&O_p(\zeta(K_1)^2\zeta(K_2)^2K_1^{-2s_1/d}K_2^{3/2} + \zeta(K_2)^2K_2^{3/2}K_2^{-s_2/d} + \zeta(K_1)^2\zeta(K_2)^2K_1^{1/2}K_1^{-s_1/d}K_2^{3/2}N^{-1} \\
&+ \zeta(K_1)\zeta(K_2)^2K_1^{-s_1/d}K_2^{2}N^{-1/}).
\end{aligned}$$

Thus,

$$\begin{aligned}
&\|[\hat{\Psi}_{w,K_2}\Omega^{-1}_{w,K_2}]^{-1/2} - [\hat{\Psi}_{w,K_2}\Omega^{-1}_{w,K_2}]^{-1/2}\| \\
=&O_p(\zeta(K_1)^2\zeta(K_2)^2K_1^{-2s_1/d}K_2^{3/2} + \zeta(K_2)^2K_2^{3/2}K_2^{-s_2/d} + \zeta(K_1)^2\zeta(K_2)^2K_1^{1/2}K_1^{-s_1/d}K_2^{3/2}N^{-1} \\
&+ \zeta(K_1)\zeta(K_2)^2K_1^{-s_1/d}K_2^{2}N^{-1/2}).
\end{aligned} \tag{A.13}$$

For the second factor in (A.12), we consider

$$\begin{aligned}
E\|N_w^{-1/2}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}\mathbf{u}_w\|^2 =&E[tr(\mathbf{u}'_wR_{w,K_2}(R'_{w,K_2}R_{w,K_2})^{-1}R'_{w,K_2}\mathbf{u}_w)] \\
\leq&\bar{\theta}^2_w\cdot tr(E[R_{w,K_2}(R'_{w,K_2}R_{w,K_2})^{-1}R'_{w,K_2}]) \\
=&\bar{\theta}^2_w\cdot tr(I_{K_2}) = O(K_2)
\end{aligned}$$

Thus

$$\|N_w^{-1/2}\hat{\Omega}^{-1/2}_{w,K_2}R'_{w,K_2}\mathbf{u}_w\| = O_p(K_2^{1/2}). \tag{A.14}$$

Combining (A.13) and (A.14) together yields,

$$\begin{aligned}
&\frac{1}{\sqrt{N_w}}\|[\hat{\Omega}^{-1}_{w,K_2}\hat{\Psi}_{w,K_2}\hat{\Omega}^{-1}_{w,K_2}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K_2}u_w - [\hat{\Omega}^{-1}_{w,K_2}\Psi_{w,K_2}\Omega^{-1}_{w,K}]^{-1/2}\hat{\Omega}^{-1}_{w,K_2}R'_{w,K}u_w\| \\
=&[O_p(\zeta(K_1)^2\zeta(K_2)^2K_1^{-2s_1/d}K_2^{3/2} + \zeta(K_2)^2K_2^{3/2}K_2^{-s_2/d} + \zeta(K_1)^2\zeta(K_2)^2K_1^{1/2}K_1^{-s_1/d}K_2^{3/2}N^{-1} \\
&+ \zeta(K_1)\zeta(K_2)^2K_1^{-s_1/d}K_2^{2}N^{-1/2})]\cdot O_p(K_2^{1/2}) \\
=&o_p(1)
\end{aligned}$$

under assumption 5.2 and 5.3. Finally, (A.11) is

$$
\begin{aligned}
&\frac{1}{\sqrt{N_w}}\|[\hat{\Omega}_{w,K_2}^{-1}\Psi_{w,K_2}\Omega_{w,K_2}^{-1}]^{-1/2}\hat{\Omega}_{w,K_2}^{-1}R'_{w,K_2}\mathbf{u}_w - [\Omega_{w,K_2}^{-1}\Psi_{w,K_2}\Omega_{w,K_2}^{-1}]^{-1/2}\Omega_{w,K_2}^{-1}R'_{w,K_2}\mathbf{u}_w\| \\
\leq&\|[\Psi_{w,K_2}\Omega_{w,K_2}^{-1}]^{-1/2}\|\|\hat{\Omega}_{w,K_2}^{-1/2} - \Omega_{w,K_2}^{-1/2}\|\|\frac{1}{\sqrt{N_w}}R'_{w,K_2}\mathbf{u}_w\|. \qquad (A.15)
\end{aligned}
$$

