Nonparametric Tests for Treatment Effect Heterogeneity*

Richard K. Crump† V. Joseph Hotz‡ Guido W. Imbens§ Oscar Mitnik¶

November 2005

Abstract

JEL Classification: C14, C21, C52

Keywords: *Average Treatment Effects, Causality, Unconfoundedness, Treatment Effect Heterogeneity*

*We are grateful for helpful comments by, and for financial support from the National Science Foundation through grant SES 0452590.

†Department of Economics, University of California, Berkeley, crump@econ.berkeley.edu.

‡Department of Economics, University of California, Los Angeles, hotz@econ.ucla.edu, http://www.econ.ucla.edu/hotz/.

§Department of Agricultural and Resource Economics, and Department of Economics, University of California, Berkeley, 661 Evans Hall, Berkeley, CA 94720-3880, imbens@econ.berkeley.edu, http://elsa.berkeley.edu/users/imbens/.

¶Department of Economics, University of Miami, omitnik@exchange.sba.miami.edu.

1 Introduction

A large part of the literature on program evaluation focuses on estimation of the average effect for the entire population or the population of treated individuals under assumptions of unconfoundedness or ignorability following the seminal work by Rubin (1973) and Rosenbaum and Rubin (1983). See Heckman, Lalonde and Smith (2000), Rosenbaum (2001), Wooldridge (2002), Imbens (2004) and Lee (2005) for surveys of this literature. The literature on testing for the presence of treatment effects in this context is much smaller.[1] In many cases researchers are interested in the effects of programs beyond point estimates of the overall average or the average for the subpopulation of treated individuals. For example, it may be of substantive interest to investigate whether there is any subpopulation for which a program or treatment has a nonzero average effect, or whether there is heterogeneity in the average effects by subpopulation. In fact, some of this interest has motivated the development of estimators for quantile treatment effects (Lehman, 1974; Doksum, 1974; Firpo, 2004).

The hypothesis that the average effect of the treatment is zero for all subpopulations is also of interest in assessing assumptions concerning selection mechanisms. In their discussion of specification tests as a tool to obtain better estimators for average treatment effects Heckman and Hotz (1989) introduced an important class of specification tests that can be interpreted as tests of the null hypothesis of zero causal effects on lagged outcomes. Heckman and Hotz focused on methods that specifically test the hypothesis of a zero effect under the maintained assumption that the effect is constant. However, the motivation for these tests suggests that the null hypothesis of interest could also be interpreted as that of zero effects for all subpopulations. Similarly, Rosenbaum (1997) discusses the use of multiple control groups to investigate the plausibility of unconfoundedness. He shows that if both control groups satisfy unconfoundedness, differences in average outcomes adjusted for difference in covariates should be close to zero. Again the hypothesis of interest can be formulated as one of zero causal effects for all subpopulations, not just a zero average effect.

In this paper we develop two nonparametric tests. The first test is for the null hypothesis that the treatment has a zero average effect for any subpopulation defined by covariates. The second test is for the null hypothesis that the average effect is identical for all subpopulations, in other words, that there is no heterogeneity in average treatment effects by covariates. Sacrificing some generality by focusing on these two specific null hypotheses we derive tests that are straightforward to implement. They are based on a series or sieve approach to nonparametric estimation for average treatment effects (Imbens, Newey and Ridder, 2004; Chen, Hong, and Tarozzi, 2004). Given the particular choice of the sieve the null hypotheses of interest can be formulated as equality restrictions on subsets of the parameters. The tests can then be implemented using standard parametric methods: the test statistics are quadratic forms in the differences in the parameter estimates with chi-squared critical values. We provide conditions on the sieves that guarantee that in large samples the tests are valid without the parametric

[1]There is a large literature on testing in the context of randomized experiments using the randomization distribution. See Rosenbaum (2001). There are also a small number of papers developing tests in other settings, e.g., Abadie (2002) in the context of instrumental variables models.

assumptions.

There is a larger literature on testing parametric restrictions on regression functions against nonparametric alternatives. Härdle and Marron (1990) study tests of parametric restrictions on comparisons of two regression functions. Their formal analysis is restricted to the case with a single regressor, although it is likely that their kernel methods can be adapted (in particular by using higher order kernels) to extend to the case with multivariate covariates. Härdle and Mammen (1993) and Horowitz and Spokoiny (2001) focus on tests of parametric models for regression functions against nonparametric alternatives. Their set up can be extended to test parametric restrictions on differences of regression functions. However, the focus in this paper on two specific tests, zero and constant conditional average treatment effects, rather than on general parametric restrictions, makes the proposed tests particularly easy to implement compared to the Härdle-Marron, Härdle-Mammen and Horowitz-Spokoiny tests. In particular p-values for our proposed tests can be based on chi-squared or normal tables, whereas Härdle and Mammen (1993) require the use of a variation of the bootstrap they call the wild bootstrap, and Horowitz and Spokoiny (2001) require simulation to calculate the p-value

We illustrate these methods using data from the Greater Avenue for INdependence (GAIN) experiments carried out in California in the nineties.

2 Set Up

Our basic framework uses the motivating example of testing zero treatment effects in a program evaluation setting, although our tests can be used more generally to test the hypotheses of constant or zero differences between regression functions estimated on separate samples. The set up is standard in the program evaluation literature and based on the potential outcome notation popularized by Rubin (1974). We have a random sample of size N from a large population. For each unit i in the sample, let W_i indicate whether the treatment of interest was received, with $W_i = 1$ if unit i receives the treatment of interest, and $W_i = 0$ if unit i receives the control treatment. Let $Y_i(0)$ denote the outcome for unit i under control and $Y_i(1)$ the outcome under treatment. We observe W_i and Y_i, where

$$Y_i \equiv Y_i(W_i) = W_i \cdot Y_i(1) + (1 - W_i) \cdot Y_i(0).$$

In addition, we observe a vector of pre-treatment variables, or covariates, denoted by X_i. Define the two conditional means, $\mu_w(x) = \mathbb{E}[Y(w)|X = x]$, the two conditional variances, $\sigma_w^2(x) = \text{Var}(Y(w)|X = x)$, the conditional average treatment effect $\tau(x) = \mathbb{E}[Y(1) - Y(0)|X = x] = \mu_1(x) - \mu_0(x)$, and the propensity score, the probability of selection into the $e(x) = \Pr(W = 1|X = x) = \mathbb{E}[W|X = x]$.

To solve the identification problem, we maintain throughout the paper the unconfoundedness assumption (Rosenbaum and Rubin, 1983), which asserts that conditional on the pre-treatment variables, the treatment indicator is independent of the potential outcomes. Formally:

Assumption 2.1 (Unconfoundedness)

$$W \perp (Y(0), Y(1)) \,\Big|\, X. \tag{2.1}$$

In addition we assume there is overlap in the covariate distributions:

Assumption 2.2 (Overlap)
For some $c > 0$,

$$c \leq e(x) \leq 1 - c.$$

In addition for estimation we often need smoothness conditions on the two regression functions $\mu_w(x)$ and the propensity score $e(x)$.

3 Testing

3.1 Introduction

In this section we discuss some statistical tests. We focus on two different hypotheses concerning the conditional average treatment effect $\tau(x)$. The first pair of hypotheses we consider

$$\mathrm{H}_0: \ \forall\, x \in \mathbb{X},\ \tau(x) = 0, \qquad \mathrm{H}_a: \ \exists\, x \in \mathbb{X},\ \text{s.t. } \tau(x) \neq 0. \tag{3.2}$$

Under the null hypothesis that average effect of the treatment is zero for all values of the covariates, whereas under the alternative there are some values of the covariates for which the effect of the treatment differs from zero.

The second pair of hypotheses is

$$\mathrm{H}_0': \ \exists\, \tau \text{ s.t. } \forall\, x \in \mathbb{X},\ \tau(x) = \tau, \qquad \mathrm{H}_a': \ \forall \tau, \exists\, x \in \mathbb{X},\ \text{s.t. } \tau(x) \neq \tau. \tag{3.3}$$

We refer to this pair as the null hypothesis of no treatment effect heterogeneity. Strictly speaking this is not correct, as we only require the average effect of the treatment to be equal to zero for all values of the covariates, allowing for distributional effects that average out to zero.

We want to contrast these hypotheses with the pair of hypotheses corresponding to zero average effect,

$$\mathrm{H}_0'': \ \mathbb{E}[\tau(X)] = 0, \qquad \mathrm{H}_a'': \ \mathbb{E}[\tau(X)] = 0. \tag{3.4}$$

Tests of the null hypothesis of a zero average effect are more commonly done, either explicitly, or implicitly through estimating the average treatment effect and its standard error. It is obviously much less restrictive than the null hypothesis of a zero conditional average effect.

The motivation for considering the two pairs of hypotheses over and above considering the hypothesis of a zero average effect consists of three parts. The first is substantive. In many cases the primary interest of the researcher may be in establishing whether the average effect of the program differs from zero. However, even if it is zero on average, there may well be subpopulations for which the effect is substantively and statistically significant. As a first step towards establishing this it can be useful to test whether there is any statistical evidence against the hypothesis that the effect of the program is zero on average for all subpopulations (the pair of hypotheses H_0 and H_a). If one finds that there is compelling evidence that the program has nonzero effect for some subpopulations, one may then further investigate which subpopulations these are, and whether the effects for these subpopulations are substantively important. As an alternative strategy one could directly estimate average effects for substantively interesting subpopulations, but there may be many such subpopulations and it can be difficult to control size when testing many null hypotheses. Our proposed strategy of a single test for zero conditional average treatment effects avoids such problems.

Second, irrespective of whether one finds evidence in favor or against a zero average treatment effect one may be concerned with the question whether there is heterogeneity in the average effect conditional on the observed covariates. If there is strong evidence in favor of heterogenous effects one may be more concerned about recommending extending the program to different populations.

The third motivation is very different. In much of the economic literature on program evaluation there is much concern about the unconfoundedness assumption. If individuals choose whether or not to participate in the program based on information that is not all observed by the researcher, it may well be that conditionally on observed covariates there is some correlation between potential outcomes and the treatment indicator as ruled out by the unconfoundedness assumption. This assumption is not directly testable. However, there are two specific sets of tests available that are suggestive of the plausibility of the assumption. Both are based on testing the effect of a pseudo treatment which is known to have no effect. The first set of tests was originally suggested by Heckman and Hotz (1989). Let us partition the vector of covariates X into two parts, a scalar V and the remainder Z, so that $X = (V, Z')'$. The idea is to take the data $(\mathbf{V}, \mathbf{W}, \mathbf{Z})$ and analyze them as if V is the outcome, W is the treatment indicator, and unconfoundedness holds conditional on Z. Since V is a pretreatment variable or covariate, we know that the effect of the treatment on V is zero for all units. If we find statistical evidence in favor of an effect of the treatment on V it must therefore be the case that the assumption of unconfoundedness conditional on Z is incorrect. This is of course not direct evidence against unconfoundedness conditional on $X = (V, Z')'$, but at the very least it suggests that unconfoundedness is a delicate assumption in this case with the presence of V essential. What makes tests of this type particularly effective is if the researcher observes a number of lagged values of the outcome. In that case one can choose V to be the one-period lagged value of the outcome. If conditional on further lags and individual characteristics one finds differences in lagged outcome distributions for those who will be treated in the future and those who will not be, it calls into question whether conditioning on all lagged outcome values will be sufficient to eliminate differences between control and treatment groups. Heckman and Hotz (1989) implement these tests by testing whether the average effect of the treatment is equal to zero, testing the pair of hypotheses in (3.4). Clearly, in this setting it would be stronger evidence in support of the unconfoundedness assumption to find that the effect of the treatment on the lagged outcome is zero for all values of Z. This corresponds to implementing tests of the pairs of hypotheses (3.2).

