NONPARAMETRIC TESTS FOR THE EFFECT OF A TREATMENT ON THE CONDITIONAL VARIANCE

Yanchun Jin*

This paper proposes nonparametric tests for the null hypothesis that a treatment has a zero effect on the conditional variance for all subpopulations characterized by the values of the covariates. Rather than the mean of an outcome, which measures the extent to which a treatment changes the level of the outcome, researchers are sometimes interested in how the treatment affects the dispersion of the outcome. We use the variance to measure dispersion and estimate the conditional variances using the series method. We provide a test rule that compares a Wald-type test statistic with the critical value of a 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.

Key words and phrases: Conditional variance, series estimation, treatment effect.

1. Introduction

Recently, treatment effect analyses have become important tools in various fields of empirical research to evaluate the impacts of policies. In this paper, we consider the effect of a treatment on the variance. Most existing studies focus on various treatment effects on the mean of the outcome of interest, such as the average treatment effect and the local average treatment effect. These parameters measure the extent to which the treatment changes the level of the outcome. However, researchers are also interested in the treatment effect on the dispersion of the 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 than they are in nonunion sectors (Freeman (1980), Gosling and Machin (1995), DiNardo et al. (1996), Card (2001)). Freeman (1980) compares the variances of the wages of union workers and nonunion workers, and finds that unionism reduces the wage differential in the organized sector, and that this difference-reducing effect within sectors is larger than the gap-increasing effect across industries. Investigating the treatment effect on the variance and how it changes across subpopulations is important to understand how unionism works.

In this paper, we provide nonparametric tests for the effect of a treatment on the variance. In particular, we consider a test for the null hypothesis that the treatment has a zero effect on the dispersion of the outcome for all subpopulations.
defined by covariates. This is useful, for example, when we want to investigate whether there is any evidence of heterogeneity in the effect of unionism on the dispersion of the outcome. Card (2001) examines the gap in the variance of wages for union and nonunion male workers defined by various skill groups, and finds that the role of unions in compressing the wage dispersion of high-skilled workers is slightly stronger than in the case of low-skilled workers. By conducting the test proposed in this paper, we can examine whether there are any subpopulations for which unions change their wage dispersion.

Although a large part of the recent literature on treatment effects focuses on estimation, studies on hypothesis testing for the treatment effect are limited. Abadie (2002) considers the distributions of the outcomes for the treatment and control groups, and tests for the null hypotheses of equality and first-order stochastic dominance using the bootstrap method. Crump et al. (2008) test for the treatment effect heterogeneity and develop tests based on series estimations. They test for the null hypothesis that the average treatment effects, conditional on the covariates, are zero for all subpopulations defined by the covariates. In addition, they propose a test for the null hypothesis that the average effect, conditional on the covariates, is constant for all subpopulations. Lee (2009) studies a nonparametric test for a null hypothesis of no distributional treatment effects for randomly censored outcomes. Hsu (2017) studies a one-sided test for the conditional treatment effect, employing Andrews and Shi’s (2013) instrumental variable approach. Chang et al. (2015) construct tests for the null hypotheses of the conditional stochastic dominance treatment effect and the positive conditional average treatment for all covariates, using test statistics based on kernel estimators.

None of the hypotheses considered in the aforementioned studies provide tests for the treatment effect on the conditional variance. In this paper, we aim to fill this gap in the literature. The null hypothesis considered here is that the difference between the conditional variances of the outcomes of the treatment group and control group is zero for all subpopulations defined by the 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 the parametric method. Therefore, it is necessary to use a nonparametric method.

We provide tests based on a two-step series approach to estimate the conditional variances. In the first step, we estimate the conditional mean function using series for the treated group and control group, and then compute their residuals. Then, we estimate the conditional variance in the second step by regressing the squared residuals on the series terms for the treated group and the control group. Given the particular series terms, the test can be considered as a test for the equality of the two coefficients of the conditional variance functions. Then, the test can be implemented using the standard parametric method. We give a Wald-type statistic with quadratic forms of the differences in the parameter estimates, and compare the test statistic to the critical value of the 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 of a standard Gaussian distribution.

By conducting a regression using squared residuals, we take into account two kinds of bias: bias of the squared residuals in the first step, and bias from the conditional variance function estimation that arises in the usual nonparametric regression analysis. Given particular series terms, the test can be viewed as a test of whether the coefficients of the treatment and control group 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 studied by Crump et al. (2008) who consider a one-step test for the conditional mean.

Our tests are similar to Hong and White’s (1995) nonparametric tests, in that they also estimate a nonparametric model using series regression and provide test statistics that converge in distribution to a standard 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 conditions for the number of series terms that ensure the validity of their tests. The increasing rates of the series terms are important also in our tests. This paper is also related to the literature on the estimation of the 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 squared residuals to estimate the 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 our hypothesis using a power series estimator of the coefficients of the conditional variance functions, which is easy to compute.

In addition, our tests for the treatment effect of the conditional variance are related to the quantile treatment effect, which also investigates the treatment effect beyond the average treatment effect. Abadie et al. (2002) and Chernozhukov and Hansen (2005) estimate conditional quantile treatment effects using instrumental variables. Firpo (2007) develops a method to identify and estimate the unconditional quantile treatment effect under unconfoundedness. Our proposed tests are particularly useful when we are interested in the treatment effect on the dispersion of the outcomes.

The rest of this paper is organized as follows. In Section 2, we describe the framework for the program evaluation analysis, and give our null hypothesis and
the alternative. Section 3 illustrates the test using a parametric model. Then, in Section 4, we extend this to a nonparametric model with series estimation and provide the test statistics. In Section 5, we give the asymptotic theorem for our test statistic, given certain assumptions. In Section 6, we conduct simulations and demonstrate the results of the test property in a finite sample. In Section 7, we consider an empirical application of 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 the potential outcomes for each unit $i$ under treatment and control, respectively. For each unit $i$, we observe $W_i$ and $Y_i$, where $Y_i = Y_i(W_i) = W_i \cdot Y_i(1) + (1 - W_i) \cdot Y_i(0)$. 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 the conditional variance is $\text{Var}[Y(1) | X = x] - \text{Var}[Y(0) | X = x]$.

Our null hypothesis is given as follows:

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

against the alternative:

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

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

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

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

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

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

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

**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 the 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 unconfoundedness (Rosenbaum and Rubin (1983)), which can be described as follows.

**Assumption 2.2. (Unconfoundedness):**

\[ W \perp \perp (Y(0), Y(1)) \mid X, \]

where \( \perp \perp \) denotes independence.

Assumption 2.2 assumes that, conditional on a set of covariates \( X \), the pair of counterfactual outcomes, \( (Y(0), Y(1)) \) is independent of \( W \).

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

**Assumption 2.3. (Overlap):**

\[ 0 < \Pr(W = 1 \mid X = x) < 1. \]

Under these assumptions, \( \text{Var}[Y(1) \mid X = x] = \text{Var}[Y \mid X = x, W = 1] \) and \( \text{Var}[Y(0) \mid X = x] = \text{Var}[Y \mid X = x, W = 0] \). Hence, \( \sigma^2_1(x) \) and \( \sigma^2_0(x) \) can be identified as computing the conditional variances for the treatment and control group, respectively. Then, we can implement our test using the 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 \( X = (X'_1, \ldots, X'_{N_1}, X'_{N_1+1}, \ldots, X'_N) \). In addition, let \( N_1 \times d \) matrix \( X_1 = (X'_1, X'_2, \ldots, X'_{N_1}) \), \( N_0 \times d \) matrix \( X_0 = (X'_{N_1+1}, X'_{N_1+2}, \ldots, X'_N) \), and let \( N_1 \) vector \( Y_1 = (Y_1, \ldots, Y_{N_1})' \), and \( N_0 \) vector \( Y_0 = (Y_{N_1+1}, \ldots, Y_N)' \).

### 3. Test statistic in parametric models

We first give a test in a standard parametric model, which helps to explain the procedure used in the nonparametric setting. Then, in Section 5, we provide the conditions on the series terms that guarantee the validity of the tests in large samples without the parametric assumptions.

To construct the test statistic, it is necessary to estimate \( \sigma^2_1(x) - \sigma^2_0(x) \). We specify \( \sigma^2_w(x) \) as the following linear function:

\[ (3.1) \quad \sigma^2_w(x) = \beta'_w x. \]

In this parametric setting, the null and alternative hypotheses are

\[ (3.2) \quad H_0 : \beta_1 = \beta_0, \]

\[ (3.3) \quad H_1 : \beta_1 \neq \beta_0. \]
where \( \beta_1 \) is in the \( K \)-dimension. This can be tested using the following Wald-type test statistic

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

where \( \hat{\beta}_w \) is an estimator of \( \beta \), \( \hat{\Omega}_w \) is the estimator of the asymptotic variance matrix of \( \hat{\beta}_w \), and \( N_1 \) and \( N_0 \) are the sample sizes of the treated and control groups, respectively.