The first factor in (A.15) is $\|[\Psi_{w,K_2}\Omega_{w,K_2}^{-1}]^{-1/2}\| = C\|I_{K_2}\| = O(K_2^{1/2})$. The second factor in (A.15) is $O_p(\zeta(K_2)K_2^{1/2}N^{-1/2})$. For the third factor in (A.15), consider

$$
\begin{aligned}
&E\|\frac{1}{\sqrt{N_w}}R'_{w,K_2}\mathbf{u}_w\|^2 \\
=&E[\frac{1}{\sqrt{N_w}}tr(\mathbf{u}'_w R_{w,K_2}R'_{w,K_2}\mathbf{u}_w)] \\
=&E[\frac{1}{\sqrt{N_w}}tr(R'_{w,K_2}\mathbf{u}_w\mathbf{u}'_w R_{w,K_2}] \\
=&tr(\frac{1}{\sqrt{N_w}}E[R'_{w,K_2}E[\mathbf{u}_w\mathbf{u}'_w|\mathbf{X_w}]R_{w,K_2}]) \\
=&\bar{\theta}_w^2 tr(\Omega_{w,K_2}) \\
=&\bar{\theta}_w^2 K_2\lambda_{\max}(\Omega_{w,K_2}) \\
=&C\dot{K}_2.
\end{aligned}
$$

Thus $|\frac{1}{\sqrt{N_w}}R'_{w,K_2}\mathbf{u}_w\| = O_p(K_2^{1/2})$. Putting these together, we have

$$
\begin{aligned}
&\frac{1}{\sqrt{N_w}}\|[\hat{\Omega}_{w,K_2}^{-1}\Psi_{w,K_2}\Omega_{w,K_2}^{-1}]^{-1/2}\hat{\Omega}_{w,K_2}^{-1}R'_{w,K_2}\mathbf{u}_w - [\Omega_{w,K_2}^{-1}\Psi_{w,K_2}\Omega_{w,K_2}^{-1}]^{-1/2}\Omega_{w,K_2}^{-1}R'_{w,K_2}\mathbf{u}_w\| \\
=&O_p(\zeta(K_2)K_2^{3/2}N^{-1/2}) = o_p(1).
\end{aligned}
$$

Hence,

$$\|[\hat{\Omega}_{0,K_2}^{-1}\hat{\Psi}_{w,K_2}\hat{\Omega}_{w,K_2}^{-1}]^{-1/2}\sqrt{N_w}(\hat{\alpha}_{w,K_2} - \alpha^*_{w,K_2}) - S_{w,K_2}\| = o_p(1).$$

Thus, the infeasible test statistic converges to standard normal distribution. □

Now, we prove Theorem 1.

Proof. Define

$$\hat{\delta}_{K_2} = \hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2},\ \delta^*_{K_2} = \alpha^*_{1,K_2} - \alpha^*_{0,K_2}.$$

Following the logic of Lemma 8 above, we can conclude that

$$T^* = \frac{(\hat{\delta}_{K_2} - \delta^*_{K_2})'\hat{V}^{-1}(\hat{\delta}_{K_2} - \delta^*_{K_2}) - K_2}{\sqrt{2K_2}} \xrightarrow{d} N(0,1).$$

To complete the proof, we need to show that $|T^* - T| = o_p(1)$. Note that under the null hypothesis, $\mu_1(x) = \mu_0(x)$, so we may choose the same approximation sequence $\alpha^0_{1,K_2} = \alpha^0_{0,K_2}$ for $\sigma^0_{1,K_2}(x) = \sigma^0_{0,K_2}(x)$. Then,

$$\begin{aligned}\|\alpha^*_{1,K_2} - \alpha^*_{0,K_2}\| &= \|\alpha^*_{1,K_2} - \alpha^0_{1,K_2} + \alpha^0_{0,K_2} - \alpha^*_{0,K_2}\| \\ &\le \|\alpha^*_{1,K_2} - \alpha^0_{1,K_2}\| + \|\alpha^0_{0,K_2} - \alpha^*_{0,K_2}\| \\ &= O(K_2^{1/2} K_2^{-s_2/d}).\end{aligned}$$