A similar set of issues comes up in Rosenbaum's (1997) discussion of the use of multiple controls groups. Rosenbaum considers a setting with two distinct potential control groups. He suggests that if biases one may be concerned with would likely be different for both groups, then evidence that the two control groups lead to similar estimates is suggestive that unconfoundedness may be appropriate. One can implement this idea by comparing the two control groups. Let $W_i = 1$ if unit i is from the treatment group, $W_i = 0$ if unit i is from the first control group and $W_i = -1$ if unit i is from the second control group. Suppose unconfoundedness holds it for both control groups. Formally, $(Y_i(0), Y_i(1)) \perp W_i | X_i, W_i \in \{0, 1\}$ (unconfoundedness relative to first control group) and $(Y_i(0), Y_i(1)) \perp W_i | X_i, W_i \in \{-1, 1\}$ (unconfoundedness relative to second control group). Then it is likely that in fact $(Y_i(0), Y_i(1)) \perp W_i | X_i$. This implies that $Y_i(0) \perp W_i | X_i, W_i \in \{-1, 0\}$ and thus $Y_i \perp W_i | X_i, W_i \in \{-1, 0\}$. This last conditional independence relation is directly testable. To carry out the test analyze the subsample with $W_i \in \{-1, 0\}$ as if $D_i = 1\{W_i = 0\}$ is a treatment indicator. If we find evidence that this pseudo treatment has a systematic effect on the outcome, it must be that for at least one of the two control groups unconfoundedness is violated. As in the Heckman-Hotz setting, the pair of hypotheses to test is that of a zero conditional average treatment effect, (3.2).

In the next section we discuss implementing the two tests in a parametric framework. In Section 3.3 we then provide conditions under which these tests can be interpreted as nonparametric tests.

3.2 Tests in Parametric Models

Here we discuss parametric versions of the tests. Suppose the regression functions are specified as

$$\mu_w(x) = \alpha_w + \beta_w' h(x),$$

for some fixed vector of functions of the covariates $h(x)$, with dimension K. The simplest case is $h(x) = x$ where we just estimate a linear model. We can estimate α_w and β_w using least squares:

$$(\hat{\alpha}_w, \hat{\beta}_w) = \arg\min \sum_{i|W_i=w} \left(Y_i - \alpha_w - \beta_w' h(X_i)\right)^2. \tag{3.5}$$

Under general heteroskedasticity, with $V(Y(w)|X) = \sigma_w^2(X)$, the normalized covariance matrix of $(\hat{\alpha}_w, \hat{\beta}_w')'$ is

$$\Omega_w = N \cdot \left(\sum_{i=1}^N h(X_i)h(X_i)'\right)^{-1} \sum_{i=1}^N \sigma_w^2(X_i)h(X_i)h(X_i)' \left(\sum_{i=1}^N h(X_i)h(X_i)'\right)^{-1}. \tag{3.6}$$

In large samples,

$$\sqrt{N} \begin{pmatrix} \hat{\alpha}_0 - \alpha_0 \\ \hat{\beta}_0 - \beta_0 \\ \hat{\alpha}_1 - \alpha_1 \\ \hat{\beta}_1 - \beta_1 \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} \Omega_0 & 0 \\ 0 & \Omega_1 \end{pmatrix}\right) \tag{3.7}$$

Let $\hat{\Omega}_0$ nd $\hat{\Omega}_1$ be consistent estimators for Ω_0 and Ω_1. In this parametric setting the first null hypothesis is

$$H_0: \ \alpha_0 = \alpha_1, \beta_0 = \beta_1,$$

with the alternative hypothesis

$$H_a: \ \alpha_0 \neq \alpha_1, \text{ or } \beta_0 \neq \beta_1.$$

This can be tested using the quadratic form

$$T = \begin{pmatrix} \hat{\alpha}_0 - \hat{\alpha}_1 \\ \hat{\beta}_0 - \hat{\beta}_1 \end{pmatrix}' (\hat{\Omega}_0 + \hat{\Omega}_1)^{-1} \begin{pmatrix} \hat{\alpha}_0 - \hat{\alpha}_1 \\ \hat{\beta}_0 - \hat{\beta}_1 \end{pmatrix}. \tag{3.8}$$

Under the null hypothesis this test statistic has in large samples a chi-squared distribution with $K+1$ degrees of freedom:

$$T \xrightarrow{d} \mathcal{X}^2(K+1). \tag{3.9}$$

We can also arrive at this test statistic in a different way. Instead of carrying out two regressions, onsider a single regression of the outcome on a constant, the treatment indicator, the covariates, and the interaction of the covariates with the treatment indicator:

$$\mathbb{E}[Y|X=x, W=w] = \alpha_0 + \gamma \cdot w + \beta_0' x + \delta' x \cdot w.$$

The coefficients in this regression satisfy $\gamma = \alpha_1 - \alpha_0$ and $\delta = \beta_1 - \beta_0$. Hence testing the null hypothesis $\alpha_0 = \alpha_1$, $\beta_0 = \beta_1$ is equivalent to testing $\gamma = 0$, $\delta = 0$. The advantage is that now the test is one of testing zero restrictions in a regression framework, and thus straightforward to implement using standard software.

The second test is very similar. The original null and alternative hypothesis (3.3) and (**??**) translate into

$$H_0': \ \beta_0 = \beta_1,$$

with the alternative hypothesis

$$H_a': \ \beta_0 \neq \beta_1.$$

Partition Ω_w into the part corresponding to the variance for $\hat{\alpha}_w$ and the part corresponding to variance for $\hat{\beta}_w$:

$$\Omega_w = \begin{pmatrix} \Omega_{w,00} & \Omega_{w,01} \\ \Omega_{w,10} & \Omega_{w,11} \end{pmatrix},$$

and partition $\hat{\Omega}_0$ and $\hat{\Omega}_1$ similarly. The test statistic is now

$$T' = \begin{pmatrix} \hat{\beta}_0 - \hat{\beta}_1 \end{pmatrix}' (\hat{\Omega}_{0,11} + \hat{\Omega}_{1,11})^{-1} \begin{pmatrix} \hat{\beta}_0 - \hat{\beta}_1 \end{pmatrix}. \tag{3.10}$$

Under the null hypothesis this test statistic has in large samples a chi-squared distribution with $K+1$ degrees of freedom:

$$T' \xrightarrow{d} \mathcal{X}^2(K). \tag{3.11}$$

Both these tests are entirely straightforward. The next section shows how these testing procedures can be used to do nonparametric tests.

3.3 Nonparametric Estimation of Regression Functions

In order to develop nonparametric extensions of the tests developed in section 3.2 we need nonparametric estimators for the two regression functions. We use the series estimator for the regression function $\mu_w(x)$ developed by Imbens, Newey and Ridder (2004) and Chen, Hong and Tarozzi (2004). Let K denote the number of terms in the series. As the basis we use power series. Let $\lambda = (\lambda_1, ..., \lambda_d)$ be a multi-index of dimension d, that is, a d-dimensional vector of non-negative integers, with norm $|\lambda| = \sum_{k=1}^d \lambda_k$, and let $x^\lambda = x_1^{\lambda_1} \ldots x_d^{\lambda_d}$. Consider a series $\{\lambda(r)\}_{r=1}^\infty$ containing all distinct vectors such that $|\lambda(r)|$ is nondecreasing. Let $p_r(x) = x^{\lambda(r)}$, where $P_r(x) = (p_1(x), ..., p_r(x))'$. Given the assumptions below the expectation $\Omega_K = \mathbb{E}[P_K(X)P_K(X)'|W = 1]$ is nonsingular for all K. Hence we can construct a sequence $R_K(x) = \Omega_K^{-1/2} P_K(x)$ with $\mathbb{E}[R_K(X)R_K(X)'|W = 1] = I_K$. Let $R_{kK}(x)$ be the kth element of the vector $R_K(x)$. It will be convenient to work with this sequence of basis function $R_K(x)$. The nonparametric series estimator of the regression function $\mu_w(x)$, given K terms in the series, is given by:

$$\hat{\mu}_w(x) = r^K(x)' \left(\sum_{W_i=w} R_K(X_i)R_K(X_i)' \right)^- \sum_{W_i=w} R_K(X_i)Y_i = R_K(x)'\hat{\gamma}_{w,K},$$

where

$$\hat{\gamma}_{w,K} = \left(\sum_{W_i=w} R_K(X_i)R_K(X_i)' \right)^- \sum_{W_i=w} R_K(X_i)Y_i.$$

Define the $N_w \times K$ matrix $R_{w,K}$ with rows equal to $R_K(X_i)$ for units with $W_i = w$, and Y_w to be the N_w vector with elements equal to Y_i for the same units, so that $\hat{\gamma}_{w,K} = (R'_{w,K}R_{w,K})^{-1}(R'_{w,K}Y_w)$. Note that we use A^- here to denote a generalized inverse of A.

Given the estimator $\hat{\mu}_{w,K}(x)$ we estimate the error variance σ_w^2 as

$$\hat{\sigma}^2_{w,K} = \frac{1}{N_w} \sum_{i|W_i=w} (Y_i - \hat{\mu}_{w,K}(X_i))^2.$$

Let $\Omega_{w,K}$ be the limiting variance of $\sqrt{N}\hat{\gamma}_{w,K}$ as the sample size increases for fixed K. We estimate this variance as

$$\hat{\Omega}_{w,K} = \hat{\sigma}^2_{w,K} \cdot (R'_{w,K}R_{w,K}/N)^{-1}.$$

We make the following assumptions.

Assumption 3.1 (Distribution of Covariates)
$X \in \mathbb{X} \subset \mathbb{R}^d$, where $\mathbb{X}$ 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 $\mathbb{X}$.

Assumption 3.2 (Propensity Score)
(i) The propensity score is bounded away from zero and one.
(ii) The propensity score is s times continuously differentiable.

Assumption 3.3 (CONDITIONAL OUTCOME DISTRIBUTIONS)
(i) The two regression functions $\mu_w(x)$ *are* t *times continuously differentiable.*
(ii) the conditional variance of $Y_i(w)$ *given* $X_i = x$ *is equal to* σ_w^2.

Assumption 3.4 (RATES FOR SERIES ESTIMATORS)
$K = N^{\nu}$, *with* $< \nu <$.

We assume homoskedasticty, although this assumption is not essential and can be relaxed to allow the conditional variance to depend on x, as long as it is bounded from above and below.

3.4 Nonparametric Tests

In this section we show how the tests discussed in Section 3.2 based on parametric regression functions can be used to test the null hypothesis against the alternative hypothesis given in (3.2) without the parametric model. Essentially we are going to provide conditions under we can apply a sequence of parametric tests identical to those discussed in Section 3.2 and obtain a test that is valid without the parametric specification.

First we focus on tests of the null hypothesis that the conditional average treatment effect $\tau(x)$ is zero for all values of the covariates, (3.2). To test this hypothesis we compare estimators for $\mu_1(x)$ and $\mu_0(x)$. Given our use of series estimators we can compare the estimated parameters $\hat{\gamma}_{0,K}$ and $\hat{\gamma}_{1,K}$. Specifically, we use as the test statistic for the test of the null hypothesis H_0 the normalized quadratic form

$$T = \left((\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K})' (\hat{\Omega}_{1,K} + \hat{\Omega}_{0,K})^{-1} (\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K}) - K \right) / \sqrt{2K}. \tag{3.12}$$

Theorem 3.1 *Suppose Assumptions ??-?? hold. Then if* $\tau(x) = 0$ *for all* $x \in \mathbb{X}$,

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

Proof: See Appendix.