Now, we conduct regressions in the treatment and control groups, respectively, to compute \( \hat{\beta}_1 \) and \( \hat{\beta}_0 \) using a least squares estimation. 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 \( \hat{\xi}_w \). Then, the residuals are \( \hat{\epsilon}_w = Y_w - \hat{\mu}_w(X_w) = Y_w - X_w \hat{\xi}_w \). Let \( \hat{\epsilon}_{1,i} \) be the \( i \)-th element of \( \hat{\epsilon}_1 \) and \( \hat{\epsilon}_{0,i} \) be the \( (i - N_1) \)-th element of \( \hat{\epsilon}_0 \). This leads to the residual-based estimator as \( \hat{\sigma}^2_w(x) = \hat{\beta}_w'x \) by solving the following problem,

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

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

However, specifying the variance function as model (3.1) is not standard, and there is no widely accepted model for variance. In addition, there is a risk of specifying the conditional mean function with a standard linear model. To get rid of misspecification, we estimate the conditional mean and 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 the 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 a standard linear model, 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 \( \sigma^2_w(x) \) by \( \hat{\sigma}^2_{w,K_2}(x) \) using \( \hat{\epsilon}_{w,i}^2 \), where \( K_1 \) and \( K_2 \) denote the number of series terms in the 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 \( p_l(x) = x^{\lambda(l)} \), and \( P_l(x) = (p_1(x), \ldots, p_l(x))' \). 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 \( P_{0,K_1} = (P'_{K_1}(X_{1N}), \ldots, P'_{K_1}(X_{N})) \) and \( N_0 \times K_1 \) matrix \( P_{0,K_1} = (P'_{K_1}(X_{1N}), \ldots, P'_{K_1}(X_{N})) \). Define the \( N_1 \times K_2 \) matrix \( P_{1,K_2} = (P'_{K_2}(X_{1}), \ldots, P'_{K_2}(X_{N})) \) and \( N_0 \times K_2 \) matrix \( P_{0,K_2} = (P'_{K_2}(X_{1N}), \ldots, P'_{K_2}(X_{N})) \).
The nonparametric series estimator of the regression function $\mu_w(x)$, given series terms $P_{w,K_1}$, is given by

$$
(4.1) \quad \hat{\mu}_{w,K_1}(x) = P_{K_1}(x)'(P_{w,K_1}'P_{w,K_1})^{-1}P_{w,K_1}'Y_w,
$$

where $A^{-}$ denotes a generalized inverse of $A$. Then, we compute the residuals by $\hat{\epsilon}_w = Y_w - \hat{\mu}_w(X_w)$. Let $\hat{\epsilon}_w^2$ denote the vector with elements that are squares of the elements of $\hat{\epsilon}_w$. Regressing $\hat{\epsilon}_w^2$ on $P_{w,K_2}$, we estimate $\sigma_w^2(x)$ by

$$
(4.2) \quad \hat{\sigma}_{w,K_2}(x) = P_{K_2}(x)'(P_{w,K_2}'P_{w,K_2})^{-1}P_{w,K_2}'\hat{\epsilon}_w^2.
$$

Let

$$
(4.3) \quad \hat{\eta}_{w,K_2} = (P_{w,K_2}'P_{w,K_2})^{-1}P_{w,K_2}'\hat{\epsilon}_w^2.
$$

In addition, we estimate the approximated limiting variance of $\sqrt{N_w}\hat{\eta}_{w,K_2}$ by $\hat{\Omega}^{-1}_{P,K_2} \hat{\Psi}_{P,K_2} \hat{\Omega}^{-1}_{P,K_2}$, where

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

where

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

is the estimate for $V_{P,K_2}$ (the variance of $\hat{\eta}_{1,K_2} - \hat{\eta}_{0,K_2}$). The 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

$$
(4.5) \quad \sigma_w^2(x) = P_{K_2}(x)'\hat{\eta}_{w,K_2}.
$$

We can approximate the distribution of $Q$ with a chi-squared distribution with degrees of freedom $K_2$. Note that only with $K_2$ increasing to infinity do the biases from the nonparametric estimation vanish.

Now, we give a normalized test statistic as

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

for the test of the null hypothesis, where $K_2$ in the numerator and $\sqrt{2K_2}$ in the denominator come from the fact that they are the mean and the standard deviation of the chi-squared distribution with degrees of freedom $K_2$.\footnote{Note that $(\hat{\eta}_{1,K_2} - \hat{\eta}_{0,K_2})'\hat{V}^{-1}_{P,K_2}(\hat{\eta}_{1,K_2} - \hat{\eta}_{0,K_2})$ is not distributed according to the chi-squared distribution because of the bias.} In the next section, we see that with $N$ and $K_2$ increasing, $T$ has asymptotically follows a standard normal distribution under the null.
5. Asymptotic theory

This section provides the asymptotic theory for our test statistic $T$. We first state the conditions in addition to Assumptions 2.1–2.3, and then we develop the 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 the variance regression functions $\sigma^2_w(x)$ are $s_2$ times continuously differentiable, with $s_1/d > 2$ and $s_2/d > 7$.
2. For $u_{w,i} = \epsilon^2_{w,i} - \sigma^2_w(X_i)$, $\theta_w(x) = E[u^2_{w,i} \mid X = x]$,
   (a) $\forall x \in \mathcal{X}$, $0 < \theta^2 \leq \theta^2_w(x) \leq \bar{\theta}^2 < \infty$, $0 < \sigma^2 \leq \sigma^2_w(x) \leq \bar{\sigma}^2 < \infty$.
   (b) $E[u^4_{w,i}] < \infty$.

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 $\Pr(W = w \mid X = x)$ 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. So $\Omega_{P1,K} = E[P_K(X)P_K(X)' \mid W = w]$ is nonsingular for all $K$. Assumption 5.2 imposes some smoothness and moment conditions on the data to ensure the asymptotic convergence of the estimators. Assumption 5.2.2-(a) ensures the nonsingularity of $\hat{\Omega}_{P,K_2}$ with probability approaching 1.

Assumption 5.3 defines the increasing rates of the number of series terms. It leads to the consistency of the conditional mean and variance estimators and the convergence of the test statistic 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 need to consider 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 the mean function estimation in the first step. In addition, the rates at which $K_2$ increases are fast enough to make the two latent conditional variance functions close under the null hypothesis, while at the same time are slow enough to maintain the approximated normal distribution. We find that the rates of the series terms need to increase quite slowly compared to the sample size. For example, with one-dimensional covariates, $K_2$ is required to increase in a rate of $CN^{1/7}$, or more slowly.
In practice, we can use “top-down” or “bottom-up” methods to select the covariates. The idea of these strategies is to choose between a model with a certain series and a model with another series. For example, the “top-down” method begins with all available covariates and tests whether these covariates are significant at a particular significance level. If they are, then we may use all of them; otherwise, the next step is to drop the least significant covariate and conduct a regression with the remaining covariates. This procedure is repeated until all the covariates meet the significance level.\(^2\) Ng and Perron (1995) describe such a strategy as a “general-to-specific modeling strategy” in the context of the lag order selection in a time series analysis. The authors conclude that the unit root test statistic based on the lag length chosen by this rule is asymptotically valid, and that the test performance is superior to that obtained by other methods. For the “bottom-up” method, we regress with one intercept and one covariate, and then select the most significant covariate. With one intercept, this selected covariate and one of all remaining covariates, we again select the most significant covariate. Repeat this procedure until all remaining covariates meet the significance level.\(^3\) In this paper, when Assumptions 5.2 and 5.3 hold, selecting covariates with these approaches, the order of series terms can be well controlled to satisfy the requirements, because the selected order increases at the same rate as the preset selection order used in the first step.

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

**Theorem 1.** Suppose Assumptions 2.1–2.3 and 5.1–5.3 hold. Then, under \(H_0\), we have,

\[
T \overset{d}{\rightarrow} N(0,1).
\]

**Proof.** See Appendix. \(\square\)

The key is the fact that the chi-squared distribution converges to the normal distribution as the degrees of freedom increase. As mentioned in Section 4, the distribution of \(Q\) can be approximated with a chi-squared distribution with degrees of freedom \(K_2\) if we ignore the bias. Given \(K_2\), there still exists bias. In an empirical application, we can test the hypothesis by comparing the 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 corresponds to the approximate asymptotic normality of \(T\) in 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 instead on the order of power series.

\(^2\) In Section 7, we repeat until all remaining covariates have \(t\)-statistics not smaller than 2 in absolute value.

\(^3\) In Sections 6 and 7, we repeat until no potential covariates have \(t\)-statistics equal to or above 2 in absolute value.
Now, we provide an informal description to understand the proof of the theorem. We first see what happens if we let \( N \) go to the infinity as in the parametric setting. Then leading \( K \) also to the infinity, we can fully see the convergence under the nonparametric setting. Define the pseudo-true values, \( \eta^*_{w,K_2} \) for \( w = 0, 1, K_2 = 1, 2, \ldots \), as

\[
\eta^*_{w,K_2} = \arg\min_{\eta} E[(\sigma^2_w(X) - P_{K_2}(X)'\eta)^2 \mid W = w] = (E[R_{K_2}(x)P_{K_2}(X)' \mid W = w])^{-1}E[P_{K_2}(X)^2_w \mid W = w]
\]

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

\begin{align}
(5.1) & \quad V^{-1/2}_{P,K_2} \cdot (\hat{\eta}_{1,K_2} - \hat{\eta}_{0,K_2}) \\
(5.2) & \quad = V^{-1/2}_{P,K_2} \cdot (\eta^*_{1,K_2} - \eta^*_{0,K_2}) + V^{-1/2}_{P,K_2} \cdot (\hat{\eta}_{0,K_2} - \eta^*_{0,K_2}). \\
(5.3) & \quad + 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{align}

Under Assumptions 2.1–2.3 and 5.1–5.3, the two terms in (5.3) converge to normal distributions with mean 0 for a given \( K_2 \). We can have the asymptotic distribution of \( T \) based on this approximate normality if (5.2) can be ignored. However, for fixed \( K_2 \), (5.2) is not equal to 0, even if \( \sigma^2_w(X) = \sigma^2_0(X) \), because the covariate distributions differ between the treatment and control groups. For large \( K_2 \), \( \sigma^2_w(x) \) is close to \( P_{K_2}(x)'\eta^*_{w,K_2} \) for all \( x \). Hence, for large enough \( K_2 \), \( P_{K_2}(x)'V^{-1/2}_{P,K_2} (\eta^*_{1,K_2} - \eta^*_{0,K_2}) \) is close to 0 for all \( x \), implying (5.2) 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 the deviation of the first term from 0 small, while at the same time slowly enough to maintain the approximation of the distribution of the two terms in (5.3) by a normal one. Note that, by Bentkus (2005, Theorem 1.1), we can approximate the distribution of a vector with a multivariate standard Gaussian distribution even with the dimension of the vector increasing. 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 presented in this paper. In the second step, conducting the regression of the residuals obtained from the first step estimation, we need to consider the error from the mean estimation. This can be explained intuitively as follows. Observe that

\begin{align}
(5.4) & \quad \hat{\epsilon}^2_{w,i} - \epsilon^2_{w,i} = (\epsilon_{w,i} + \mu_w(X_i) - \hat{\mu}_{w,K}(X_i))^2 - \epsilon^2_{w,i} \\
(5.5) & \quad = 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.
\end{align}

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

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,

\[
\sigma^2_1(x) - \sigma^2_0(x) = \rho_N \cdot \Delta(x),
\]

with \( \Delta(x) \) \( s_2 \) times continuously differentiable, \(|\Delta(x_0)| = C_0 > 0 \) for some \( x_0 \), and \( \rho_N^{-1} = O(N^{1/2-3\nu/2_1-3\nu_2/2-\varepsilon}) \) for some \( \varepsilon > 0 \). Then, as \( N \to \infty \), for all \( M \),

\[
\Pr(T \geq M) \to 1.
\]

Proof. See Appendix. \( \square \)

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 that when the true model is that of (4.5) for a fixed \( K_2 = k \), the nonparametric test will decrease the power of the test. The nonparametric test checks for additional parameters larger than the necessary \( k \), whereas under the parametric model, this would be zero. Therefore, this additional procedure reduces the power.