by Lemma 6 (ii) and

$$\begin{aligned}\|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\| &= \|\hat{\alpha}_{1,K_2} - \alpha^0_{1,K_2} + \alpha^0_{0,K} - \hat{\alpha}_{0,K_2}\| \\ &\le \|\hat{\alpha}_{1,K_2} - \alpha^0_{1,K_2}\| + \|\alpha^0_{0,K_2} - \hat{\alpha}_{0,K_2}\| \\ &= O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2}N^{-1}) + O_p(K_2^{1/2}N^{-1/2}) + O(K_2^{-s_2/d}) \\ &+ K_2^{1/2}N^{-1/2} + K_2^{-s_2/d})\end{aligned}$$

by Lemma 6 (iii). Thus,

$$\begin{aligned}|T^* - T| &= |(\frac{\hat{\delta}_{K_2} - \delta^*_{K_2})'\hat{V}^{-1}(\hat{\delta}_{K_2} - \delta^*_{K_2}) - K_2}{\sqrt{2K_2}} - \frac{\hat{\delta}'_{K_2}\hat{V}^{-1}\hat{\delta}_{K_2} - K_2}{\sqrt{2K_2}}| \\ &\le \frac{2}{\sqrt{2K_2}}|(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'\hat{V}^{-1}(\alpha^*_{1,K_2} - \alpha^*_{0,K_2})| &\text{(A.16)} \\ &+ \frac{1}{\sqrt{2K_2}}|(\alpha^*_{1,K_2} - \alpha^*_{0,K_2})'\hat{V}^{-1}(\alpha^*_{1,K_2} - \alpha^*_{0,K_2})|. &\text{(A.17)}\end{aligned}$$

For (A.16),

$$\begin{aligned}&\frac{2}{\sqrt{2K_2}}|(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'\hat{V}^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})| \\ =&\frac{N}{\sqrt{2K_2}}N \cdot |tr((\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'[N\hat{V}]^{-1}(\alpha^*_{1,K_2} - \alpha^*_{0,K_2}))| \\ \le&\frac{2}{\sqrt{2K_2}}N \cdot \lambda_{max}([N\hat{V}]^{-1})\|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\|\|\alpha^*_{1,K_2} - \alpha^*_{0,K_2}\| \\ =&\frac{2}{\sqrt{2K_2}}N \cdot [C + o_p(1)] \cdot [O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2}N^{-1}) + O_p(K_2^{1/2}N^{-1/2}) + O(K_2^{-s_2/d}) \\ &+ K_2^{1/2}N^{-1/2} + K_2^{-s_2/d})][O(K_2^{1/2}K_2^{-s_2/d})] \\ =&O_p(\zeta(K_1)K_1^{-s_1/d}K_2^{-s_2/d}N + \zeta(K_1)K_1^{1/2}K_2^{-s_2/d} + K_2^{1/2}K_2^{-s_2/d}N^{1/2} + K_2^{-2s_2/d}N).\end{aligned}$$

For (A.17),

$$\frac{1}{\sqrt{2K_2}}|(\alpha^*_{1,K_2} - \alpha^*_{0,K_2})'\hat{V}^{-1}(\alpha^*_{1,K_2} - \alpha^*_{0,K_2})|$$

$$
\begin{aligned}
=&\frac{1}{\sqrt{2K_2}} \cdot N \cdot |tr((\alpha^*_{1,K_2} - \alpha^*_{0,K_2})'[N\hat{V}]^{-1}(\alpha^*_{1,K_2} - \alpha^*_{0,K_2}))| \\
\leq&\frac{1}{\sqrt{2K_2}} \cdot N \cdot \lambda_{max}([N\hat{V}]^{-1})\|\alpha^*_{1,K_2} - \alpha^*_{0,K_2}\|^2 \\
=&N \cdot [C + o_p(1)][O(K_2^{1/2}K_2^{-s_2/d})]^2 \\
=&O(K_2^{1/2}K_2^{-2s_2/d}N).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&|T^* - T| \\
=&O_p(\zeta(K_1)K_1^{-s_1/d}K_2^{-s_2/d}N + \zeta(K_1)K_1^{1/2}K_2^{-s_2/d} + K_2^{1/2}K_2^{-s_2/d}N^{1/2} + O(K_2^{1/2}K_2^{-2s_2/d}N) \\
=&o_p(1),
\end{aligned}
$$

by 5.2 and 5.3. Hence the result follows. □

Now, we prove Theorem 2.