To understand the result it is useful to decompose the difference $\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K}$ into three parts. Define the pseudo-true values $\gamma^*_{w,K}$, for $w = 0, 1$, as

$$\gamma^*_{w,K} = \arg\min \mathbb{E}\left[R_K(X) R_K(X)' | W = w\right]^{-1} \mathbb{E}\left[R_K(X) Y | W = w\right],$$

so that for fixed K, as $N \to \infty$, $\hat{\gamma}_{w,K} \to \gamma^*_{w,K}$. Then

$$\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K} = \left(\gamma^*_{1,K} - \gamma^*_{0,K}\right) + \left(\hat{\gamma}_{1,K} - \gamma^*_{1,K}\right) - \left(\hat{\gamma}_{0,K} - \gamma^*_{0,K}\right).$$

For fixed K, in large samples, the last two terms are normally distributed, centered around zero. The asymptotic distribution of T is based on this normality. It ignores the first term, the difference $\left(\gamma^*_{1,K} - \gamma^*_{0,K}\right)$. For fixed K this difference is not equal to zero even if $\mu_0(x) = \mu_1(x)$ because the covariate distributions differ in the two treatment groups. To see the justification for ignoring the difference in large samples, recall that under the null hypothesis $\mu_0(x) = \mu_1(x)$ for all x. For large enough K it must be that $\mu_w(x)$ is close to $R_K(x)\gamma_{w,K}$ for all x. Hence it follows that for large enough K it must be that for all x, $R_K(x)(\gamma_{1,K} - \gamma_{0,K})$ is close to zero, implying $\gamma_{0,K}$ and $\gamma_{1,K}$ must be close. The formal result then shows that we can increase K fast enough to make this difference small, while at the same time increasing K slowly enough to maintain the close approximation of the distribution of $\hat{\gamma}_{w,K} - \gamma^*_{w,K}$ by a normal one. A key result here is Theorem 1.3 in Götze (1991) that ensures that convergence to multivariate normality is fast enough to hold even with the dimension of the vector increasing.

In large samples the test statistic has a standard normal distribution if the null hypothesis is correct. However, we would only reject the null hypothesis if the two regression functions are far apart, which corresponds to large positive values of the test statistic. Hence we recommend using critical values for the test based on one-sided tests.

In practice we may wish to modify the testing procedure slightly. Instead of calculating T we can calculate the quadratic form

$$Q = (\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K})'(\hat{\Omega}_{1,K} + \hat{\Omega}_{0,K})^{-1}(\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K}) = \sqrt{2K} \cdot T + K,$$

and compare this to the critical values of a Chi-squared distribution with K degrees of freedom. In large samples this would lead to approximately the same decision rule since $(Q - K)/\sqrt{2K}$ is approximately standard normal if Q has a Chi-squared distribution with degrees of freedom equal to K for large K. This modification would make the testing procedure identical to the one discussed in Section 3.2, which is what one would do if the parametric model

$$\mu_w(x) = R_K(x)'\gamma_{w,K},$$

is correctly specified. This makes the tests particularly simple to apply. However, in large samples the tests do not rely on the correct specification, instead relying on the increasingly flexible specification as K increases with the sample size.

Next, we consider tests of the null hypothesis (3.3) against the alternative hypothesis given in (**??**). For this test we partition $\hat{\gamma}_{w,K}$ as

$$\hat{\gamma}_{w,K} = \begin{pmatrix} \hat{\gamma}_{w0,K} \\ \hat{\gamma}_{w1,K} \end{pmatrix},$$

and the matrix $\hat{\Omega}_{w,K}$ as

$$\hat{\Omega}_{w,K} = \begin{pmatrix} \hat{\Omega}_{w,00,K} & \hat{\Omega}_{w,01,K} \\ \hat{\Omega}_{w,10,K} & \hat{\Omega}_{w,11,K} \end{pmatrix}.$$

The test statistic is then:

$$T' = \Big((\hat{\gamma}_{11,K} - \hat{\gamma}_{01,K})'(\hat{\Omega}_{1,11,K} + \hat{\Omega}_{0,11,K})^{-1}(\hat{\gamma}_{11,K} - \hat{\gamma}_{01,K}) - (K-1)\Big) \Big/ \sqrt{2(K-1)}. \quad (3.13)$$

Theorem 3.2 *Suppose Assumptions ??-?? hold. Then if $\tau(x) = \tau_0$ for some τ_0 and for all $x \in \mathbb{X}$,*

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

Proof: See Appendix.

In practice we may again wish to use the Chi-squared approximation. Now we calculate the quadratic form

$$Q = (\hat{\gamma}_{11,K} - \hat{\gamma}_{01,K})'(\hat{\Omega}_{1,11,K} + \hat{\Omega}_{0,11,K})^{-1}(\hat{\gamma}_{11,K} - \hat{\gamma}_{01,K}) = \sqrt{2(K-1)} \cdot T + K - 1,$$

and compare this to the critical values of a Chi-squared distribution with $K - 1$ degrees of freedom.

3.5 Comparison with Other Tests

Our approach differs from that in other discussions of testing of parametric null hypotheses against nonparametric alternatives such as Härdle and Marron (1990), Härdle and Mammen (1993) and Horowitz and Spokoiny (2001). Both those papers focus on integrated or averaged squared differences of the estimated parametric and nonparametric regression functions. The limiting distribution of the test statistic is not a standard distribution and requires a variation on the bootstrap (Härdle and Mammen, 1993) or simulation (Horowitz and Spokoiny, 2001). In contrast, the limiting distribution of the test statistics in our case is normal or chi-squared, so p-values can be obtained easily from standard tables.

To further understand the relation between the tests, and in fact to motivate our proposed test statistic, consider the simple average of the squared difference between the two regression functions:

$$S = \frac{1}{N} \sum_i ((\hat{\mu}_1(X_i) - \hat{\mu}_0(X_i))^2 \,.$$

Given the series regression estimator, $\hat{\mu}_w(x) = R_K(x)'\hat{\gamma}_w$ so that

$$S = \frac{1}{N} \sum_i (\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K})' R_K(X_i) R_K(X_i)'(\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K})$$

$$= (\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K})' \frac{1}{N} \sum_i \left(R_K(X_i) R_K(X_i)' \right) (\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K}).$$

Hence the averaged squared difference leads to a quadratic form in the difference in the estimated parameters. Within the class of test statistics that are of the form

$$S' = (\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K})' \Sigma (\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K}),$$

a natural choice for the weight matrix Σ is the inverse of the variance of the difference $\hat{\gamma}_{1,K} - \hat{\gamma}_{0,K}$, and this gives us the statistic T in (3.12).

4 Application

4.1 The GAIN Data

In the empirical section of this paper we use data from the Great Avenue to INdependence (GAIN) experiments. These experimental evaluations of job training and job search assistance programs took place in the 1990's in a number of locations in California. As the evaluations are based on formal randomization within the sites, evaluations of the within-site effects are straightforward.

We conduct the tests for zero and constant average treatment effects using data from four locations, Los Angeles, Riverside, Alameda and San Diego with 1400, 1040, 601, and 1154 observations. For each individual we have individual background characteristics, including gender, age, ethnicity (hispanic, black, or other), an indicator for high school graduation, an indicator for the presence of exactly one child (all individuals have at least one child), and an indicator for the presence of children under the age of 6, and ten quarters of earnings data. Table 1 presents summary statistics for the covariates by site. All earnings data are in thousands of dollars per quarter. The levels of earnings are averages for the subsamples with positive earnings only.

4.2 Tests of Zero and Constant Average Treatment Effcts

We then carry out the tests developed in this paper. We implement the tests by including all seven individual characteristics linearly plus a quadratic term for age, plus all ten quarterly earnings variables and ten indicators for zero earnings in each quarter. This leads to a total of twenty eight covariates in the regressions, plus an intercept.

The first test is for the null hypothesis of $\tau(x) = 0$. We include in the set of covariates all 28 covariates listed in Table 1. The chi-squared version of the test has 29 degrees of freedom. For three out of the four counties we get a clear rejection at the 5% level, with only the test statistic for LA smaller than conventional critical values.

The second test is for the null hypothesis of a constant average treatment effect. Again we reject the null at conventional levels for three out of the four counties.

For comparison purposes we also include a simple test for the null hypothesis that the average effect of the treatment is equal to zero. This test is based on the statistic calculated as the difference in average outcomes for the treatment and control groups divided by the standard error of this difference. Here it is interesting that the test fails to reject for three out of the four counties, with only the Riverside data leading to a clear rejection of a zero average treatment effect.

The combination of the tests suggests that these training programs had statistically highly significant effects on earnings, with the effects varying considerably by observable characteristics. A simple analysis that focused exclusively on average effects for the overall population would have missed effects for two out of the four counties.

5 Conclusion

Appendix

Before proving Theorem 3.2 we present a couple of preliminary results.

Lemma A.1 *Suppose Assumptions XX-XX hold. Then* (i)

$$\left\|\hat{\Omega}_{w,K} - \Omega_{w,K}\right\| = O_p\left(\zeta(K)K^{\frac{1}{2}}N^{-\frac{1}{2}}\right),$$

and (ii) *The eigenvalues of* $\Omega_{w,K}$ *are bounded and bounded away from zero and* (iii) *The eigenvalues of* $\hat{\Omega}_{w,K}$ *are bounded and bounded away from zero if* $O_p\left(\zeta(K)K^{\frac{1}{2}}N^{-\frac{1}{2}}\right) = o_p(1)$.

Proof: We will generalize the proof found in Imbens, Newey and Ridder (2004). For (i) we will show

$$\mathbb{E}\left[\left\|\hat{\Omega}_{w,K} - \Omega_{w,K}\right\|^2\right] \leq C \cdot \zeta(K)^2 K/N$$

so that the result follows by Markov's inequality.

$$\begin{aligned}
& \mathbb{E}\left[\left\|\hat{\Omega}_{w,K} - \Omega_{w,K}\right\|^2\right] \\
= \; & \mathbb{E}\left[\left\|\left(R'_{w,K}R_{w,K}/N_w\right) - \Omega_{w,K}\right\|^2\right] \\
= \; & \mathbb{E}\left[\operatorname{tr}\left(\left(R'_{w,K}R_{w,K}/N_w\right) - \Omega_{w,K}\right)'\left(\left(R'_{w,K}R_{w,K}/N_w\right) - \Omega_{w,K}\right)\right] \\
= \; & \mathbb{E}\left[\operatorname{tr}\left(R'_{w,K}R_{w,K}R'_{w,K}R_{w,K}/N_w^2 - \Omega_{w,K}\left(R'_{w,K}R_{w,K}/N_w\right) - \left(R'_{w,K}R_{w,K}/N_w\right)\Omega_{w,K} + \Omega^2_{w,K}\right)\right] \\
= \; & \operatorname{tr}\left(\mathbb{E}\left[R'_{w,K}R_{w,K}R'_{w,K}R_{w,K}/N_w^2\right] - \Omega_{w,K}\mathbb{E}\left[R'_{w,K}R_{w,K}/N_w\right] - \mathbb{E}\left[R'_{w,K}R_{w,K}/N_w\right]\Omega_{w,K} + \Omega^2_{w,K}\right) \\
= \; & \operatorname{tr}\left(\mathbb{E}\left[R'_{w,K}R_{w,K}R'_{w,K}R_{w,K}/N_w^2\right] - 2\cdot\Omega^2_{w,K} + \Omega^2_{w,K}\right) \\
= \; & \operatorname{tr}\left(\mathbb{E}\left[R'_{w,K}R_{w,K}R'_{w,K}R_{w,K}/N_w^2\right]\right) - \operatorname{tr}\left(\Omega^2_{w,K}\right)
\end{aligned}$$