6. Monte Carlo experiment

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

In the following two experiments, we consider three different sample sizes: \( N_0 = 45, N_1 = 55; N_0 = 200, N_1 = 300; \) and \( N_0 = 450, N_1 = 550 \), with 10000 repetitions. In order to show the performance of the tests under different series terms, we give the result when the power series with order \( K = 2, 3, 4 \) are used. 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 the “bottom-up” method described in Section 5.

First, we simulate the property of the test statistic under the null hypothesis. Our data-generating process is as follows:

\[
X_i \sim U[0, 1]
\]
\[
\mu_0(X_i) = X_i,
\]
\[
\mu_1(X_i) = 5X_i^2 + 2X_i,
\]
\[
\epsilon_{0,i} = \tau_0 \sqrt{\sigma^2_0(X_i)},
\]
\[
\epsilon_{1,i} = \tau_1 \sqrt{\sigma^2_1(X_i)/2},
\]
\[
Y_{0,i} = \mu_0(X_i) + \epsilon_{0,i} = x + \tau_0 \sqrt{e^{X_i}}
\]
\[
Y_{1,i} = \mu_1(X_i) + \epsilon_{1,i} = 5X_i^2 + 2X_i + \tau_1 \sqrt{e^{X_i}/2},
\]

where \( \tau_0 \) and \( \tau_1 \) are parameters in constructing the variance function, with \( \tau_0 \sim N(0, 1), \tau_1 \sim \text{Student’s } t \text{ distribution with degrees of freedom } 4. \)

\footnote{The variance of Student’s } t \text{ distribution with degrees of freedom } \nu, \text{ for } \nu > 2, \text{ is } \nu/(\nu - 2).
generated above, we have

\[(6.1) \quad \text{Var}[Y_0 \mid X = x] = \text{Var}[Y_0 \mid X = x] = e^x \quad \text{for all } x \in X,\]

and the null hypothesis holds. We use this set of data to illustrate the size properties of our test. In this setting, the conditions mentioned in Sections 2 and 5 are satisfied.

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 levels 10%, 5%, and 1%, respectively, and the right panel shows the probabilities of the nonparametric test with test statistic $Q$ under the same significance levels. The numbers in brackets are the average numbers of selected series terms in the 10000 repetitions when using the “bottom-up” method. From the table, we see that empirical rejection probabilities are sensitive to the choice of the order of the power series terms. With the number of series terms increasing, the tests expose distortions, which might be attributed to testing for additional coefficients for two functions. In these sample settings, tests with two series terms perform well, and when the “bottom-up” method is used, two series terms are selected on average (specifically, 2.0121, 2.0123, 2.0132). In addition, the table shows that the chi-squared test with $Q$ performs better than the test with $T$.

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

\[
X_i \sim U[0, 1] \\
\mu_0(X_i) = X_i, \\
\mu_1(X_i) = 5X_i^2 + 2X_i, \\
\epsilon_{0,i} = \tau_0 \sqrt{\sigma_0^2(X_i)}, \\
\epsilon_{1,i} = \tau_1 \sqrt{\sigma_1^2X_i^2/2}, \\
Y_{0,i} = \mu_0(X_i) + \epsilon_{0,i} = x + \tau_0 \sqrt{e^{X_i}}, \\
Y_{1,i} = \mu_1(X_i) + \epsilon_{1,i} = 5X_i^2 + 2X_i + \tau_1 \sqrt{X_i^2/2},
\]

In this case, we have

\[
\text{Var}[Y_0 \mid X = x] = e^x, \quad \text{Var}[Y_1 \mid X = x] = x^2,
\]

that is,

\[
\exists x \in X, \quad \text{Var}[Y_1 \mid X = x] - \text{Var}[Y_0 \mid X = x] \neq 0.
\]

Table 2 presents the results of this experiment, which demonstrate the power of our tests for the null hypothesis $\sigma_1^2(x) - \sigma_0^2(x) = 0$, for all $x \in X$. In this case, we can see that our test is quite powerful, regardless of the order of the power series.
Table 1. Probabilities of rejecting $H_0$.

| Sample size | Order of power series | Test with T | Test with Q |
|-------------|-----------------------|-------------|-------------|
|             |                       | 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 $H_0$.

| Sample size | Order of power series | Test with T | Test with Q |
|-------------|-----------------------|-------------|-------------|
|             |                       | Nominal size | 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 |

7. Empirical application

We apply the test approach proposed in this paper to wage data of 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 the tests on 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; see for example, a survey
written by Johnson (1975). Freeman’s (1980) important study led to a renewed investigation of the relationship between unionism and inequality. Analyzing cross-sectional micro data on workers in union and nonunion sectors, Freeman found that unionism reduces white-collar/blue-collar wage differentials in the organized sector, which overwhelms the increase in the dispersion of wages across industries. Then studies were conducted using variants of the framework to illustrate the effect of unions more completely, by considering union coverage rates (DiNardo et al. (1996)), the union effect across different types of workers (Card (2001)), and unobserved skill differences (Lemieux (1993)). Card (2001) divides workers into 10 equal-sized skill groups in the nonunion sector to estimate the overall effect of unions on the variance of wages, relative to what would be observed if all workers were paid according to the existing nonunion wage structure. 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 subpopulations with covariates beyond particular skill characteristics. In the following, we consider an empirical application regarding the effects of unionism on the dispersion of wages using US data. We analyze whether there is significant evidence that unionism in the United States differs in terms of the dispersion of wages between union workers and nonunion workers with particular characteristics.

We analyze data from the National Longitudinal Survey (Youth Sample), which contains data on full-time working males who had completed their schooling by 1980, and who 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 to 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).

As in Crump et al. (2008), we conduct two specifications of the test with covariates selected by the “top-down” and “bottom-up” methods described in Section 5. In both specifications, we specify the 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
### Table 3. Tests for zero conditional average treatment effects and zero conditional variance effects.

| Year | Test for conditional mean | Test for conditional variance |
|------|--------------------------|-------------------------------|
|      | \( Q_{\text{mean}} \) \( (K_1) \) & \( Q_{\text{mepval}} \) & \( T_{\text{mean}} \) & \( T_{\text{mepval}} \) & \( Q_{\text{var}} \) \( (K_2) \) & \( Q_{\text{varpval}} \) & \( T_{\text{var}} \) & \( T_{\text{varpval}} \) |
| 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 |

**top-down**

| Year | Test for conditional mean | Test for conditional variance |
|------|--------------------------|-------------------------------|
| 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 |

**bottom-up**

The 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. This results in \( K_2 \) covariates in this step.

The results of these four versions of the tests are reported in Table 3.\(^5\) We also provide the zero conditional average effect of unionism on wages as proposed by Crump et al. (2008) and our test for a zero effect of unionism on conditional variance. The \( Q \) statistic for the chi-squared test and the \( T \) 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 variable selection procedure. The null hypothesis of a zero conditional average treatment effect is rejected at the 1% level for each year, while for conditional variance, there 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 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 1980, 1982–1983, and 1985–1987 are consistent with the conclusion of Card et al. (2004) that unions reduce the variance of wages for men.

---

\(^5\) In the test of a zero average treatment effect for 1983 data using the “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.
8. Conclusion

In this paper, we developed nonparametric tests for the null hypothesis that the conditional variances of the outcomes for the treatment group and the control group are identical, for all subpopulations defined by the covariates. We gave the 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 1980 to 1987, we find unionism has no treatment effect on the wage dispersion of any subpopulation. 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, $\Omega_{P,1,K} = E[P_K(X)P_K(X)' \mid W = 1]$ is nonsingular for all $K$ (Newey (1994)). We can construct a sequence of basis functions $R_{K_1}(x) = \Omega_{P,1,K_1}^{-1} P_{K_1}(x)$ with $E[R_{K_1}(X)R_{K_1}(X)' \mid W = 1] = I_{K_1}$ and $R_{K_2}(x) = \Omega_{P,1,K_2}^{-1/2} P_{K_2}(x)$ with $E[R_{K_2}(X)R_{K_2}(X)' \mid W = 1] = I_{K_2}$. Below we prove the theorem by the sequence of basis functions, $R_{K_1}(x)$, instead of $P_K(x)$, 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 $N_w \times K_1$ matrix $R_{1,K_r} = (R_{K_1}(X_1)', \ldots, R_{K_r}(X_{N_1}))'$ and $R_{0,K_r} = (R_{K_1}(X_{N_1+1})', \ldots, R_{K_r}(X_{N}))'$. Then, nonparametric series estimator of $\mu_w(x)$, given $K_1$ terms in the series, is given by

$$\hat{\mu}_{w,K_1}(x) = R_{K_1}(x)'(R_{w,K_1}' R_{w,K_1})^{-1}(R_{w,K_1} Y_w) = R_{K_1}(x)' \hat{\gamma}_{w,K_1},$$

where $\hat{\gamma}_{w,K_1} = (R_{w,K_1}' R_{w,K_1})^{-1}(R_{w,K_1}' Y_w)$. Then nonparametric series estimator of $\sigma^2_w(x)$, given $K_2$ terms in the series, is given by

$$\hat{\sigma}^2_{w,K_2}(x) = R_{K_2}(x)'(R_{w,K_2}' R_{w,K_2})^{-1}(R_{w,K_2}' \hat{\epsilon}_w^2) = R_{K_2}(x)' \hat{\alpha}_{w,K_2},$$

where $\hat{\alpha}_{w,K_2} = (R_{w,K_2}' R_{w,K_2})^{-1}(R_{w,K_2}' \hat{\epsilon}_w^2)$.