Proof.

$$
\begin{aligned}
&\rho_N \cdot \sup_{x\in\mathcal{X}} |\Delta(x)| = \sup_{x\in\mathcal{X}} |\sigma_1^2(x) - \sigma_0^2(x)| \\
\leq& \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{1,K_2} - \sigma_1^2(x)| + \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{0,K_2} - \sigma_0^2(x)| + \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\hat{\alpha}_{1,K_2} - R_{K_2}(x)'\alpha^0_{1,K_2}| \\
&+ \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\hat{\alpha}_{0,K_2} - R_{K_2}(x)'\alpha^0_{0,K_2}| + \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\hat{\alpha}_{1,K_2} - R_{K_2}(x)'\hat{\alpha}_{0,K_2}| \\
\leq& \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{1,K_2} - \sigma_1^2(x)| + \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{0,K_2} - \sigma_0^2(x)| + \sup_{x\in\mathcal{X}} \|R_{K_2}(x)\| \cdot \|\hat{\alpha}_{1,K_2} - \alpha 0_{1,K_2}\| \\
&+ \sup_{x\in\mathcal{X}} \|R_{K_2}(x)\| \cdot \|\hat{\alpha}_{0,K_2} - \alpha^0_{0,K_2}\| + \sup_{x\in\mathcal{X}} \|R_{K_2}(x)\| \cdot \|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\| \\
=& \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{1,K_2} - \sigma_1^2(x)| + \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{0,K_2} - \sigma_0^2(x)| \\
&+ \zeta(K_2)\|\hat{\alpha}_{1,K_2} - \alpha^0_{1,K_2}\| + \zeta(K_2)\|\hat{\alpha}_{0,K_2} - \alpha^0_{0,K_2}\| + \zeta(K_2)\|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\|
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&\|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\| \\
\geq&\zeta(K_2)^{-1} \cdot \rho_N \cdot \sup_{x\in\mathcal{X}} |\Delta(x)| - \zeta(K_2)^{-1} \cdot \rho_N \cdot \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{1,K_2} - \sigma_1^2(x)| \\
&- \zeta(K_2)^{-1} \cdot \rho_N \cdot \sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{0,K_2} - \sigma_0^2(x)| - \|\hat{\alpha}_{1,K_2} - \alpha^0_{1,K_2}\| - \|\hat{\alpha}_{0,K_2} - \alpha^0_{0,K_2}\| \\
\geq&\zeta(K_2)^{-1} \cdot \rho_N \cdot C_0\Bigg(1 - \frac{\sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{1,K_2} - \sigma_1^2(x)|}{\rho_N \cdot C_0} - \frac{\sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{0,K_2} - \sigma_0^2(x)|}{\rho_N \cdot C_0}
\end{aligned}
$$

$$- \zeta(K_2)\frac{\|\hat\alpha_{1,K_2} - \alpha^0_{1,K_2}\|}{\rho_N \cdot C_0} - \zeta(K_2)\frac{\|\hat\alpha_{0,K_2} - \alpha^0_{0,K_2}\|}{\rho_N \cdot C_0}\Bigg).$$

Under assumption 5.1 and 5.3, we have

$$\begin{aligned}
&\frac{\sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{1,K_2} - \sigma_1^2(x)|}{\rho_N \cdot C_0} = O(K_2^{-s_2/d}) \cdot O(N^{1/2-3v/2_1-3v_2/2-\varepsilon}) = o(1),\\
&\frac{\sup_{x\in\mathcal{X}} |R_{K_2}(x)'\alpha^0_{0,K_2} - \sigma_0^2(x)|}{\rho_N \cdot C_0} = O(K_2^{-s_2/d}) \cdot O(N^{1/2-3v/2_1-3v_2/2-\varepsilon}) = o(1),\\
&\zeta(K_2)\frac{\|\hat\alpha_{1,K_2} - \alpha^0_{1,K_2}\|}{\rho_N \cdot C_0} = O(K_2)\cdot[O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2}N^{-1})\\
&\quad + O_p(K_2^{1/2}N^{-1/2}) + O(K_2^{-s_2/d})] \cdot O(N^{1/2-3v/2_1-3v_2/2-\varepsilon}) = o(1),\\
&\zeta(K_2)\frac{\|\hat\alpha_{0,K_2} - \alpha^0_{0,K_2}\|}{\rho_N \cdot C_0} = O(K_2)\cdot[O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2}N^{-1})\\
&\quad + O_p(K_2^{1/2}N^{-1/2}) + O(K_2^{-s_2/d})] \cdot O(N^{1/2-3v/2_1-3v_2/2-\varepsilon}) = o(1).
\end{aligned}$$