The second term is

$$\operatorname{tr}(\Omega^2_{w,K}) = \sum_{k=1}^{K}\sum_{l=1}^{K}\left(\mathbb{E}\left[R_{kK}(X)R_{lK}(X)|W=w\right]\right)^2 \tag{A.1}$$

The first term is

$$\begin{aligned}
& \operatorname{tr}\left(\mathbb{E}\left[R'_{w,K}R_{w,K}R'_{w,K}R_{w,K}\right]/N_w^2\right) \\
= \; & \mathbb{E}\left[\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\sum_{i|W_i=w}^{N}R_{kK}(X_i)R_{lK}(X_i)\right)^2\right]/N_w^2 \\
= \; & \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i|W_i=w}^{N}\sum_{j|W_j=w}^{N}\mathbb{E}\left[R_{kK}(X_i)R_{lK}(X_i)R_{lK}(X_j)R_{kK}(X_j)|W=w\right]/N_w^2
\end{aligned}$$

We can then partition this expression into terms with $i=j$,

$$\sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i|W_i=w}^{N}\mathbb{E}\left[R_{kK}(X_i)^2R_{lK}(X_i)^2|W=w\right]/N_w^2 \tag{A.2}$$

and with terms $i \neq j$,

$$N_w(N_w-1)\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\mathbb{E}\left[R_{kK}(X)R_{lK}(X)|W=w\right]\right)^2/N_w^2 \tag{A.3}$$

Combining equations (1), (2) and (3) yields,

$$
\begin{aligned}
\mathbb{E}\left[\left\|\hat{\Omega}_{w,K}-\Omega_{w,K}\right\|^2\right] &= \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i|W_i=w}^{N}\mathbb{E}\left[R_{kK}(X_i)^2R_{lK}(X_i)^2|W=w\right]/N_w^2 \\
&\quad +N_w(N_w-1)\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\mathbb{E}\left[R_{kK}(X)R_{lK}(X)|W=w\right]\right)^2/N_w^2 \\
&\quad -\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\mathbb{E}\left[R_{kK}(X)R_{lK}(X)|W=w\right]\right)^2 \\
&= \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i|W_i=w}^{N}\mathbb{E}\left[R_{kK}(X_i)^2R_{lK}(X_i)^2|W=w\right]/N_w^2 \\
&\quad -\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\mathbb{E}\left[R_{kK}(X)R_{lK}(X)|W=w\right]\right)^2/N_w \\
&< \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i|W_i=w}^{N}\mathbb{E}\left[R_{kK}(X_i)^2R_{lK}(X_i)^2|W=w\right]/N_w^2 \\
&= \frac{1}{N_w^2}\sum_{i|W_i=w}^{N}\mathbb{E}\left[\sum_{k=1}^{K}R_{kK}(X_i)^2\sum_{l=1}^{K}R_{lK}(X_i)^2|W=w\right] \\
&\leq \frac{1}{N_w^2}\sum_{i|W_i=w}^{N}\zeta(K)^2\cdot\mathbb{E}\left[\sum_{l=1}^{K}R_{lK}(X_i)^2|W=w\right] \\
&= \frac{1}{N_w^2}\sum_{i|W_i=w}^{N}\zeta(K)^2\cdot\sum_{l=1}^{K}\mathbb{E}\left[R_{lK}(X_i)^2|W=w\right] \\
&= \frac{1}{N_w}\zeta(K)^2\cdot\operatorname{tr}\left(\Omega_{w,K}\right) \\
&\leq \frac{1}{N_w}\zeta(K)^2\cdot K\cdot\lambda_{max}\left(\Omega_{w,K}\right) \\
&\leq C\cdot K\zeta(K)^2/N
\end{aligned}
$$

where the fifth line follows by

$$\zeta(K)=\sup_x\|R_K(x)\|=\sup_x\left(\sum_{k=1}^{K}R_{kK}^2(x)\right)^{\frac{1}{2}}$$

which then implies that

$$\sum_{k=1}^{K}R_{kK}^2(x)\leq\zeta(K)^2.$$

The eighth line follows since the maximum eigenvalue of $\Omega_{w,K}$ is $O(1)$ (see below).

For (ii), let us first show that for any two positive semi-definite matrices A and B, and conformable vectors c and d, if $A\geq B$ in a positive semi-definite sense, then for

$$\lambda_{min}(A)=\min_{c'c=1}c'Ac=c^{*\prime}Ac^*,\quad \lambda_{min}(B)=\min_{d'd=1}d'Bd=d^{*\prime}Bd^*,$$

we have that,

$$\lambda_{min}(A) = c^{*\prime}Ac^* \geq c^{*\prime}Bc^* \geq d^{*\prime}Bd^* = \lambda_{min}(B).$$

Now, let $f_w(x) = f_{X|W}(x|W = w)$ and recall that $\Omega_{w,K} = \mathbb{E}\left[R_K(X)R_K(X)'|W = w\right]$ where $\Omega_{1,K}$ is normalized to equal I_K. Next define

$$c(x) = f_0(x)/f_1(x)$$

and note that by Assumptions 8.1 and 8.2 we have that

$$0 < \underline{c} \leq c(x) \leq \bar{c}.$$

Thus we may define $c(x) \equiv \underline{c} + \tilde{c}(x)$ so that,

$$\begin{aligned}
\Omega_{0,K} &= \mathbb{E}\left[R_K(x)R_K(x)'|W = 0\right] \\
&= \int R_K(x)R_K(x)' f_0(x)dx \\
&= \int R_K(x)R_K(x)' c(x) f_1(x)dx \\
&= \int R_K(x)R_K(x)' \left(\underline{c} + \tilde{c}(x)\right) f_1(x)dx \\
&= \underline{c} \int R_K(x)R_K(x)' f_1(x)dx + \int R_K(x)R_K(x)'\tilde{c}(x) f_1(x)dx \\
&= \underline{c} \cdot \Omega_{1,K} + \int R_K(x)R_K(x)'\tilde{c}(x) f_1(x)dx \\
&= \underline{c} \cdot \Omega_{1,K} + \tilde{C}
\end{aligned}$$

Where $\tilde{C}$ is a positive semi-definite matrix. Thus,

$$\Omega_{0,K} \geq \underline{c} \cdot \Omega_{1,K} \Rightarrow \lambda_{min}\left(\Omega_{0,K}\right) \geq \underline{c} \cdot \lambda_{min}\left(\Omega_{1,K}\right) = \underline{c}$$

so that the minimum eigenvalue of $\Omega_{0,K}$ is bounded away from zero. The minimum eigenvalue of $\Omega_{1,K}$ is bounded away from zero by construction. Then note that for a positive definite matrix A, $1/\lambda_{min}(A) = \lambda_{max}(A^{-1})$, so that the eigenvalues of $\Omega_{w,K}$ are also bounded from above.

For (iii) consider the minimum eigenvalue of $\hat{\Omega}_{w,K}$:

$$\begin{aligned}
\lambda_{min}\left(\hat{\Omega}_{w,K}\right) &= \min_{c'c=1} c'\left(\hat{\Omega}_{w,K}\right)c \\
&= \min_{c'c=1}\left(c'\left(\Omega_{w,K}\right)c + c'\left(\hat{\Omega}_{w,K} - \Omega_{w,K}\right)c\right) \\
&\geq \min_{c'c=1} c'\left(\Omega_{w,K}\right)c + \min_{d'd=1} d'\left(\hat{\Omega}_{w,K} - \Omega_{w,K}\right)d \\
&= \lambda_{min}\left(\Omega_{w,K}\right) + \lambda_{min}\left(\hat{\Omega}_{w,K} - \Omega_{w,K}\right) \\
&\geq \lambda_{min}\left(\Omega_{w,K}\right) - \left\|\hat{\Omega}_{w,K} - \Omega_{w,K}\right\| \\
&= \lambda_{min}\left(\Omega_{w,K}\right) - O_p\left(\zeta(K)K^{\frac{1}{2}}N^{-\frac{1}{2}}\right)
\end{aligned}$$

Where the fifth line follows since for a symmetric matrix A

$$\|A\|^2 = \operatorname{tr}\left(A^2\right) \geq \lambda_{min}\left(A^2\right) = \lambda_{min}(A)^2,$$

and since the norm is nonnegative

$$\|A\| \geq -\lambda_{min}(A).$$

The last line follows by part (i). □

Newey (1994) showed that $\zeta(K)$ is $O(K)$, so this lemma implies that if $K^3/N \to 0$ (as implied by Assumption XX), $\|\hat{\Omega}_{w,K} - \Omega_{w,K}\| = o_p(1)$.

Next, define the pseudo true value $\gamma^*_{w,K}$ as

$$\gamma^*_{w,K} \equiv \left(\mathbb{E}[R_K(X)R_K(X)'|W=w]\right)^{-1}\mathbb{E}[R_K(X)Y|W=w] = \Omega^{-1}_{w,K}\mathbb{E}[R_K(X)Y|W=w].$$

and

$$\tilde{\gamma}_{w,K} \equiv \gamma^*_{w,K} + \Omega^{-1}_{w,K}R'_{w,K}\varepsilon_w/N_w$$

where

$$\varepsilon_w \equiv Y_w - \mu_w(\mathbf{X}).$$

Then we can write $\sqrt{N_w}(\tilde{\gamma}_{w,K} - \gamma^*_{w,K})$ as

$$\Omega^{-1}_{w,K}\frac{1}{\sqrt{N_w}}R'_{w,K}\varepsilon_w = \frac{1}{\sqrt{N_w}}\sum_{i|W_i=w}^{N}\Omega^{-1}_{w,K}R_K(X_i)\varepsilon_i$$

with

$$\mathbb{E}[\Omega^{-1}_{w,K}R_K(X_i)\varepsilon_i] = \Omega^{-1}_{w,K}\mathbb{E}\left[R_K(X_i)\mathbb{E}\left[\varepsilon_i|X_i\right]\right] = 0$$

and

$$\mathbb{V}\left[\Omega^{-1}_{w,K}R_K(X_i)\varepsilon_i\right] = \sigma^2_w \cdot \Omega^{-1}_{w,K}$$

Therefore,

$$S_{w,K} \equiv \frac{1}{\sqrt{N_w}}\sum_{i|W_i=w}^{N}\left[\sigma^2_w \cdot \Omega_{w,K}\right]^{-\frac{1}{2}}R_K(X_i)\varepsilon_i \equiv \frac{1}{\sqrt{N_w}}\sum_{i|W_i=w}^{N}Z_i$$

is a normalized summation of N_w independent random vectors distributed with expectation $\mathbf{0}$ and variance-covariance matrix I_K.
Denote the distribution of $S_{w,K}$ by Q_{N_w} and define $\beta_3 \equiv \sum_{i=1}^{N}\mathbb{E}\left\|\frac{Z_i}{\sqrt{N}}\right\|^3$. Then, by Theorem 1.3, Götze (1991), provided that $K \geq 6$,

$$sup_{\mathcal{A}\in A_K}|Q_{N_w}(\mathcal{A}) - \Phi(\mathcal{A})| \leq C_K\beta_3 N^{-\frac{1}{2}}$$

where A_K is the class of all measurable convex sets in K-dimensional Euclidean space, C_K is $O(K)$, and Φ is a multivariate standard Gaussian distribution.