In addition,

$$\Omega_{w,K_r} \equiv E[R_{K_r}(X)R_{K_r}(X)' \mid W = w],$$

$$\Psi_{w,K_1} \equiv E[\sigma^2_w(X)R_{K_1}(X)R_{K_1}(X)' \mid W = w]$$

and

$$\Psi_{w,K_2} \equiv E[\theta^2_w(X)R_{K_2}(X)R_{K_2}(X)' \mid W = w],$$

and we estimate them by

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

Finally, we have
\[ N \cdot V = \frac{1}{\pi_0} \Omega^{-1}_{0,K_2} \Psi_{0,K_2} \hat{\Omega}^{-1}_{0,K_2} + \frac{1}{\pi_1} \Omega^{-1}_{1,K_2} \Psi_{1,K_2} \hat{\Omega}^{-1}_{1,K_2} \]

and estimator
\[ \hat{\Psi} = \frac{\Omega^{-1}_{0,K_2} \hat{\Psi}_{0,K_2} \hat{\Omega}^{-1}_{0,K_2}}{N_0} + \frac{\Omega^{-1}_{1,K_2} \hat{\Psi}_{1,K_2} \hat{\Omega}^{-1}_{1,K_2}}{N_1} \]

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 \( A \), \( \| A \| = \text{tr}(A' A) \). 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 \( \epsilon^1_i = ((Y_1 - \mu_1(X_1))^2, (Y_2 - \mu_1(X_2))^2, \ldots, (Y_{N_1} - \mu_1(X_{N_1}))^2)' \) and \( \epsilon^2_i = ((Y_{N_1+1} - \mu_2(X_{N_1+1}))^2, (Y_{N_1+2} - \mu_2(X_{N_1+2}))^2, \ldots, (Y_N - \mu_2(X_N))^2)' \). Let \( \epsilon_w = Y_w - \mu_w(X_w) \) and \( u_w = \epsilon^2_w - \sigma^2_w(X_w) \).

First, we establish asymptotic normality for pseudo statistics. We define,
\[ \alpha^*_{w,K_2} \equiv (E[R_{K_2}(X)R_{K_2}(X)' | W = w])^{-1} E[R_{K_2}(X)\epsilon^2_{w,i} | W = w] \]
\[ = \Omega^{-1}_{w,K_2} E[R_{K_2}(X)\epsilon^2_{w,i} | W = w] \]

Now we consider asymptotic normality of \( \sqrt{N_w} \cdot \frac{1}{N_w} \Omega^{-1}_{w,K_2} R'_{w,K_2} u_w \).
\[ \sqrt{N_w} \cdot \frac{1}{N_w} \Omega^{-1}_{w,K_2} R'_{w,K_2} u_w = \frac{1}{\sqrt{N_w}} \sum_{i=1}^N \Omega^{-1}_{w,K_2} 1(W_i = w) R_{K_2}(X_i) u_{w,i}. \]

with
\[ E[\Omega^{-1}_{w,K_2} R'_{w,K_2} u_w] = \Omega^{-1}_{w,K_2} E[1(W_i = w) R_{K_2}(X_i) E[u_{w,i} | X_i, W_i = w]] \]
\[ = \Omega^{-1}_{w,K_2} E[1(W_i = w) R_{K_2}(X_i) E[u_{w,i} | X_i]] = 0, \]

and
\[ \text{Var}[\Omega^{-1}_{w,K_2} R'_{w,K_2} u_w] \]
\[ = \Omega^{-1}_{w,K_2} E[1(W_i = w) u^2_{w,i} R_{K_2}(X_i) R_{K_2}(X_i)'] \Omega^{-1}_{w,K_2} \]
\[ = \Omega^{-1}_{w,K_2} E[1(W_i = w)^2 R_{K_2}(X_i) R_{K_2}(X_i)'] E[u^2_{w,i} | X_i, W_i = w] \Omega^{-1}_{w,K_2} \]
\[ = \Omega^{-1}_{w,K_2} E[1(W_i = w) \theta^2_{w}(X_i) R_{K}(X_i) R_{K}(X_i)'] \Omega^{-1}_{w,K_2} \]
\[ = \Omega^{-1}_{w,K_2} E[\theta^2_{w}(X_i) R_{K}(X_i) R_{K}(X_i)'] | W_i = w] \Omega^{-1}_{w,K_2} \cdot \Pr(W_i = w) \]
\[ = \Omega^{-1}_{w,K_2} E[\theta^2_{w}(X_i) R_{K}(X_i) R_{K}(X_i)'] \Omega^{-1}_{w,K_2} \cdot \pi_w \]
\[ = \Omega^{-1}_{w,K_2} \Psi_{w,K_2} \Omega^{-1}_{w,K_2} \cdot \pi_w. \]
Therefore,
\[
\text{Var} \left[ \sqrt{N_w} \cdot \frac{1}{N_w} \Omega_{w,K}^{-1} R'_{w,K_2} u_w \right] = \frac{1}{N_w} N \cdot \Omega_{w,K_2}^{-1} \Psi_{w,K_2} \Omega_{w,K_2}^{-1} \cdot \pi_w \\
\rightarrow \Omega_{w,K_2}^{-1} \Psi_{w,K_2} \Omega_{w,K_2}^{-1}.
\]

Se define
\[
S_{w,K_2} = \frac{1}{\sqrt{N_w}} \sum_{i=1}^{N} \Omega_{w,K}^{-1}(W_i = w) R_K(X_i) u_{w,i}
\]
\[
= \frac{1}{\sqrt{N_w}} \sum_{i=1}^{N} \left[ \Omega_{w,K_2}^{-1} \Psi_{w,K_2} \Omega_{w,K_2}^{-1} \cdot \pi_w \right]^{-1/2} \Omega_{w,K_2}^{-1} 1(W_i = w) R_K(X_i) u_{w,i}
\]
\[
= \frac{1}{\sqrt{N_w}} \sum_{i=1}^{N} Z_i.
\]

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,
\[
\sup_{A \in A_{K_2}} |Q_{N_w}(A) - \Phi(A)| = o(1),
\]
where \(A_{K_2}\) is the class of all measurable convex sets in \(K_2\)-dimensional Euclidean space and \(\Phi\) 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,
\[
\beta_3 \equiv N_w^{-3/2} \sum_{i=1}^{N} E\|Z_i\|^3
\]
\[
= N_w^{-3/2} \sum_{i=1}^{N} E\left[\|\Omega_{w,K}^{-1} \Psi_{w,K} \Omega_{w,K_2}^{-1} \cdot \pi\|^{-1/2} 1(W_i = w) R_K(X_i) u_{w,i}\|^3\right]
\]
\[
\leq \beta^2 \cdot N_w^{-3/2} \sum_{i=1}^{N} E\left[\|\Omega_{w,K_2}^{-1} \cdot \pi\|^{-1/2} \cdot 1(W_i = w) R_K(X_i) u_{w,i}\|^3\right]
\]
\[
\leq C N^{-3/2} \sum_{i=1}^{N} E\|\Omega_{w,K_2}^{-1/2} R_K(X_i) u_{w,i}\|^3
\]
\[
\leq CN^{-3/2} \sum_{i=1}^{N} E[\lambda_{\max}(\Omega_{w,K}^{-1/2}) R_{K}^{2}(X_i) u_{w,i}]^3 \\
\leq CN^{-3/2} \sum_{i=1}^{N} \lambda_{\max}(\Omega_{w,K}^{-1/2})^3 \zeta(K_2)^3 E|u_{w,i}|^3 \\
\leq C\zeta(K_2)^3 N^{-1/2}.
\]

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, \(\sup_{A \in A_{K_2}} |Q_{N_w}(A) - \Phi(A)| = o(1)\). □

Now, we have a multivariate asymptotic normality, under which we may further proceed a univariate standard Gaussian distribution. We consider the quadratic form \(S'_{w,K} S_{w,K}\),
\[
S'_{w,K} S_{w,K} = \sum_{j=1}^{K_2} \left( \frac{1}{\sqrt{N_w}} \sum_{i=1}^{N} Z_{ij} \right)^2,
\]
where \(Z_{ij}\) is the \(j\)th element of the vector \(Z_i\). Thus, \(S'_{w,K} S_{w,K}\) 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,
\[
\sup_c |\Pr(S'_{w,K} S_{w,K} \leq c) - \chi^2_{K_2}(c)| \to 0.
\]

**Proof.** Define the set \(A(c) \equiv \{S \in \mathbb{R}^{K_2} \mid S' S \leq c\}\), and it is a measurable, convex set in \(\mathbb{R}^{K_2}\). For \(Z \sim N(0, I_{K_2})\),
\[
\sup_c |\Pr(S'_{w,K} S_{w,K} \leq c) - \chi^2_{K_2}(c)| \\
= \sup_c |\Pr(S'_{w,K} S_{w,K} \leq c) - \Pr(Z' Z \leq c)| \\
= \sup_c |\Pr(S_{w,K} \in A(c)) - \Pr(Z' Z \leq c)| \\
\leq \sup_{A \in A_{K_2}} |Q_{N_w}(A) - \Phi(A)| \\
= o(1). \quad \Box
\]

The normalized version of \(S'_{w,K} S_{w,K}\) 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,

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

Proof.

\[
\sup_c \left| \Pr \left( \frac{S'_{w,K_2}S_{w,K_2} - K_2}{\sqrt{2K_2}} \leq c \right) - \Phi(c) \right| = \sup_c \left| \Pr(S'_{w,K_2}S_{w,K_2} \leq K + c\sqrt{2K_2}) - \Phi(c) \right|
\]

\[
\leq \sup_c \left| \Pr(S'_{w,K_2}S_{w,K} \leq K_2 + c\sqrt{2K_2}) - \chi^2(K_2 + c\sqrt{2K_2}) \right|
\]

\[
+ \sup_c \left| \chi^2(K_2 + c\sqrt{2K_2}) - \Phi(c) \right|
\]

\[
= \sup_c \left| \Pr(S'_{w,K_2}S_{w,K_2} \leq K_2 + c\sqrt{2K_2}) - \chi^2(K_2 + c\sqrt{2K_2}) \right|
\]

\[
+ \sup_c \left| \Pr \left( \frac{Z/Z - K_2}{\sqrt{2K_2}} \leq c \right) - \Phi(c) \right|
\]

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 \( \nu_2 < 0 \) it also converges to zero. So the result holds. \( \square \)

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 \( \Omega_{w,K_r} \) 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, \( O_p(\zeta(K_2)K_2^{1/2}N^{-1/2}) = o_p(1) \). \( \square \)

Lemma 5. Suppose Assumptions 2.1–2.3 and 5.1–5.3 hold. Then,

(i) there is a sequence of vector \( \gamma^0_{w,K_1} \), 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) \( \sup_x |\mu_w(x) - R_{K_1}(x)\gamma_{w,K_1}| = \sup_x |\mu_w(x) - \mu_{w,K_1}(x)| = O(\zeta(K_1)K_1^{-s_1/d}) + O_p(\zeta(K_1)K_1^{1/2}N^{-1}) \).