Hence, $\|\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2}\| \geq \zeta(K_2)^{-1}\cdot\rho_N\cdot C_0$ with probability going to 1 as $N\to\infty$.

$$N^{1/2}\zeta(K_1)^{-3/2}K_2^{-1/2}\|\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2}\| \geq N^{1/2}\zeta(K_1)^{-3/2}K_2^{-1/2}\zeta(K_2)^{-1}\cdot\rho_N\cdot C_0$$

with probability going to 1 as $N\to\infty$. Since

$$N^{1/2}\zeta(K_1)^{-3/2}K_2^{-1/2}\zeta(K_2)^{-1}\cdot\rho_N\cdot C_0 \geq CN^{1/2}\zeta(K_1)^{-3/2}K_2^{-1/2}\zeta(K_2)^{-1}\cdot N^{-1/2+3v/2_1+3v_2/2+\varepsilon} \geq CN^{\varepsilon},$$

for any M',

$$Pr(N^{1/2}\zeta(K_1)^{-3/2}K_2^{-1/2}\|\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2}\| > M') \to 1. \tag{A.18}$$

Next, we show

$$Pr\left(\frac{\tilde C\cdot(\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2})'V^{-1}(\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2}) - K_2}{\sqrt{2K_2}} > M\right) \to 1$$

for an arbitrary positive constant $\tilde C$. Let $\underline{\lambda}$ and $\bar\lambda$ be the minimum and maximum eigenvalues of $[NV]^{-1}$, respectively. $\underline{\lambda}$ is bounded away from 0 and $\bar\lambda$ is bounded.

$$\begin{aligned}
&Pr\left(\frac{\tilde C\cdot(\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2})'V^{-1}(\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2}) - K_2}{\sqrt{2K_2}} > M\right)\\
=&Pr\left(\frac{\tilde CN\cdot(\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2})'[NV]^{-1}(\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2}) - K_2}{\sqrt{2K_2}} > M\right)\\
=&Pr\left(\tilde CN\cdot(\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2})'[NV]^{-1}(\hat\alpha_{1,K_2} - \hat\alpha_{0,K_2}) > \sqrt{2K_2}M + K_2\right)
\end{aligned}$$

$$\begin{aligned}
\geq & Pr\left(\underline{\lambda}\tilde{C}N \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) > \sqrt{2K_2}M + K_2\right) \\
= & Pr\left(N\zeta(K_1)^{-3}K_2^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) > (\underline{\lambda}\tilde{C})^{-1}\zeta(K_2)^{-3}(\sqrt{2}MK_2^{-1/2} + 1)\right) \\
= & Pr\left(N^{1/2}\zeta(K_1)^{-3/2}K_2^{-1/2}\|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\| > (\underline{\lambda}\tilde{C})^{-1/2}\zeta(K_2)^{-3/2}(\sqrt{2}MK_2^{-1/2} + 1)^{1/2}\right)
\end{aligned}$$

Since for any M, for large enough N, we have

$$(\underline{\lambda}\tilde{C})^{-1/2}\zeta(K_2)^{-3/2}(\sqrt{2}MK_2^{-1/2} + 1)^{1/2} < 2(\underline{\lambda}\tilde{C})^{-1/2},$$

it follows that for large N,

$$\begin{aligned}
& Pr\left(\frac{\tilde{C} \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'V^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M\right) \\
\leq & Pr\left(N^{1/2}\zeta(K_1)^{-3/2}K_2^{-1/2}\|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\| > 2(\underline{\lambda}\tilde{C})^{-1/2}\right) \\
\to & 1.
\end{aligned} \tag{A.19}$$

by (A.18). Then, we show that

$$Pr(T > M) = Pr\left(\frac{(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'\hat{V}^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M\right).$$

Let $\hat{\underline{\lambda}}$ be the minimum eigenvalue of $[N\hat{V}^{-1}]$. Let B_1 denote the event that $\hat{\underline{\lambda}} > \underline{\lambda}/2$. $Pr(B_1) \to 1$ as $N \to \infty$ by lemma 7. In addition, let B_2 be the event

$$\frac{(\underline{\lambda}/2)N(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M$$

.