Lemma A.2 *Suppose that $K(N) = N^{\nu}$ where $\nu < \frac{2}{11}$. Then,*

$$\sup_{\mathcal{A}\in A_K}|Q_{N_w}(\mathcal{A}) - \Phi(\mathcal{A})| \to 0$$

Proof: First we will show that β_3 is $O(K^{\frac{9}{2}}N^{-\frac{1}{2}})$

$$\begin{aligned}\beta_3 \equiv \sum_{i|W_i=w}^{N} \mathbb{E}\left\|\frac{Z_i}{\sqrt{N_w}}\right\|^3 &= N_w^{-\frac{3}{2}} \sum_{i|W_i=w}^{N} \mathbb{E}\left\|\left[\sigma_w^2 \cdot \Omega_{w,K}\right]^{-\frac{1}{2}} R_K(X_i)\varepsilon_i\right\|^3 \\ &= \left(N_w \cdot \sigma_w^2\right)^{-\frac{3}{2}} \sum_{i|W_i=w}^{N} \mathbb{E}\left\|\Omega_{w,K}^{-\frac{1}{2}} R_K(X_i)\varepsilon_i\right\|^3 \\ &\leq \left(N_w \cdot \sigma_w^2\right)^{-\frac{3}{2}} \sum_{i|W_i=w}^{N} \mathbb{E}\left[\|\Omega_{w,K}^{-\frac{1}{2}}\|^3 \, \|R_K(X_i)\varepsilon_i\|^3\right]\end{aligned}$$

First, consider

$$\|\Omega_{w,K}^{-\frac{1}{2}}\|^3 = \left[\mathrm{tr}(\Omega_{w,K}^{-1})\right]^{\frac{3}{2}} \leq \left[K \cdot \lambda_{max}(\Omega_{w,K}^{-1})\right]^{\frac{3}{2}} \leq C \cdot K^{\frac{3}{2}}$$

which is $O(K^{\frac{3}{2}})$ because $\lambda_{min}(\Omega_{w,K})$ is bounded away from zero by Lemma 0.1. Next, consider

$$\mathbb{E}\,\|R_K(X_i)\varepsilon_i\|^3 \leq \sup_x \|R_K(x)\|^3 \cdot \mathbb{E}\,|\varepsilon_i|^3 \leq C \cdot K^3$$

where the third moment of ε_i is bounded by Assumption XX and so the factor is $O(K^3)$. Since σ_w^2 is also bounded by Assumption XX, β_3 is $O(K^{\frac{9}{2}}N^{-\frac{1}{2}})$. Thus,

$$C_K\beta_3 N_w^{-\frac{1}{2}} = K \sum_{i|W_i=w}^{N} \mathbb{E}\left\|\frac{Z_i}{\sqrt{N_w}}\right\|^3 N_w^{-\frac{1}{2}} \leq C \cdot K \cdot K^{\frac{9}{2}} N_w^{-\frac{1}{2}} \cdot N_w^{-\frac{1}{2}} = C \cdot K^{\frac{11}{2}} N^{-1}$$

and the result follows. □

We may proceed further to detail conditions under which the quadratic form, $S'_{w,K}S_{w,K}$, properly normalized, converges to a univariate standard Gaussian distribution. The quadratic form $S'_{w,K}S_{w,K}$ can be written as

$$S'_{w,K}S_{w,K} = \sum_{j=1}^{K}\Big(\frac{1}{\sqrt{N_w}}\sum_{i|W_i=w}^{N} Z_{ij}\Big)^2$$

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 uncorrelated, squared random variables with each random variable converging to a standard Gaussian distribution by the previous result. Intuitively, this sum should converge to a χ^2 random variable with K degrees of freedom.

Lemma A.3 *Under Assumptions XX-XX,*

$$\sup_c \left|\mathrm{P}(S'_{w,K}S_{w,K} \leq c) - \chi^2_K(c)\right| \to 0.$$

Proof: Define the set $A(c) \equiv \left\{S \in \mathbb{R}^K \mid S'S \leq c\right\}$. Note that $A(c)$ is a measurable convex set in $\mathbb{R}^K$. Also note that for $Z_K \sim \mathcal{N}(0, I_K)$, we have that $\chi^2_K(c) = \mathrm{P}(Z'_K Z_k \leq c)$. Then,

$$\begin{aligned}\sup_c \left|\mathrm{P}[S'_{w,K}S_{w,K} \leq c] - \chi^2_K(c)\right| &= \sup_c \left|\mathrm{P}(S'_{w,K}S_{w,K} \leq c) - \mathrm{P}(Z'_K Z_K \leq c)\right| \\ &= \sup_c \left|\mathrm{P}(S_{w,K} \in A(c)) - \mathrm{P}(Z_K \in A(c))\right| \\ &\leq \sup_{\mathcal{A}\in A_K} \left|Q_{N_w}(\mathcal{A}) - \Phi(\mathcal{A})\right| \\ &\leq C_K\beta_3 N^{-\frac{1}{2}} \\ &= O(K^{\frac{11}{2}}N^{-1})\end{aligned}$$

which is o(1) for $\nu < \frac{2}{11}$ by Lemma 0.2. □

The proper normalization of the quadratic form yields the studentized version, $(S'_{w,K}S_{w,K} - K)/\sqrt{2K}$. This converges to a standard Gaussian distribution by the following lemma.

Lemma A.4 *Under Assumptions XX-XX,*

$$\sup_c \left| \mathrm{P}\left(\frac{S'_{w,K}S_{w,K} - K}{\sqrt{2K}} \leq c \right) - \Phi(c) \right| \to 0.$$

Proof:

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

The first term goes to zero by Lemma 0.3. For the second term we may apply the Berry-Esséen Theorem which yields,

$$\sup_c \left| \mathrm{P}\left(\frac{Z'_K Z_K - K}{\sqrt{2K}} \leq c \right) - \Phi(c) \right| \leq C \cdot K^{-\frac{1}{2}}.$$

Thus for $\nu > 0$ the right-hand side converges to zero as well and the result is established. □

In order to proceed we need the following selected results from Imbens, Newey and Ridder (2004). These results establish convergence rates for the estimators of the regression function.

Lemma A.5 (Imbens, Newey and Ridder (2004)): *Suppose Assumptions XX-XX hold. Then,*

(i) there is a sequence $\gamma^0_{w,K}$ such that

$$\sup_x \left| \mu_w(x) - R_K(x)'\gamma^0_{w,K} \right| \equiv \sup_x \left| \mu_w(x) - \mu^0_{w,K} \right| = O(K^{-\frac{s}{d}})$$

(ii)

$$\sup_x \left| R_K(x)'\gamma^*_{w,K} - R_K(x)'\gamma^0_{w,K} \right| \equiv \sup_x \left| \mu^*_{w,K} - \mu^0_{w,K} \right| = O_p\left(\zeta(K)^2 K^{-\frac{s}{d}} \right)$$

(iii)

$$\left\| \gamma^*_{w,K} - \gamma^0_{w,k} \right\| = O(\zeta(K)K^{-\frac{s}{d}})$$

(iv)

$$\left\| \hat{\gamma}_{w,K} - \gamma^0_{w,k} \right\| = O_p(K^{\frac{1}{2}}N^{-\frac{1}{2}} + K^{-\frac{s}{d}})$$

The following lemma describes the limiting distribution of the infeasible test statistic.

Lemma A.6 *Under Assumptions XX-XX,*

$$\left(N_w \cdot \left(\hat{\gamma}_{w,K} - \gamma^*_{w,K} \right)' \left(\hat{\sigma}^2_{w,K} \cdot \hat{\Omega}^{-1}_{w,K} \right)^{-1} \left(\hat{\gamma}_{w,K} - \gamma^*_{w,K} \right) - K \right) / \sqrt{2K} \xrightarrow{d} \mathcal{N}(0,1)$$

Proof: We need only show that,

$$\left\| \left[\hat{\sigma}^2_{w,K} \cdot \hat{\Omega}^{-1}_{w,K}\right]^{-\frac{1}{2}} \sqrt{N_w} \left(\hat{\gamma}_{w,K} - \gamma^*_{w,K}\right) - S_{w,K} \right\| = o_p(1).$$

then the result follows by Lemmas 0.2, 0.3, and 0.4.

First, notice that we can rewrite $\hat{\gamma}_{w,K}$ as

$$\hat{\gamma}_{w,K} = \gamma^*_{w,K} + \hat{\Omega}^{-1}_{w,K} R'_{w,K} \varepsilon_{w,K} / N_w$$

where

$$\varepsilon_{w,K} \equiv Y_w - R_{w,K} \gamma^*_{w,K},$$

with ith row equal to

$$\varepsilon_{Ki} = Y_i - R_K(X_i)' \gamma^*_{w,K}.$$

Then,

$$\begin{aligned}
& \left\| \left[\hat{\sigma}^2_{w,K} \cdot \hat{\Omega}^{-1}_{w,K}\right]^{-\frac{1}{2}} \sqrt{N_w} \left(\hat{\gamma}_{w,K} - \gamma^*_{w,K}\right) - S_{w,K} \right\| \\
= & \left\| \left[\hat{\sigma}^2_{w,K} \cdot \hat{\Omega}^{-1}_{w,K}\right]^{-\frac{1}{2}} \sqrt{N_w} \cdot \hat{\Omega}^{-1}_{w,K} \cdot R'_{w,K} \varepsilon_{w,K} / N_w - \left[\sigma^2_w \cdot \Omega_{w,K}\right]^{-\frac{1}{2}} \sqrt{N_w} \cdot R'_{w,K} \varepsilon_w / N_w \right\| \\
= & \left\| \hat{\sigma}^{-1}_{w,K} \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_{w,K} / \sqrt{N_w} - \sigma^{-1}_w \Omega^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} \right\| \\
= & \Big\| \hat{\sigma}^{-1}_{w,K} \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_{w,K} / \sqrt{N_w} - \hat{\sigma}^{-1}_{w,K} \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} \\
& \quad + \hat{\sigma}^{-1}_{w,K} \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} - \sigma^{-1}_w \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} \\
& \quad + \sigma^{-1}_w \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} - \sigma^{-1}_w \Omega^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} \Big\| \\
\leq & \left\| \hat{\sigma}^{-1}_{w,K} \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_{w,K} / \sqrt{N_w} - \hat{\sigma}^{-1}_{w,K} \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} \right\| \\
& + \left\| \hat{\sigma}^{-1}_{w,K} \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} - \sigma^{-1}_w \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} \right\| \\
& + \left\| \sigma^{-1}_w \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} - \sigma^{-1}_w \Omega^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} \right\| \\
= & \left| \hat{\sigma}^{-1}_{w,K} \right| \left\| \hat{\Omega}^{-\frac{1}{2}}_{w,K} R'_{w,K} \left(\varepsilon_{w,K} - \varepsilon_w\right) / \sqrt{N_w} \right\| & \text{(A.4)} \\
& + \left| \hat{\sigma}^{-1}_{w,K} - \sigma^{-1}_w \right| \left\| \hat{\Omega}^{-\frac{1}{2}}_{w,K} \cdot R'_{w,K} \varepsilon_w / \sqrt{N_w} \right\| & \text{(A.5)} \\
& + \left| \sigma^{-1}_w \right| \left\| \left(\hat{\Omega}^{-\frac{1}{2}}_{w,K} - \Omega^{-\frac{1}{2}}_{w,K}\right) R'_{w,K} \varepsilon_w / \sqrt{N_w} \right\| & \text{(A.6)}
\end{aligned}$$