**Proof.** The proofs can be found in Crump et al. (2008) Lemma A.6. Because \( O_p(\zeta(K_1)K_1^{-1/2}N^{-1/2}) = o_p(1) \) 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) \( \sup_x |\sigma_w^2(x) - R_{K_2}(x)\sigma_{w,K_2}^0| = \sup_x |\sigma_w^2(x) - \sigma_{w,K_2}^0(x)| = O(K_2^{-s_2/d}) \),

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

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

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

(v) \( \sup_x |\sigma_w^2(x) - R_{K_2}(x)\hat{\gamma}_{w,K_2}| = \sup_x |\sigma_w^2(x) - \hat{\gamma}_{w,K_2}^2| \equiv \sup_x |\sigma_w^2(x) - \hat{\gamma}_{w,K_2}^2| = O(\zeta(K_1)\zeta(K_2)K_1^{-s_1/d} + \zeta(K_1)\zeta(K_2)K_1^{1/2}N^{-1} + \zeta(K_2)K_2^{1/2}N^{-1/2} + \zeta(K_2)K_2^{-s_2/d}) \).

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

\[
\Pr(|\epsilon_{w,i} - \bar{\epsilon}_{w,i}| > M) \leq \frac{E|\epsilon_{w,i} - \bar{\epsilon}_{w,i}|}{M} \leq \frac{E[(\epsilon_{w,i} + \mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i))^2 - \epsilon_{w,i}^2]}{M} \leq \frac{E|2\epsilon_{w,i}(\mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i))|}{M} + \frac{E|\mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i)|^2}{M} \leq 2 \cdot \sup_i |\mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i)| \cdot \frac{1}{M} E|\epsilon_{w,i}| + \left( \sup_i |\mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i)| \right)^2 \leq C \cdot \sup_i |\mu_w(X_i) - \hat{\mu}_{w,K_1}(X_i)| + O(\zeta(K_1)K_1^{-s_1/d}) + O_p(\zeta(K_1)K_1^{1/2}N^{-1/2})^2 \leq O(\zeta(K_1)K_1^{-s_1/d} + O_p(\zeta(K_1)K_1^{1/2}N^{-1})^2.
\]

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

\[
\|\hat{\alpha}_{w,K_2} - \alpha_{w,K_2}^0\| = \left\| \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1} R_{w,K_2} P_{w,K_2} \hat{\epsilon}_{w}^2 - \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1} R_{w,K_2} P_{w,K_2} \alpha_{w,K_2}^0 \right\| = \left\| \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1} R_{w,K_2} (\hat{\epsilon}_{w}^2 - R_{w,K_2} \alpha_{w,K_2}^0) \right\|
\]
\[
\begin{align*}
&\leq \lambda_{\max}(\hat{\Omega}_{w,K_2}^{-1/2}) \cdot \left\| \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1/2} R_{w,K_2}' (\epsilon_{w}^2 - R_{w,K_2} \alpha_{w,K_2}^0) \right\| \\
&= \left\| \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1/2} R_{w,K_2}' (\hat{\epsilon}_{w}^2 - \epsilon_{w}^2 + \epsilon_{w}^2) \\
&\quad - \sigma_w^2(X_w) + \sigma_w^2(X_w) - R_{w,K_2} \alpha_{w,K_2}^0 \right\| \\
&= \left\| \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1/2} R_{w,K_2}' \left( \epsilon_{w}^2 - \epsilon_{w}^2 \right) \right\| \\
&+ \left\| \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1/2} R_{w,K_2} u_w \right\| \\
&+ \left\| \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1/2} R_{w,K_2} \left( \sigma_w^2(X_w) - R_{w,K_2} \alpha_{w,K_2}^0 \right) \right\| \\
&\leq \left\| \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1/2} R_{w,K_2}' (\hat{\epsilon}_{w}^2 - \epsilon_{w}^2) \right\|^2 \\
&= \frac{1}{N_w} E \left[ (\hat{\epsilon}_{w}^2 - \epsilon_{w}^2)' R_{w,K_2} (R_{w,K_2}' R_{w,K_2})^{-1} R_{w,K_2}' (\hat{\epsilon}_{w}^2 - \epsilon_{w}^2) \right] \\
&\leq \frac{1}{N_w} E \left[ (\hat{\epsilon}_{w}^2 - \epsilon_{w}^2)' (\hat{\epsilon}_{w}^2 - \epsilon_{w}^2) \right] \\
&\leq E \| \epsilon_{w,i}^2 - \epsilon_{w,i}^2 \| \\
&= O(\zeta(K_1)^2 K_1^{-2s_1/d}) + O_p(\zeta(K_1)^2 K_1 N^{-2}).
\end{align*}
\]

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, \( \left\| \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1/2} R_{w,K_2}' (\epsilon_{w}^2 - \epsilon_{w}^2) \right\| = O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2} N^{-1}). \)

For (A.2), we have,

\[
E \left\| \frac{1}{N_w} \hat{\Omega}_{w,K_2}^{-1/2} R_{w,K_2} u_w \right\|^2 \\
= E \left[ \text{tr} \left( \frac{1}{N_w} u_w' R_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} R_{w,K_2} u_w \right) \right] \\
= \frac{1}{N_w} E \left[ \text{tr} (R_{w,K_2} (R_{w,K_2}' R_{w,K_2})^{-1} R_{w,K_2} u_w u_w') \right] \\
= \frac{1}{N_w} \text{tr} (E[R_{w,K_2} (R_{w,K_2}' R_{w,K_2})^{-1} R_{w,K_2}' E[u_w u_w'] | X])
\]

\( \lambda_{\max}(\hat{\Omega}_{w,K}^{-1}) = \lambda_{\max}(\Omega_{w,K}^{-1}) + O_p(\zeta(K)K^{1/2} N^{-1/2}) \)
\[
\leq \theta^2 \cdot \frac{1}{N_w} E[\text{tr}(R_{w,K2}(R'_{w,K2} R_{w,K2})^{-1} R'_{w,K2})] \\
= \theta^2 \cdot \frac{1}{N_w} K_2 \\
\leq CK_2 N^{-1},
\]
and so \( \| \frac{1}{N_w} \hat{\Omega}_{w,K2}^{-1/2} R'_{w,K2} u_w \| = O_p(K_2^{1/2} N^{-1/2}) \).

For (A.3), we have,

\[
\left\| \frac{1}{N_w} \hat{\Omega}_{w,K2}^{-1/2} R'_{w,K2} (\sigma^2_w(X_w) - R_{w,K2} \alpha^0_{w,K2}) \right\|^2 \\
= \frac{1}{N_w} (\sigma^2_w(X_w) - R_{w,K2} \alpha^0_{w,K2})' R_{w,K2} (R'_{w,K2} R_{w,K2})^{-1} \\
\times R_{w,K2} (\sigma^2_w(X_w) - R_{w,K2} \alpha^0_{w,K2}) \\
\leq \frac{1}{N_w} (\sigma^2_w(X_w) - R_{w,K2} \alpha^0_{w,K2})' (\sigma^2_w(X_w) - R_{w,K2} \alpha^0_{w,K2}) \\
\leq \left( \sup_x (\sigma^2_w(x) - R_{w,K2} \alpha^0_{w,K2}) \right)^2 \\
\leq CK_2^{-2s_2/d}
\]
by (i), and so \( \| \frac{1}{N_w} \hat{\Omega}_{w,K2}^{-1/2} R'_{w,K2} (\sigma^2_w(X_w) - R_{w,K2} \alpha^0_{w,K2}) \| = O(K_2^{-s_2/d}) \). Combining these, we have,

\[
\| \hat{\alpha}_{w,K} - \alpha_{w,K} \| = [O(1) + o_p(0)] \\
\cdot [O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2} N^{-1}) + O_p(K_2^{1/2} N^{-1/2}) + O(K_2^{-s_2/d})] \\
= O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2} N^{-1}) + O_p(K_2^{1/2} N^{-1/2}) + O(K_2^{-s_2/d}).
\]
Finally, for (v), we have,

\[
\sup_x |\sigma^2_w(x) - \hat{\sigma}^2_{w,K2}(x)| \leq \sup_x |\sigma^2_w,K_2(x) - \sigma^2_{w,K2}(x)| + \sup_x |\sigma^2_{w,K2}(x) - \hat{\sigma}^2_{w,K2}(x)|.
\]
The first term is \( O(K_2^{-s_2/d}) \) by (i). For the second term, we have

\[
\sup_x |\sigma^2_{w,K2}(x) - \hat{\sigma}^2_{w,K2}(x)| \\
= \sup_x |R_{K2}(x)'(\hat{\alpha}_{w,K2} - \hat{\alpha}_{w,K2})| \\
\leq \sup_x \|R_{K2}(x)\| \cdot \|\hat{\alpha}_{w,K2} - \hat{\alpha}_{w,K2}\| \\
= \zeta(K_2) [O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2} N^{-1}) \\
+ O_p(K_2^{1/2} N^{-1/2}) + O(K_2^{-s_2/d})],
\]
where we use the result from (iv). Thus,

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

**Lemma 7.** Suppose Assumptions 2.1–2.3 and 5.1–5.3 hold. Then,

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

(ii) the eigenvalues of \(\Psi_{w,K_2}\) 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,

$$
\tilde{\Psi}_{w,K_2} = \frac{R'_{w,K_2} \tilde{M}_{w,K_2} R_{w,K_2}}{N_w},
$$

where \(\tilde{M}_{w,K} = \text{diag}\{u^2_{w,i}, \text{ for } i \text{ with } W_i = w\}\). Then,

$$
E\|\hat{\Psi}_{w,K_2} - \Psi_{w,K_2}\|^2 = E\|\hat{\Psi}_{w,K_2} - \hat{\Psi}_{w,K_2} + \hat{\Psi}_{w,K_2} - \Psi_{w,K_2}\|^2
\leq 2E\|\hat{\Psi}_{w,K_2} - \hat{\Psi}_{w,K_2}\|^2 + 2E\|\hat{\Psi}_{w,K_2} - \Psi_{w,K_2}\|^2
\leq 2E\left\|\frac{R'_{w,K_2}(\tilde{M}_{w,K_2} - \tilde{M}_{w,K_2})R_{w,K_2}}{N_w}\right\|^2 + 2E\|\hat{\Psi}_{w,K_2} - \Psi_{w,K_2}\|^2.
$$