$$\begin{aligned}
& Pr\left(\frac{\tilde{C} \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'V^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M\right) \\
= & Pr\left(\frac{\tilde{C}N \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'[NV]^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M\right) \\
\leq & Pr\left(\frac{\bar{\lambda}\tilde{C}N \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M\right) \\
\to & 1
\end{aligned}$$

as $N \to \infty$ by (A.19). Let $\tilde{C}$ be $\underline{\lambda}/2\bar{\lambda}^{-1}$, then $Pr(B_2) \to 1$ as $N \to \infty$. Thus, $Pr(B_1 \cap B_2) \to 1$ as $N \to \infty$. Note that the event $B_1 \cap B_2$ implies that

$$T = \frac{(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'\hat{V}^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}}$$

$$
\begin{aligned}
&\geq \frac{\hat{\underline{\lambda}} N(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} \\
&> \frac{(\underline{\lambda}/2) N(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} \\
&> M.
\end{aligned}
$$

Hence, $Pr(T > M) \to 1$.

□

References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association 97*(457), 284–292.

Andrews, D. W. and X. Shi (2013). Inference based on conditional moment inequalities. *Econometrica 81*(2), 609–666.

Bentkus, V. (2005). A Lyapunov-type bound in R^d. *Theory of Probability* & *Its Applications 49*(2), 311–323.

Card, D. (2001). The effect of unions on wage inequality in the US labor market. *Industrial & Labor Relations Review 54*(2), 296–315.

Card, D., T. Lemieux, and W. C. Riddell (2004). Unions and wage inequality. *Journal of Labor Research 25*(4), 519–559.

Chang, M., S. Lee, and Y.-J. Whang (2015). Nonparametric tests of conditional treatment effects with an application to single-sex schooling on academic achievements. *The Econometrics Journal 18*(3), 307–346.

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2007). Additional Proofs for: Nonparametric tests for treatment effect heterogeneity. `http://moya.bus.miami.edu/~omitnik/PDF_Documents/ReStat_08_08_additional_proofs.pdf`. Accessed August 28, 2016].

Crump, R. K., V. J. Hotz, G. W. Imbens, and O. A. Mitnik (2008). *Nonparametric tests for treatment effect heterogeneity.*

DiNardo, J., N. Fortin, and T. Lemieux (1996). Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica 64*(5), 1001–1044.

Fan, J. and Q. Yao (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika 85*(3), 645–660.

Freeman, R. B. (1980). Unionism and the dispersion of wages. *Industrial & Labor Relations Review 34*(1), 3–23.

Gosling, A. and S. Machin (1995). Trade unions and the dispersion of earnings in British establishments, 1980–90. *Oxford Bulletin of Economics and Statistics 57*(2), 167–184.

Hong, Y. and H. White (1995). Consistent specification testing via nonparametric series regression. *Econometrica: Journal of the Econometric Society*, 1133–1159.

Hsu, Y.-C. (2011). Consistent tests for conditional treatment effects. *Unpublished manuscript, Department of Economics, University of Missouri at Columbia*.

Imbens, G. W., W. K. Newey, and G. Ridder (2005). Mean-square-error calculations for average treatment effects. *Unpublished manuscript, Department of Economics, Harvard University*.

Johnson, G. E. (1975). Economic analysis of trade unionism. *The American Economic Review 65*(2), 23–28.

Lee, M.-J. (2009). Non-parametric tests for distributional treatment effect for randomly censored responses. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71*(1), 243–264.

Lemieux, T. (1993). Unions and wage inequality in Canada and the United States. In *Small differences that matter: Labor markets and income maintenance in Canada and the United States*, pp. 69–108. University of Chicago Press.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators. *Econometrica*, 1349–1382.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168.

Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

Song, Q. and L. Yang (2009). Spline confidence bands for variance functions. *Journal of Nonparametric Statistics 21*(5), 589–609.

Vella, F., M. Verbeek, et al. (1998). Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. *Journal of Applied Econometrics 13*(2), 163–183.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data, 2nd ed.* MIT press.

Yu, K. and M. Jones (2004). Likelihood-based local linear estimation of the conditional variance function. *Journal of the American Statistical Association 99*(465), 139–144.

Zhu, L., Y. Dong, and R. Li (2013). Semiparametric estimation of conditional heteroscedasticity via single-index modeling. *Statistica Sinica 23*(3), 1215.