First, consider equation (4),

$$\begin{aligned}
& \left| \hat{\sigma}^{-1}_{w,K} \right| \left\| \hat{\Omega}^{-\frac{1}{2}}_{w,K} R'_{w,K} \left(\varepsilon_{w,K} - \varepsilon_w\right) / \sqrt{N_w} \right\| \\
= & \left(\sigma^{-1}_w + o_p\left(N^{-\frac{1}{2}}\right)\right) \cdot \left\| \hat{\Omega}^{-\frac{1}{2}}_{w,K} R'_{w,K} \left(\varepsilon_{w,K} - \varepsilon_w\right) / \sqrt{N_w} \right\| \\
= & \left(O\left(1\right) + o_p\left(N^{-\frac{1}{2}}\right)\right) \cdot \left\| \hat{\Omega}^{-\frac{1}{2}}_{w,K} R'_{w,K} \left(\varepsilon_{w,K} - \varepsilon_w\right) / \sqrt{N_w} \right\|
\end{aligned}$$

where

$$\begin{aligned}
& \mathbb{E}\left\|\hat{\Omega}_{w,K}^{-\frac{1}{2}} R_{w,K}' \left(\varepsilon_{w,K}-\varepsilon_w\right)/\sqrt{N_w}\right\|^2 \\
= \ & \mathbb{E}\left[\frac{1}{N_w}\operatorname{tr}\left(\left(\varepsilon_{w,K}-\varepsilon_w\right)' R_{w,K}\hat{\Omega}_{w,K}^{-1} R_{w,K}'\left(\varepsilon_{w,K}-\varepsilon_w\right)\right)\right] \\
= \ & \mathbb{E}\left[\left(\left(\varepsilon_{w,K}-\varepsilon_w\right)' R_{w,K}\left(R_{w,K}' R_{w,K}\right)^{-1} R_{w,K}'\left(\varepsilon_{w,K}-\varepsilon_w\right)\right)\right] \\
\leq \ & \mathbb{E}\left[\left(\varepsilon_{w,K}-\varepsilon_w\right)'\left(\varepsilon_{w,K}-\varepsilon_w\right)\right] \\
= \ & \mathbb{E}\left[\left(\mu_w(\mathbf{X})-R_{w,K}\gamma_{w,K}^*\right)'\left(\mu_w(\mathbf{X})-R_{w,K}\gamma_{w,K}^*\right)\right] \\
\leq \ & N_w\cdot\sup_x\left|\mu_w(x)-R_K(x)'\gamma_{w,K}^*\right|^2 \\
\leq \ & N_w\cdot\sup_x\left(\left|\mu_w(x)-R_K(x)'\gamma_{w,K}^0\right|+\left|R_K(x)'\gamma_{w,K}^0-R_K(x)'\gamma_{w,K}^*\right|\right)^2 \\
= \ & N_w\left(O\left(K^{-\frac{s}{d}}\right)+O\left(\zeta(K)^2K^{-\frac{s}{d}}\right)\right)^2 \\
= \ & O\left(N\right)\cdot\left(O\left(\zeta(K)^2K^{-\frac{s}{d}}\right)\right)^2
\end{aligned}$$

so that equation (4) is $O_p\left(N^{\frac{1}{2}}\zeta(K)^2K^{-\frac{s}{d}}\right)$ by Markov's inequality and consistency of the sample variance. The third line follows since $(I_{N_w}-R_{w,K}(R_{w,K}'R_{w,K})^{-1}R_{w,K}')$ is a projection matrix and so it is positive semi-definite. The seventh line follows from Lemma 0.5 (i) and (ii).

Now consider equation (5),

$$\left|\hat{\sigma}_{w,K}^{-1}-\sigma_w^{-1}\right|\left\|\hat{\Omega}_{w,K}^{-\frac{1}{2}}\cdot R_{w,K}'\varepsilon_w/\sqrt{N_w}\right\|$$

The first factor is $o_p(N^{-\frac{1}{2}})$ and

$$\begin{aligned}
& \mathbb{E}\left\|\hat{\Omega}_{w,K}^{-\frac{1}{2}}\cdot R_{w,K}'\varepsilon_w/\sqrt{N_w}\right\|^2 \\
= \ & \mathbb{E}\left[\frac{1}{N_w}\operatorname{tr}\left(\varepsilon_w' R_{w,K}\hat{\Omega}_{w,K}^{-1}R_{w,K}'\varepsilon_w\right)\right] \\
= \ & \mathbb{E}\left[\operatorname{tr}\left(\varepsilon_w' R_{w,K}\left(R_{w,K}'R_{w,K}\right)^{-1}R_{w,K}'\varepsilon_w\right)\right] \\
= \ & \mathbb{E}\left[\operatorname{tr}\left(R_{w,K}\left(R_{w,K}'R_{w,K}\right)^{-1}R_{w,K}'\varepsilon_w\varepsilon_w'\right)\right] \\
= \ & \operatorname{tr}\left(\mathbb{E}\left[R_{w,K}\left(R_{w,K}'R_{w,K}\right)^{-1}R_{w,K}'\mathbb{E}\left[\varepsilon_w\varepsilon_w'|\mathbf{X}\right]\right]\right) \\
= \ & \sigma_w^2\cdot\operatorname{tr}\left(\mathbb{E}\left[R_{w,K}\left(R_{w,K}'R_{w,K}\right)^{-1}R_{w,K}'\right]\right) \\
= \ & \sigma_w^2\cdot\mathbb{E}\left[\operatorname{tr}\left(R_{w,K}\left(R_{w,K}'R_{w,K}\right)^{-1}R_{w,K}'\right)\right] \\
= \ & \sigma_w^2\cdot\mathbb{E}\left[\operatorname{tr}\left(\left(R_{w,K}'R_{w,K}\right)^{-1}R_{w,K}'R_{w,K}\right)\right] \\
= \ & \sigma_w^2\cdot\operatorname{tr}\left(I_K\right) \\
= \ & \sigma_w^2\cdot K
\end{aligned}$$

so that the second factor is $O\left(K^{\frac{1}{2}}\right)$. Thus equation (5) is $o_p\left(K^{\frac{1}{2}}N^{-\frac{1}{2}}\right)$.

Finally, consider equation (6),

$$\begin{aligned}
& \left|\sigma_w^{-1}\right|\left\|\left(\hat{\Omega}_{w,K}^{-\frac{1}{2}}-\Omega_{w,K}^{-\frac{1}{2}}\right)R_{w,K}'\varepsilon_w/\sqrt{N_w}\right\| \\
\leq \ & C\cdot\left\|\hat{\Omega}_{w,K}^{-\frac{1}{2}}-\Omega_{w,K}^{-\frac{1}{2}}\right\|\left\|R_{w,K}'\varepsilon_w/\sqrt{N_w}\right\|
\end{aligned}$$

The first factor is $O_p\left(\zeta(K)K^{\frac{1}{2}}N^{-\frac{1}{2}}\right)$ by Lemma 0.1 and the continuous mapping theorem, and

$$\begin{aligned}
& \mathbb{E}\left\|R'_{w,K}\varepsilon_w/\sqrt{N_w}\right\|^2 \\
= & \ \mathbb{E}\left[\frac{1}{N_w}\mathrm{tr}\left(\varepsilon'_w R_{w,K}R'_{w,K}\varepsilon_w\right)\right] \\
= & \ \mathbb{E}\left[\frac{1}{N_w}\mathrm{tr}\left(R'_{w,K}\varepsilon_w\varepsilon'_w R_{w,K}\right)\right] \\
= & \ \mathrm{tr}\left(\frac{1}{N_w}\mathbb{E}\left[R'_{w,K}\mathbb{E}\left[\varepsilon_w\varepsilon'_w|\mathbf{X}\right]R_{w,K}\right]\right) \\
= & \ \sigma^2_w\cdot\mathrm{tr}\left(\mathbb{E}\left[R'_{w,K}R_{w,K}/N_w\right]\right) \\
= & \ \sigma^2_w\cdot\mathrm{tr}\left(\Omega_{w,K}\right) \\
\leq & \ \sigma^2_w\cdot K\cdot\lambda_{max}\left(\Omega_{w,K}\right) \\
\leq & \ C\cdot K
\end{aligned}$$

so that the second factor is $O\left(K^{\frac{1}{2}}\right)$ by Assumption XX, Lemma 0.1 (ii) and Markov's inequality. Thus, equation (3) is $O_p\left(\zeta(K)KN^{-\frac{1}{2}}\right)$.

Combining these results yields:

$$\begin{aligned}
& \left\|\left[\hat{\sigma}^2_{w,K}\cdot\hat{\Omega}^{-1}_{w,K}\right]^{-\frac{1}{2}}\sqrt{N_w}\left(\hat{\gamma}_{w,K}-\gamma^*_{w,K}\right)-S_{w,K}\right\| \\
= & \ O_p\left(N^{\frac{1}{2}}\zeta(K)^2K^{-\frac{s}{d}}\right)+o_p\left(K^{\frac{1}{2}}N^{-\frac{1}{2}}\right)+O_p\left(\zeta(K)KN^{-\frac{1}{2}}\right) \\
= & \ O_p\left(N^{\frac{1}{2}}K^{(2-\frac{s}{d})}\right)+o_p\left(K^{\frac{1}{2}}N^{-\frac{1}{2}}\right)+O_p\left(K^2N^{-\frac{1}{2}}\right)
\end{aligned}$$

All three terms are $o_p(1)$ by Assumption XX and for $\frac{s}{d}>\frac{4\nu+1}{2\nu}$. □

Proof of Theorem 8.1: From the previous lemma we have that [2]

$$T^*\equiv\left(N_w\cdot\left((\hat{\gamma}_{1,K}-\hat{\gamma}_{0,K})-(\gamma^*_{1,K}-\gamma^*_{0,K})\right)'\cdot\hat{V}^{-1}\cdot\left((\hat{\gamma}_{1,K}-\hat{\gamma}_{0,K})-(\gamma^*_{1,K}-\gamma^*_{0,K})\right)-K\right)/\sqrt{2K}$$

converges in distribution to a $\mathcal{N}(0,1)$ random variable, where $\hat{V}$ is defined as

$$\hat{V}\equiv(\hat{\sigma}^2_{0,K}\cdot\hat{\Omega}^{-1}_{0,K}+\hat{\sigma}^2_{1,K}\cdot\hat{\Omega}^{-1}_{1,K}).$$

To complete the proof we must show that under our assumptions $|T^*-T|=o_p(1)$, where T is defined as

$$T\equiv\left(N_w\cdot(\hat{\gamma}_{1,K}-\hat{\gamma}_{0,K})'\cdot\hat{V}^{-1}\cdot(\hat{\gamma}_{1,K}-\hat{\gamma}_{0,K})-K\right)/\sqrt{2K}.$$

Note that under the null hypothesis $\mu_1(x)=\mu_0(x)$ so we may choose the same approximating sequence $\gamma^0_{1,K}=\gamma^0_{0,K}$ for $\mu^0_{1,K}(x)=\mu^0_{0,K}(x)$. Then,

$$\begin{aligned}
\left\|\gamma^*_{1,K}-\gamma^*_{0,K}\right\| &= \left\|\gamma^*_{1,K}-\gamma^0_{1,K}+\gamma^0_{0,K}-\gamma^*_{0,K}\right\| \\
&\leq \left\|\gamma^*_{1,K}-\gamma^0_{1,K}\right\|+\left\|\gamma^0_{0,K}-\gamma^*_{0,K}\right\| \\
&= O(\zeta(K)K^{-\frac{s}{d}}) \qquad \text{(A.7)}
\end{aligned}$$