(A.4)

(A.5)

Consider (A.4),

$$
E\left\|\frac{R'_{w,K_2}(\tilde{M}_{w,K_2} - \tilde{M}_{w,K_2})R_{w,K_2}}{N_w}\right\|^2
= \frac{1}{N_w^2} \sum_{k=1}^{K_2} \sum_{l=1}^{K_2} \sum_{i=1}^{N} \sum_{j=1}^{N} E[1(W_i = w)1(W_j = w)(\hat{u}^2_{w,i} - u^2_{w,i})
\times R_{k,K_2}(X_i)R_{l,K_2}(X_i)R_{k,K_2}(X_j)R_{l,K_2}(X_j)],
$$
Following the steps in the proof of Lemma A.1(iv) in Crump et al. (2007), we can know the last term is

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

So for (A.4), we have

\[
E \left\| \left( \hat{M}_{w,K_2} - \hat{M}_{w,K_2} \right) R_{w,K_2} \right\|^2 = \frac{1}{N_w^2} \sum_{k=1}^{K_2} \sum_{l=1}^{K_2} \sum_{i=1}^{N} \sum_{j=1}^{N} E[R_{kK_2}(X_i)R_{lK_2}(X_i)R_{kK_2}(X_j)R_{lK_2}(X_j)] = O(K_2).
\]

(9)

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

\[
E \left\| \hat{\Psi}_{w,K_2} - \Psi_{w,K_2} \right\|^2 = O_p(\zeta^2 K_2 N^{-1}).
\]

Combining (9) and (10) yields

\[
\left\| \hat{\Psi}_{w,K_2} - \Psi_{w,K_2} \right\|
\]
\[= [O_p(\zeta(K_1)\zeta(K_2)K_1^{-s_1/d} + \zeta(K_1)\zeta(K_2)K_1^{1/2}N^{-1}) \]
\[+ \zeta(K_2)K_2^{1/2}N^{-1/2} + \zeta(K_2)K_2^{-s_2/d})]O(K_2) \]
\[+ O_p(\zeta(K_2)^2K_2^{1/2}N^{-1/2}) \]
\[= O_p(\zeta(K_1)^2\zeta(K_2)^2K_1^{-s_1/d}K_2 + \zeta(K_2)^2K_2^2K_2^{-s_2/d}) \]
\[+ \zeta(K_1)^2\zeta(K_2)^2K_1^{1/2}K_1^{-s_1/d}K_2N^{-1} \]
\[+ \zeta(K_1)\zeta(K_2)^2K_1^{-s_1/d}K_2K_2^{3/2}N^{-1/2} \]
\[+ \zeta(K_1)\zeta(K_2)^2K_1^{1/2}K_2^{3/2}N^{-3/2} \]
\[+ \zeta(K_2)^2K_2^{1/2}N^{-1/2}). \]

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

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

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

We can prove (iv) as Lemma A.2 in Crump et al. (2008). For (v),
\[\lambda_{\min}(\hat{\Omega}_{w,K_2}^{-1}\hat{\Psi}_{w,K_2}\hat{\Omega}_{w,K_2}^{-1}) \]
\[\geq \lambda_{\min}(\Psi_{w,K_2}) \cdot \lambda_{\min}(\Omega_{w,K_2}^{-1})^2 \]
\[ \geq [\lambda_{\min}(\Psi_{w,K_2}) + o_p(1)]\lambda_{\min}(\Omega_{w,K_2}^{-1})^2 \]
\[ \geq [C + o_p(1)][C + o_p(1)]^2 \]
\[ = C + o_p(1). \]

Thus,
\[ \lambda_{\min}(N \cdot \hat{V}) = \min_{d'd=1} d' \left( \frac{N}{N_0} \hat{\Omega}_{0,K_2}^{-1} \hat{\Psi}_{0,K_2} \hat{\Omega}_{0,K_2}^{-1} + \frac{N}{N_1} \hat{\Omega}_{1,K_2}^{-1} \hat{\Psi}_{1,K_2} \hat{\Omega}_{1,K_2}^{-1} \right) d \]
\[ \geq \frac{N}{N_0} \min_{d'd=1} d'(\hat{\Omega}_{0,K_2}^{-1} \hat{\Psi}_{0,K_2} \hat{\Omega}_{0,K_2}^{-1})d + \frac{N}{N_1} \min_{d'd=1} d'(\hat{\Omega}_{1,K_2}^{-1} \hat{\Psi}_{1,K_2} \hat{\Omega}_{1,K_2}^{-1})d \]
\[ = \lambda_{\min}(\hat{\Omega}_{0,K_2}^{-1} \hat{\Psi}_{0,K_2} \hat{\Omega}_{0,K_2}^{-1}) + \lambda_{\min}(\hat{\Omega}_{1,K_2}^{-1} \hat{\Psi}_{1,K_2} \hat{\Omega}_{1,K_2}^{-1}) \]

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,
\[ N_w(\hat{\alpha}_{w,K_2} - \alpha_{w,K_2}^*)' [\hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1}]^{-1} (\hat{\alpha}_{w,K_2} - \alpha_{w,K_2}^*) - K_2 \]
\[ \overset{d}{\rightarrow} N(0, 1). \]

**Proof.** Because Lemma 3 holds, we need only show that
\[ (A.8) \quad \| (\hat{\Omega}_{0,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{0,K_2}^{-1})^{1/2} N_w(\hat{\alpha}_{w,K_2} - \alpha_{w,K_2}^*) - S_{w,K_2} \| = o_p(1). \]

Let \( u_{w,K_2}^* \equiv \epsilon_{w,K_2}^2 - R_{w,K_2} \alpha_{w,K_2}^* \), then
\[ \hat{\alpha}_{w,K_2} - \alpha_{w,K_2}^* = (R'_{w,K_2} R_{w,K_2})^{-1}(R'_{w,K_2}(R_{w,K_2} \alpha_{w,K_2}^* + u_{w,K_2}^*)) - \alpha_{w,K_2}^* \]
\[ = \frac{1}{N_w} \cdot \hat{\Omega}_{w,K_2}^{-1} R'_{w,K_2} u_{w,K_2} \]

The left side of (A.8) can be written as
\[ (A.9) \quad \leq \frac{1}{\sqrt{N_w}} \| (\hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1})^{1/2} \hat{\Omega}_{w,K_2}^{-1} R_{w,K_2} u_{w,K_2} \|
\]
\[ - [\hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1}]^{1/2} \hat{\Omega}_{w,K_2}^{-1} R_{w,K_2} u_{w,K_2} \|
\]
\[ (A.10) \quad + \frac{1}{\sqrt{N_w}} \| [\hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1}]^{1/2} \hat{\Omega}_{w,K_2}^{-1} R_{w,K_2} u_{w,K_2} \|
\]
\[ - [\hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1}]^{1/2} \hat{\Omega}_{w,K_2}^{-1} R_{w,K_2} u_{w,K_2} \|
\]
\[ (A.11) \quad + \frac{1}{\sqrt{N_w}} \| [\hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1}]^{1/2} \hat{\Omega}_{w,K_2}^{-1} R_{w,K_2} u_{w,K_2} \|\]
For (A.9),

$$\frac{1}{\sqrt{N_w}} \left| \left[ \Omega_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right]^{-1/2} \hat{\Omega}_{w,K_2}^{-1/2} R'_{w,K_2} u_w^* \right| - \left[ \Omega_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right]^{-1/2} \hat{\Omega}_{w,K_2}^{-1/2} R'_{w,K_2} u_w^* \right| \leq \lambda_{\max} \left( \hat{\Omega}_{w,K_2}^{-1/2} \hat{\Psi}_{w,K_2}^{-1/2} \right) \lambda_{\max} \left( \Omega_{w,K_2}^{-1/2} K_2^{-1/2} \right) \leq \left( [C + o_p(1)] \right)^2 K_2^{1/2} = O_p(K_2^{1/2}).$$

For the second factor we have,

$$E \left| \frac{\hat{\Omega}_{w,K_2}^{-1/2} R'_{w,K_2} (u_w^* - u_w)}{\sqrt{N_w}} \right|^2 \leq N_w \left( \sup_x |\sigma_w^2(x) - R_{K_2}(x)' \alpha_{w,K_2}^0| + \sup_x |R_{K_2}(x)' \alpha_{w,K_2}^0 - R_{K_2}(x)' \alpha_{w,K_2}^*| \right)^2$$

$$= N_w \left( \sup_x |\sigma_w^2(x) - R_{K_2}(x)' \alpha_{w,K_2}^0| + \sup_x |R_{K_2}(x)' \alpha_{w,K_2}^0 - R_{K_2}(x)' \alpha_{w,K_2}^*| \right)^2$$

Then, equation (A.9) is

$$\frac{1}{\sqrt{N_w}} \left| \left[ \Omega_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right]^{-1/2} \hat{\Omega}_{w,K_2}^{-1/2} R'_{w,K_2} u_w^* \right| - \left[ \Omega_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right]^{-1/2} \hat{\Omega}_{w,K_2}^{-1/2} R'_{w,K_2} u_w^* \right| \leq \lambda_{\max} \left( \hat{\Omega}_{w,K_2}^{-1/2} \hat{\Psi}_{w,K_2}^{-1/2} \right) \lambda_{\max} \left( \Omega_{w,K_2}^{-1/2} K_2^{-1/2} \right) \leq \left( [C + o_p(1)] \right)^2 K_2^{1/2} = O_p(K_2^{1/2}).$$
For the first factor in (A.12), we consider

$$\frac{1}{\sqrt{N_w}} \left| \left| \hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right| - \left| \left| \hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right| - \left| \left| \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \right| \r
Combining (A.13) and (A.14) together yields,
\[
\frac{1}{\sqrt{N_w}} \left[ \hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right]^{-1/2} \hat{\Omega}_{w,K_2}^{-1} R'_{w,K_2} u_w \\
- \left[ \Omega_{w,K_2}^{-1} \Psi_{w,K_2} \Omega_{w,K_2}^{-1} \right]^{-1/2} \Omega_{w,K_2}^{-1} R'_{w,K_2} u_w \\
= \left[ O_p(\zeta(K_1)^2 \zeta(K_2)^2 K_1^{-2s_1/d} K_2^{3/2} + \zeta(K_2)^2 K_2^{-s_2/d} \\
+ \zeta(K_1)^2 \zeta(K_2)^2 K_1^{1/2} K_2^{-s_1/d} K_2^{3/2} N^{-1} \\
+ \zeta(K_1) \zeta(K_2)^2 K_1^{-s_1/d} K_2^{1/2} N^{-1/2}) \right] O_p(K_2^{1/2}) \\
= o_p(1)
\]
under Assumption 5.2 and 5.3. Finally, (A.11) is
\[
(A.15) \quad \frac{1}{\sqrt{N_w}} \left[ \hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right]^{-1/2} \hat{\Omega}_{w,K_2}^{-1} R'_{w,K_2} u_w \\
- \left[ \Omega_{w,K_2}^{-1} \Psi_{w,K_2} \Omega_{w,K_2}^{-1} \right]^{-1/2} \Omega_{w,K_2}^{-1} R'_{w,K_2} u_w \\
\leq \left\| \left[ \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right]^{-1/2} \hat{\Omega}_{w,K_2}^{-1} - \Omega_{w,K_2}^{-1} \right\| \left\| \frac{1}{\sqrt{N_w}} R'_{w,K_2} u_w \right\|.
\]
The first factor in (A.15) is \( \left\| \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right\| = C \left\| I_{K_2} \right\| = O(K_2^{1/2}) \). 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
\[
E \left\| \frac{1}{\sqrt{N_w}} R'_{w,K_2} u_w \right\|^2 \\
= E \left[ \frac{1}{\sqrt{N_w}} \text{tr}(u'_w R_{w,K_2} R'_{w,K_2} u_w) \right] \\
= E \left[ \frac{1}{\sqrt{N_w}} \text{tr}(R'_{w,K_2} u_w u'_w R_{w,K_2}) \right] \\
= \text{tr} \left( \frac{1}{\sqrt{N_w}} E[R'_{w,K_2} E[u_w u'_w | X_w] R_{w,K_2}] \right) \\
= \delta^2 w \text{tr}(\Omega_{w,K_2}) \\
= \delta^2 w K_2 \lambda_{\max}(\Omega_{w,K_2}) \\
= C K_2.
\]
Thus \( \left\| \frac{1}{\sqrt{N_w}} R'_{w,K_2} u_w \right\| = O_p(K_2^{1/2}) \). Putting these together, we have
\[
\frac{1}{\sqrt{N_w}} \left[ \hat{\Omega}_{w,K_2}^{-1} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1} \right]^{-1/2} \hat{\Omega}_{w,K_2}^{-1} R'_{w,K_2} u_w \\
- \left[ \Omega_{w,K_2}^{-1} \Psi_{w,K_2} \Omega_{w,K_2}^{-1} \right]^{-1/2} \Omega_{w,K_2}^{-1} R'_{w,K_2} u_w \\
= O_p(\zeta(K_2)^3 N^{-1/2}) = o_p(1).
\]
Hence,
\[
\|\hat{\Omega}_{0,K_2}^{-1}_{w,K_2} \hat{\Psi}_{w,K_2} \hat{\Omega}_{w,K_2}^{-1}_{d} - 1/2 \sqrt{N_w} (\hat{\alpha}_{w,K_2} - \alpha_{w,K_2}^*) - S_{w,K_2} \| = o_p(1).
\]
Thus, the infeasible test statistic converges to standard normal distribution. \qed

Now, we prove Theorem 1.

**Proof.** Define
\[
\hat{\delta}_{K_2} = \hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}, \quad \delta_{K_2}^* = \alpha_{1,K_2}^* - \alpha_{0,K_2}^*.
\]
Following the logic of Lemma 8 above, we can conclude that
\[
T^* = \frac{(\hat{\delta}_{K_2} - \delta_{K_2}^*)' \hat{\Omega}^{-1} (\hat{\delta}_{K_2} - \delta_{K_2}^*) - K_2}{\sqrt{2K_2}} \xrightarrow{d} N(0,1).
\]
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 \(\alpha_{1,K_2}^0 = \alpha_{0,K_2}^0\) for \(\sigma_{1,K_2}^0(x) = \sigma_{0,K_2}^0(x)\). Then,
\[
\|\alpha_{1,K_2}^0 - \alpha_{0,K_2}^0\| = \|\alpha_{1,K_2}^0 - \alpha_{1,K_2}^0 + \alpha_{0,K}^0 - \alpha_{0,K_2}^0\|
\leq \|\alpha_{1,K_2}^0 - \alpha_{1,K_2}^0\| + \|\alpha_{0,K}^0 - \alpha_{0,K_2}^0\|
= O(K_2^{1/2} K_2^{-s_2/d}).
\]
by Lemma 6 (ii) and
\[
\|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\| = \|\hat{\alpha}_{1,K_2} - \alpha_{1,K_2}^0 + \alpha_{0,K}^0 - \hat{\alpha}_{0,K_2}\|
\leq \|\hat{\alpha}_{1,K_2} - \alpha_{1,K_2}^0\| + \|\alpha_{0,K}^0 - \hat{\alpha}_{0,K_2}\|
= O_p[\zeta(K_1) K_1^{-s_1/d} + \zeta(K_1) K_1^{1/2} N^{-1}] + O_p(K_2^{1/2} N^{-1/2})
+ O(K_2^{-s_2/d}) + K_2^{1/2} N^{-1/2} + K_2^{-s_2/d})
\]
by Lemma 6 (iii). Thus,
\[
|T^* - T| = \frac{\|\hat{\delta}_{K_2} - \delta_{K_2}^*\| \hat{\Omega}^{-1} (\hat{\delta}_{K_2} - \delta_{K_2}^*) - K_2}{\sqrt{2K_2}} \xrightarrow{d} \frac{\hat{\delta}_{K_2} - \delta_{K_2}^*}{\sqrt{2K_2}}
\]
(A.16) \quad \leq \frac{2}{\sqrt{2K_2}} |(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})' \hat{\Omega}^{-1} (\alpha_{1,K_2}^* - \alpha_{0,K_2}^*)|

(A.17) \quad + \frac{1}{\sqrt{2K_2}} |(\alpha_{1,K_2}^* - \alpha_{0,K_2}^*)' \hat{\Omega}^{-1} (\alpha_{1,K_2} - \alpha_{0,K_2})|.

For (A.16),
\[
\frac{2}{\sqrt{2K_2}} |(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})' \hat{\Omega}^{-1} (\alpha_{1,K_2}^* - \alpha_{0,K_2}^*)|
\]
\[
\frac{N}{\sqrt{2K_2}}N \cdot |\text{tr}((\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})' [N\hat{V}]^{-1}(\alpha^*_{1,K_2} - \alpha^*_{0,K_2}))| \\
\leq \frac{2}{\sqrt{2K_2}}N \cdot \lambda_{\max}([N\hat{V}]^{-1})\|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\|\|\alpha^*_{1,K_2} - \alpha^*_{0,K_2}\| \\
= \frac{2}{\sqrt{2K_2}}N \cdot [C + o_p(1)] \cdot [O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2}N^{-1}) \\
+ O_p(K_2^{1/2}N^{-1/2}) + O(K_2^{-s_2/d}) \\
+ K_2^{1/2}N^{-1/2} + K_2^{-s_2/d})][O(K_2^{1/2}K_2^{-s_2/d})] \\
= O_p(\zeta(K_1)K_1^{-s_1/d}K_2^{-s_2/d}N + \zeta(K_1)K_1^{1/2}K_2^{-s_2/d} \\
+ K_2^{1/2}K_2^{-s_2/d}N^{1/2} + K_2^{-2s_2/d}N).
\]

For (A.17),
\[
\frac{1}{\sqrt{2K_2}} |(\alpha^*_{1,K_2} - \alpha^*_{0,K_2})' \hat{V}^{-1}(\alpha^*_{1,K_2} - \alpha^*_{0,K_2})| \\
= \frac{1}{\sqrt{2K_2}} \cdot N \cdot |\text{tr}((\alpha^*_{1,K_2} - \alpha^*_{0,K_2})' [N\hat{V}]^{-1}(\alpha^*_{1,K_2} - \alpha^*_{0,K_2}))| \\
\leq \frac{1}{\sqrt{2K_2}} \cdot N \cdot \lambda_{\max}([N\hat{V}]^{-1})\|\alpha^*_{1,K_2} - \alpha^*_{0,K_2}\|^2 \\
= N \cdot [C + o_p(1)][O(K_2^{1/2}K_2^{-s_2/d})]^2 \\
= O(K_2^{1/2}K_2^{-2s_2/d}N).
\]

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

by 5.2 and 5.3. Hence the result follows. \(\square\)

Now, we prove Theorem 2.