[2]For simplicity of notation we assume $N_1=N_0$

by Lemma 0.5 (iii), and

$$\begin{aligned}
\|\hat\gamma_{1,K}-\hat\gamma_{0,K}\| &= \left\|\hat\gamma_{1,K}-\gamma^0_{1,K}+\gamma^0_{0,K}-\hat\gamma_{0,K}\right\| \\
&\leq \left\|\hat\gamma_{1,K}-\gamma^0_{1,K}\right\|+\left\|\gamma^0_{0,K}-\hat\gamma_{0,K}\right\| \\
&= O_p(K^{\frac{1}{2}}N^{-\frac{1}{2}}+K^{-\frac{s}{d}}) \qquad \text{(A.8)}
\end{aligned}$$

by Lemma 0.5 (iv). So then,

$$\begin{aligned}
|T^*-T| &= \Big|\Big(N_w\cdot\left((\hat\gamma_{1,K}-\hat\gamma_{0,K})-(\gamma^*_{1,K}-\gamma^*_{0,K})\right)'\hat V^{-1}\left((\hat\gamma_{1,K}-\hat\gamma_{0,K})-(\gamma^*_{1,K}-\gamma^*_{0,K})\right)-K\Big)/\sqrt{2K} \\
&\quad -\Big(N_w\cdot(\hat\gamma_{1,K}-\hat\gamma_{0,K})'\hat V^{-1}(\hat\gamma_{1,K}-\hat\gamma_{0,K})-K\Big)/\sqrt{2K}\Big| \\
&= \frac{N_w}{\sqrt{2K}}\cdot\Big|\left((\hat\gamma_{1,K}-\hat\gamma_{0,K})-(\gamma^*_{1,K}-\gamma^*_{0,K})\right)'\hat V^{-1}\left((\hat\gamma_{1,K}-\hat\gamma_{0,K})-(\gamma^*_{1,K}-\gamma^*_{0,K})\right) \\
&\quad -(\hat\gamma_{1,K}-\hat\gamma_{0,K})'\hat V^{-1}(\hat\gamma_{1,K}-\hat\gamma_{0,K})\Big| \\
&= \frac{N_w}{\sqrt{2K}}\cdot\left|-2\cdot(\hat\gamma_{1,K}-\hat\gamma_{0,K})'\hat V^{-1}\left(\gamma^*_{1,K}-\gamma^*_{0,K}\right)+\left(\gamma^*_{1,K}-\gamma^*_{0,K}\right)'\hat V^{-1}\left(\gamma^*_{1,K}-\gamma^*_{0,K}\right)\right| \\
&\leq \frac{N_w}{\sqrt{2K}}\cdot\left(2\cdot\left|(\hat\gamma_{1,K}-\hat\gamma_{0,K})'\hat V^{-1}\left(\gamma^*_{1,K}-\gamma^*_{0,K}\right)\right|+\left|\left(\gamma^*_{1,K}-\gamma^*_{0,K}\right)'\hat V^{-1}\left(\gamma^*_{1,K}-\gamma^*_{0,K}\right)\right|\right)
\end{aligned}$$

Consider the first term,

$$\begin{aligned}
2\cdot\left|(\hat\gamma_{1,K}-\hat\gamma_{0,K})'\hat V^{-1}(\gamma^*_{1,K}-\gamma^*_{0,K})\right| &= 2\cdot\left|\mathrm{tr}\left((\hat\gamma_{1,K}-\hat\gamma_{0,K})'\hat V^{-1}(\gamma^*_{1,K}-\gamma^*_{0,K})\right)\right| \\
&\leq 2\cdot\|\hat\gamma_{1,K}-\hat\gamma_{0,K}\|\cdot\left\|\gamma^*_{1,K}-\gamma^*_{0,K}\right\|\cdot\lambda_{max}(\hat V^{-1}) \\
&\leq C\cdot\|\hat\gamma_{1,K}-\hat\gamma_{0,K}\|\cdot\left\|\gamma^*_{1,K}-\gamma^*_{0,K}\right\|+o_p(1) \\
&= \left(O_p(K^{\frac{1}{2}}N^{-\frac{1}{2}}+K^{-\frac{s}{d}})\cdot O(\zeta(K)K^{-\frac{s}{d}})\right)
\end{aligned}$$

Where the third line follows from Lemma 0.1 (iii) and Assumption XX. The last line follows from equations (7) and (8).

Now, consider the second term,

$$\begin{aligned}
\left|(\gamma^*_{1,K}-\gamma^*_{0,K})'\hat V^{-1}(\gamma^*_{1,K}-\gamma^*_{0,K})\right| &= \left|\mathrm{tr}\left((\gamma^*_{1,K}-\gamma^*_{0,K})'\hat V^{-1}(\gamma^*_{1,K}-\gamma^*_{0,K})\right)\right| \\
&\leq \left\|\gamma^*_{1,K}-\gamma^*_{0,K}\right\|^2\cdot\lambda_{max}(\hat V^{-1}) \\
&\leq C\cdot\left\|\gamma^*_{1,K}-\gamma^*_{0,K}\right\|^2+o_p(1) \\
&= O(\zeta(K)^2K^{-\frac{2s}{d}})
\end{aligned}$$

Where the third line follows from Lemma 0.1 (iii) and Assumption XX. The last line follows from equation (7).

So then,

$$\begin{aligned}
|T^*-T| &= \frac{N_w}{\sqrt{2K}}\cdot\left(O_p(K^{\frac{1}{2}}N^{-\frac{1}{2}}+K^{-\frac{s}{d}})\cdot O(\zeta(K)K^{-\frac{s}{d}})+O(\zeta(K)^2K^{-\frac{2s}{d}})\right) \\
&= O_p\left(N^{\frac{1}{2}}\zeta(K)K^{-\frac{s}{d}}\right)+O_p\left(N\zeta(K)K^{-(\frac{1}{2}+\frac{2s}{d})}\right)+O\left(N\zeta(K)^2K^{-(\frac{1}{2}+\frac{2s}{d})}\right)
\end{aligned}$$

For $\frac{s}{d}>\frac{2\nu+1}{2\nu}$ all three terms are $o_p(1)$ and the result follows. □

Before proving Theorem 8.2 we need the following lemma.

Lemma A.7

$$\lambda_{max}\left(\hat{V}^{-1}\right) \geq \lambda_{max}\left(\left([\hat{V}]_{-1}\right)^{-1}\right)$$

Where $[\quad]_{-1}$ is defined as the $(K-1)$ x $(K-1)$ matrix created by eliminating the first row and column of a $(K \text{ x } K)$ square matrix.

Proof: The proof follows by the Interlacing Theorem given below.

If A is an n x n positive semi-definite Hermitian matrix with eigenvalues $\lambda_1 \geq ... \geq \lambda_n$, B is a k x k principal submatrix of A with eigenvalues $\tilde{\lambda}_1 \geq ... \geq \tilde{\lambda}_k$, then

$$\lambda_i \geq \tilde{\lambda}_i \geq \lambda_{i+n-k}, \quad i = 1, ..., k.$$

In our case, $\hat{V}$ is positive definite, symmetric and thus positive definite, Hermitian. So then, by the Interlacing Theorem

$$\lambda_{min}\left(\hat{V}\right) \leq \lambda_{min}\left([\hat{V}]_{-1}\right) \Longrightarrow \lambda_{max}\left(\hat{V}^{-1}\right) \geq \lambda_{max}\left(\left([\hat{V}]_{-1}\right)^{-1}\right)$$

□

Proof of Theorem 8.2: When the conditional average treatment effect is constant we may choose the two approximating sequences, $\gamma^0_{0,K}$ and $\gamma^0_{1,K}$, to differ only by way of the first element (the coefficient of the constant term in the approximating sequence). In other words, if $\mu_1(x) - \mu_0(x) = \tau_0$ for all $x \in \mathbb{X}$, then the coefficients of the power series terms involving x^r such that $r > 0$ should be identical for $w = 0, 1$, so that their difference no longer varies with x.

Thus, a natural strategy to test the null hypothesis of a constant conditional average treatment effect is to compare the last $K-1$ elements of $\hat{\gamma}_{1,K}$ and $\hat{\gamma}_{0,K}$ and to reject the null hypothesis when these elements are sufficiently different.

By the proof of Lemma 0.6 we have that,

$$\left\| \left[\hat{\sigma}^2_{w,K} \cdot \hat{\Omega}^{-1}_{w,K}\right]^{-\frac{1}{2}} \sqrt{N_w}\left(\hat{\gamma}_{w,K} - \gamma^*_{w,K}\right) - S_{w,K} \right\| = o_p(1)$$

so that by Lemma 0.2,

$$\left[\hat{\sigma}^2_{w,K} \cdot \hat{\Omega}^{-1}_{w,K}\right]^{-\frac{1}{2}} \sqrt{N_w}\left(\hat{\gamma}_{w,K} - \gamma^*_{w,K}\right) \xrightarrow{d} \mathcal{N}\left(0, I_K\right).$$

By the joint convergence in distribution of the elements of $\hat{\gamma}_{w,K}$ we may partition this result as

$$\left[\hat{\sigma}^2_{w,K} \cdot \left[\hat{\Omega}^{-1}_{w,K}\right]_{1,1}\right]^{-\frac{1}{2}} \sqrt{N_w}\left(\hat{\gamma}_{w,K,1} - \gamma^*_{w,K,1}\right) \xrightarrow{d} \mathcal{N}\left(0, 1\right)$$

and

$$\left[\hat{\sigma}^2_{w,K} \cdot \left[\hat{\Omega}^{-1}_{w,K}\right]_{-1}\right]^{-\frac{1}{2}} \sqrt{N_w}\left(\left[\hat{\gamma}_{w,K}\right]_{-1} - \left[\gamma^*_{w,K}\right]_{-1}\right) \xrightarrow{d} \mathcal{N}\left(0, I_{K-1}\right) \tag{A.9}$$

where $\hat{\gamma}_{w,K,i}$ is the ith element of $\hat{\gamma}_{w,K}$, $[M]_{i,j}$ is the (i,j) element of a matrix M, and where $[\quad]_{-1}$ also represents the $(K-1)$ vector created by eliminating the first element of a vector of dimension K as well as the $(K-1)$ x $(K-1)$ matrix created by eliminating the first row and column of a $(K$ x $K)$ square matrix.