**Proof.**
\[
\rho_N \cdot \sup_{x \in \mathcal{X}} |\Delta(x)| = \sup_{x \in \mathcal{X}} |\sigma^2_1(x) - \sigma^2_0(x)| \\
\leq \sup_{x \in \mathcal{X}} |R_{K_2}(x)'\alpha^0_{1,K_2} - \sigma^2_1(x)| + \sup_{x \in \mathcal{X}} |R_{K_2}(x)'\alpha^0_{0,K_2} - \sigma^2_0(x)| \\
+ \sup_{x \in \mathcal{X}} |R_{K_2}(x)'\hat{\alpha}_{1,K_2} - R_{K_2}(x)'\alpha^0_{1,K_2}| \\
+ \sup_{x \in \mathcal{X}} |R_{K_2}(x)'\hat{\alpha}_{0,K_2} - R_{K_2}(x)'\alpha^0_{0,K_2}| \\
+ \sup_{x \in \mathcal{X}} |R_{K_2}(x)'\hat{\alpha}_{1,K_2} - R_{K_2}(x)'\hat{\alpha}_{0,K_2}|.
\]
Thus,

\[ \| \hat{\alpha}_{1,K_2} - \alpha_{0,K_2} \| \]
\[ \geq \zeta(K_2)^{-1} \cdot \rho_N \cdot \sup_{x \in \mathcal{X}} |\Delta(x)| - \zeta(K_2)^{-1} \cdot \rho_N \cdot \sup_{x \in \mathcal{X}} |R_{K_2}(x)'\alpha_{1,K_2}^0 - \sigma_1^2(x)| \]
\[ - \zeta(K_2)^{-1} \cdot \rho_N \cdot \sup_{x \in \mathcal{X}} |R_{K_2}(x)'\alpha_{0,K_2}^0 - \sigma_0^2(x)| \]
\[ - \| \hat{\alpha}_{1,K_2} - \alpha_{1,K_2}^0 \| - \| \hat{\alpha}_{0,K_2} - \alpha_{0,K_2}^0 \| \]
\[ \geq \zeta(K_2)^{-1} \cdot \rho_N \cdot C_0 \left( 1 - \frac{\sup_{x \in \mathcal{X}} |R_{K_2}(x)'\alpha_{1,K_2}^0 - \sigma_1^2(x)|}{\rho_N \cdot C_0} \right. \]
\[ - \frac{\sup_{x \in \mathcal{X}} |R_{K_2}(x)'\alpha_{0,K_2}^0 - \sigma_0^2(x)|}{\rho_N \cdot C_0} \]
\[ \left. - \zeta(K_2) \frac{\| \hat{\alpha}_{1,K_2} - \alpha_{1,K_2}^0 \|}{\rho_N \cdot C_0} - \zeta(K_2) \frac{\| \hat{\alpha}_{0,K_2} - \alpha_{0,K_2}^0 \|}{\rho_N \cdot C_0} \right). \]

Under Assumption 5.1 and 5.3, we have

\[ \sup_{x \in \mathcal{X}} |R_{K_2}(x)'\alpha_{1,K_2}^0 - \sigma_1^2(x)| \]
\[ \frac{1}{\rho_N \cdot C_0} = O(K_2^{-s_2/d}) \cdot O(N^{1/2-3\nu/21-3\nu_2/2-\varepsilon}) = o(1), \]
\[ \sup_{x \in \mathcal{X}} |R_{K_2}(x)'\alpha_{0,K_2}^0 - \sigma_0^2(x)| \]
\[ \frac{1}{\rho_N \cdot C_0} = O(K_2^{-s_2/d}) \cdot O(N^{1/2-3\nu/21-3\nu_2/2-\varepsilon}) = o(1), \]
\[ \zeta(K_2) \frac{\| \hat{\alpha}_{1,K_2} - \alpha_{1,K_2}^0 \|}{\rho_N \cdot C_0} = O(K_2) \cdot [O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2}N^{-1}) \]
\[ + O_p(K_2^{1/2}N^{-1/2}) + O(K_2^{-s_2/d})] \]
\[ \cdot O(N^{1/2-3\nu/21-3\nu_2/2-\varepsilon}) = o(1), \]
\[ \zeta(K_2) \frac{\| \hat{\alpha}_{0,K_2} - \alpha_{0,K_2}^0 \|}{\rho_N \cdot C_0} = O(K_2) \cdot [O_p(\zeta(K_1)K_1^{-s_1/d} + \zeta(K_1)K_1^{1/2}N^{-1}) \]
\[ + O_p(K_2^{1/2}N^{-1/2}) + O(K_2^{-s_2/d})] \]
\[ \cdot O(N^{1/2-3\nu/21-3\nu_2/2-\varepsilon}) = o(1). \]
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

\[
N^{1/2} \zeta(K_1)^{-3/2} K_2^{-1/2} \zeta(K_2)^{-1} \cdot \rho_N \cdot C_0 \\
\geq C N^{1/2} \zeta(K_1)^{-3/2} K_2^{-1/2} \zeta(K_2)^{-1} \cdot N^{-1/2 + 3\nu_1/2 + 3\nu_2/2 + \varepsilon} \geq C N^\varepsilon,
\]

for any \(M'\),

(A.18) \[\Pr[N^{1/2} \zeta(K_1)^{-3/2} K_2^{-1/2} \|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\| > M'] \to 1.\]

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 \(\lambda\) and \(\bar{\lambda}\) be the minimum and maximum eigenvalues of \([NV]^{-1}\), respectively. \(\lambda\) is bounded away from 0 and \(\bar{\lambda}\) is bounded.

\[
\Pr\left(\frac{\tilde{C} \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})' [NV]^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M\right)
\]

\[
= \Pr\left(\tilde{C} N \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})' [NV]^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) > \sqrt{2K_2}M + K_2\right)
\]

\[
\geq \Pr(\tilde{\lambda} C N \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})' (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) > \sqrt{2K_2}M + K_2)
\]

\[
= \Pr(N \zeta(K_1)^{-3/2} K_2^{-1/2} (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})' (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) > (\tilde{\lambda} C)^{-1} \zeta(K_2)^{-3} (\sqrt{2MK_2^{-1/2}} + 1))
\]

\[
= \Pr(N^{1/2} \zeta(K_1)^{-3/2} K_2^{-1/2} \|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\| > (\tilde{\lambda} C)^{-1/2} \zeta(K_2)^{-3/2} (\sqrt{2MK_2^{-1/2}} + 1)^{1/2}).
\]

Since for any \(M\), for large enough \(N\), we have

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

it follows that for large \(N\),

\[
\Pr\left(\frac{\tilde{C} \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})' V^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M\right)
\]

\[
\leq \Pr(N^{1/2} \zeta(K_1)^{-3/2} K_2^{-1/2} \|\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}\| > 2(\tilde{\lambda} C)^{-1/2})
\]

(A.19) \[\to 1.\]
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 \( \hat{\lambda} \) be the minimum eigenvalue of \([N\hat{V}]^{-1}\). Let \( B_1 \) denote the event that \( \hat{\lambda} > \lambda/2 \). \( \Pr(B_1) \to 1 \) as \( N \to \infty \) by lemma 7. In addition, let \( B_2 \) be the event
\[
\left( \frac{\lambda}{2} \right) N(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'([N\hat{V}]^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2) > M.
\]

\[
\Pr \left( \frac{\hat{\lambda} \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'\hat{V}^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M \right)
= \Pr \left( \frac{\hat{\lambda} \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'\hat{V}^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M \right)
\leq \Pr \left( \frac{\hat{\lambda} \cdot (\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'\hat{V}^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}} > M \right)
\to 1
\]
as \( N \to \infty \) by (A.19). Let \( \hat{\lambda} \) be \( \lambda/2 \hat{\lambda}^{-1} \), 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
\[
T = \frac{(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'\hat{V}^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2}{\sqrt{2K_2}}
\geq \frac{\hat{\lambda} N(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'([N\hat{V}]^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2)}{\sqrt{2K_2}}
\geq \frac{(\lambda/2) N(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2})'([N\hat{V}]^{-1}(\hat{\alpha}_{1,K_2} - \hat{\alpha}_{0,K_2}) - K_2)}{\sqrt{2K_2}}
> M.
\]

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

**Acknowledgements**

I thank my supervisors, Yoshihiko Nishiyama and Ryo Okui, for their continuing guidance and support. I also appreciate the insightful comments from the editor, two anonymous referees, and the 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.
References

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models, *J. Am. Stat. Assoc.*, **97**(457), 284–292.

Abadie, A., Angrist, J. and Imbens, G. (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings, *Econometrica*, **70**(1), 91–117.

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

Bentkus, V. (2005). A Lyapunov-type bound in $R^d$, *Theory Probab. Appl.*, **49**(2), 311–323.

Card, D. (2001). The effect of unions on wage inequality in the US labor market, *Ind. Labor Relat. Rev.*, **54**(2), 296–315.

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

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

Chernozhukov, V. and Hansen, C. (2005). An IV model of quantile treatment effects, *Econometrica*, **73**(1), 245–261.

Crump, R. K., Hotz, V. J., Imbens, G. W. and Mitnik, O. A. (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., Hotz, V. J., Imbens, G. W. and Mitnik, O. A. (2008). Nonparametric tests for treatment effect heterogeneity, *Rev. Econ. Stat.*, **90**(3), 389–405.

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

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

Firpo, S. (2007). Efficient non-parametric estimation of quantile treatment effects, *Econometrica*, **75**(1), 259–276.

Freeman, R. B. (1980). Unionism and the dispersion of wages, *Ind. Labor Relat. Rev.*, **34**(1), 3–23.

Gosling, A. and Machin, S. (1995). Trade unions and the dispersion of earnings in British establishments, 1980–90, *Oxf. Bull. Econ. Stat.*, **57**(2), 167–184.

Hong, Y. and White, H. (1995). Consistent specification testing via nonparametric series regression, *Econometrica*, **63**(5), 1133–1159.

Hsu, Y.-C. (2017). Consistent tests for conditional treatment effects, *The Econometrics Journal*, **20**(1), 1–22.

Imbens, G. W., Newey, W. K. and Ridder, G. (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, *Am. Econ. Rev.*, **65**(2), 23–28.

Lee, M.-J. (2009). Non-parametric tests for distributional treatment effect for randomly censored responses, *J. R. Stat. Soc. Ser. B (Statistical Methodology)*, **71**(1), 243–264.

Lemieux, T. (1993). Unions and wage inequality in Canada and the United States, *Small Differences that Matter: Labor Markets and Income Maintenance in Canada and the United States* (eds. D. Card and R. B. Freeman), pp. 69–108, University of Chicago Press.

Newey, W. K. (1994). The asymptotic variance of semiparametric estimators, *Econometrica*, **62**(6), 1349–1382.

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

Ng, S. and Perron, P. (1995). Unit root tests in arma models with data-dependent methods for the selection of the truncation lag, *J. Am. Stat. Assoc.*, **90**(429), 268–281.

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

Song, Q. and Yang, L. (2009). Spline confidence bands for variance functions, *J. Nonparametr. Stat.*, **21**(5), 589–609.
Vella, F., Verbeek, M. et al. (1998). Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men, *J. Appl. Econ.*, **13**(2), 163–183.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., MIT Press.

Yu, K. and Jones, M. (2004). Likelihood-based local linear estimation of the conditional variance function, *J. Am. Stat. Assoc.*, **99**(465), 139—144.

Zhu, L., Dong, Y. and Li, R. (2013). Semiparametric estimation of conditional heteroscedasticity via single-index modeling, *Statistica Sinica*, **23**(3), 1235–1255.