By the independence of the control and treated observations, marginal convergence in distribution of (1) for $w = \{0,1\}$ implies joint convergence in distribution. Then, applying the Cramer-Wold device

$$\left([\hat{V}]_{-1}\right)^{-\frac{1}{2}} \sqrt{N_w} \left([\hat{\gamma}_{1,K}]_{-1} - [\hat{\gamma}_{0,K}]_{-1} - \left(\left[\gamma^*_{1,K}\right]_{-1} - \left[\gamma^*_{0,K}\right]_{-1}\right)\right) \xrightarrow{d} \mathcal{N}\left(0, I_{K-1}\right)$$

where

$$[\hat{V}]_{-1} = \hat{\sigma}^2_{0,K} \cdot \left[\hat{\Omega}^{-1}_{0,K}\right]_{-1} + \hat{\sigma}^2_{1,K} \cdot \left[\hat{\Omega}^{-1}_{1,K}\right]_{-1}$$

so that by the continuous mapping theorem (or following the logic of Lemma 0.3) we have that

$$N_w \left(\hat{\delta} - \delta^*\right)' \left([\hat{V}]_{-1}\right)^{-1} \left(\hat{\delta} - \delta^*\right) \xrightarrow{d} \chi^2_{K-1}$$

where

$$\hat{\delta} = [\hat{\gamma}_{1,K}]_{-1} - [\hat{\gamma}_{0,K}]_{-1}$$

and

$$\delta^* = \left[\gamma^*_{1,K}\right]_{-1} - \left[\gamma^*_{0,K}\right]_{-1}.$$

We need only show that

$$N_w \left\{\left(\hat{\delta} - \delta^*\right)' \left([\hat{V}]_{-1}\right)^{-1} \left(\hat{\delta} - \delta^*\right) - \hat{\delta}' \left([\hat{V}]_{-1}\right)^{-1} \hat{\delta}\right\} = o_p(1) \tag{A.10}$$

to obtain the asymptotic distribution of our feasible test statistic. To continue further we will again use the results summarized in Lemma 0.5. Specifically, note that

$$\begin{aligned}
\left\|\left[\gamma^*_{w,K}\right]_{-1} - \left[\gamma^0_{w,K}\right]_{-1}\right\|^2 &= \sum_{i=2}^{K} \left(\gamma^*_{w,K,i} - \gamma^0_{w,K,i}\right)^2 \\
&\leq \sum_{i=1}^{K} \left(\gamma^*_{w,K,i} - \gamma^0_{w,K,i}\right)^2 \\
&= \left\|\gamma^*_{w,K} - \gamma^0_{w,K}\right\|^2 \\
&= O\left(\left[\zeta(K) K^{-\frac{s}{d}}\right]^2\right) \qquad \text{(A.11)}
\end{aligned}$$

by Lemma 0.5 (iii) and

$$\begin{aligned}
\left\|[\hat{\gamma}_{w,K}]_{-1} - \left[\gamma^0_{w,K}\right]_{-1}\right\|^2 &= \sum_{i=2}^{K} \left(\hat{\gamma}_{w,K,i} - \gamma^0_{w,K,i}\right)^2 \\
&\leq \sum_{i=1}^{K} \left(\hat{\gamma}_{w,K,i} - \gamma^0_{w,K,i}\right)^2 \\
&= \left\|\hat{\gamma}_{w,K} - \gamma^0_{w,K}\right\|^2 \\
&= O_p\left(\left[K^{\frac{1}{2}} N^{-\frac{1}{2}} + K^{-\frac{s}{d}}\right]^2\right). \qquad \text{(A.12)}
\end{aligned}$$

by Lemma 0.5 (iv). We may choose the last $(K-1)$ elements of the approximating sequence to be equal, $[\gamma_{1,K}^0]_{-1} = [\gamma_{1,K}^0]_{-1}$. This allows us to bound $\hat{\delta}$ and δ^* by the following

$$\begin{aligned} \left\|\hat{\delta}\right\| &= \left\|[\hat{\gamma}_{1,K}]_{-1} - [\hat{\gamma}_{0,K}]_{-1}\right\| \\ &= \left\|[\hat{\gamma}_{1,K}]_{-1} - \left[\gamma_{1,K}^0\right]_{-1} + \left[\gamma_{0,K}^0\right]_{-1} - [\hat{\gamma}_{0,K}]_{-1}\right\| \\ &\leq \left\|[\hat{\gamma}_{1,K}]_{-1} - \left[\gamma_{1,K}^0\right]_{-1}\right\| + \left\|\left[\gamma_{0,K}^0\right]_{-1} - [\hat{\gamma}_{0,K}]_{-1}\right\| \\ &= O_p\left(K^{\frac{1}{2}} N^{-\frac{1}{2}} + K^{-\frac{s}{d}}\right) \end{aligned} \tag{A.13}$$

by equation (12). Also,

$$\begin{aligned} \|\delta^*\| &= \left\|\left[\gamma_{1,K}^*\right]_{-1} - \left[\gamma_{0,K}^*\right]_{-1}\right\| \\ &= \left\|\left[\gamma_{1,K}^*\right]_{-1} - \left[\gamma_{1,K}^0\right]_{-1} + \left[\gamma_{0,K}^0\right]_{-1} - \left[\gamma_{0,K}^*\right]_{-1}\right\| \\ &\leq \left\|\left[\gamma_{1,K}^*\right]_{-1} - \left[\gamma_{1,K}^0\right]_{-1}\right\| + \left\|\left[\gamma_{0,K}^0\right]_{-1} - \left[\gamma_{0,K}^*\right]_{-1}\right\| \\ &= O\left(\zeta(K) K^{-\frac{s}{d}}\right) \end{aligned} \tag{A.14}$$

by equation (11).

Next we can rewrite the left-hand side of equation (10) as

$$N_w \left\{ \left(\hat{\delta} - \delta^*\right)' \left([\hat{V}]_{-1}\right)^{-1} \left(\hat{\delta} - \delta^*\right) - \hat{\delta}' \left([\hat{V}]_{-1}\right)^{-1} \hat{\delta} \right\} = N_w \left\{ \delta^{*\prime} \left([\hat{V}]_{-1}\right)^{-1} \delta^* - 2 \cdot \hat{\delta}' \left([\hat{V}]_{-1}\right)^{-1} \delta^* \right\}$$

Consider the first term,

$$\begin{aligned} \left|\delta^{*\prime} \left([\hat{V}]_{-1}\right)^{-1} \delta^*\right| &= \left|\operatorname{tr}\left(\delta^{*\prime} \left([\hat{V}]_{-1}\right)^{-1} \delta^*\right)\right| \\ &\leq \|\delta^*\|^2 \cdot \lambda_{max}\left(\left([\hat{V}]_{-1}\right)^{-1}\right) \\ &\leq \|\delta^*\|^2 \cdot \lambda_{max}(\hat{V}^{-1}) \\ &\leq C \cdot \|\delta^*\|^2 + o_p(1) \\ &= O\left(\zeta(K)^2 K^{-\frac{2s}{d}}\right) \end{aligned}$$

The third line follows from Lemma 0.7. The fourth line follows from Lemma 0.1 (iii) and Assumption XX. The last line follows from equation (14). Now consider the second term,

$$\begin{aligned} 2 \cdot \left|\hat{\delta}' \left([\hat{V}]_{-1}\right)^{-1} \delta^*\right| &= 2 \cdot \left|\operatorname{tr}\left(\hat{\delta}' \left([\hat{V}]_{-1}\right)^{-1} \delta^*\right)\right| \\ &\leq 2 \cdot \left\|\hat{\delta}\right\| \cdot \|\delta^*\| \cdot \lambda_{max}\left(\left([\hat{V}]_{-1}\right)^{-1}\right) \\ &\leq 2 \cdot \left\|\hat{\delta}\right\| \cdot \|\delta^*\| \cdot \lambda_{max}(\hat{V}^{-1}) \\ &\leq C \cdot \left\|\hat{\delta}\right\| \cdot \|\delta^*\| + o_p(1) \\ &= O_p\left(K^{\frac{1}{2}} N^{-\frac{1}{2}} + K^{-\frac{s}{d}}\right) \cdot O\left(\zeta(K) K^{-\frac{s}{d}}\right) \end{aligned}$$

The third line follows from Lemma 0.7. The fourth line follows from Lemma 0.1 (iii) and Assumption XX. The last line follows from equations (13) and (14). Thus the left-hand side of equation (10) is of order

$$O(N) \cdot \left[O\left(\zeta(K)^2 K^{-\frac{2s}{d}}\right) + O_p\left(K^{\frac{1}{2}} N^{-\frac{1}{2}} + K^{-\frac{s}{d}}\right) \cdot O\left(\zeta(K) K^{-\frac{s}{d}}\right).\right]$$

For $\frac{s}{d} > \frac{3\nu+1}{2\nu}$ all three terms are $o_p(1)$ and

$$N_w cdot \hat{\delta}' \left([\hat{V}]_{-1}\right)^{-1} \hat{\delta} \xrightarrow{d} \chi^2_{K-1}.$$

Finally, by Lemma 0.4 replacing, K with $(K-1)$, we have that

$$\frac{1}{\sqrt{2(K-1)}} \left[N_w \cdot \left(\left([\hat{\gamma}_{1,K}]_{-1} - [\hat{\gamma}_{0,K}]_{-1} \right)' \left([\hat{V}]_{-1}\right)^{-1} \left([\hat{\gamma}_{1,K}]_{-1} - [\hat{\gamma}_{0,K}]_{-1} \right) \right) - (K-1) \right] \xrightarrow{d} \mathcal{N}(0,1)$$

□

References

Abadie, A., (2002), "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, Vol 97, 284-292.

Abadie, A., and G. Imbens, (2005a), "Large Sample Properties of Matching Estimators for Average Treatment Effects,"forthcoming, *Econometrica*.

Bentkus, V., (2005), "A Lyapunov-type Bound in R^d," *Theory of Probability and Applications*, Vol 49(2), 311-322.

Bitler, M., J. Gelbach, and H. Hoynes (2002) "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," unpublished paper, Department of Economics, University of Maryland.

Bolthausen, E., and F. Götze (1993), "The Rate of Convergence for Multivariate Sampling Statistics," *The Annals of Statistics*, V. 21: 1692-1710.

Chen, X., Hong, H., and Tarozzi, A., (2004), "Semiparametric Efficiency in GMM Models of Nonclassical Measurement Error, Missing Data and Treatment Effects."Working Paper.

Crump, R., V. J. Hotz, V. J., G. Imbens, and O.Mitnik, (2005), "Moving the Goalposts: Addressing Limited Overlap in Estimation of Average Treatment Effects by Changing the Estimand,"unpublished manuscript, Department of Economics, UC Berkeley.

Doksum, K., (1974), "Empirical Probability Plots and Statistical Inference for Nonlinear Models in the Two-Sample Case," *The Annals of Statistics*, Vol 2, 267-277.

Firpo, S., (2004), "Efficient Semiparametric Estimation of Quantile Treatment Effects"Working Paper.

Götze, F., (1991), "On the Rate of Convergence in te Multivariate CLT," *The Annals of Probability*, Vol 19(2), 724-739.

Hahn, J., (1998), "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects,"*Econometrica* 66 (2), 315-331.

Härdle, W., and E. Mammen, (1993), "Comparing Nonparametric Versus Parametric Regression Fits," *The Annals of Statistics*, Vol 21(4), 1926-1947.

Härdle, W., and J. Marron, (1990), "Semiparametric Comparison of Regression Curves," *The Annals of Statistics*, Vol 18(1), 63-89.

Heckman, J., and V. J. Hotz, (1989), "Alternative Methods for Evaluating the Impact of Training Programs," (with discussion), *Journal of the American Statistical Association.*, 84(804): 862-874.

Heckman, J., R. Lalonde, and J. Smith (2000), "The Economics and Econometrics of Active Labor Markets Programs," in A. Ashenfelter and D. Card eds. Handbook of Labor Economics, vol. 3. New York: Elsevier Science.

Hirano, K., G. Imbens, and G. Ridder, (2003), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score,"*Econometrica*, 71(4): 1161-1189. July

Horowitz, J., and V. Spokoiny, (2001), "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69(3): 599-631.

Imbens, G., (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review, *Review of Economics and Statistics*, 86(1): 1-29.

Imbens, G., W. Newey and G. Ridder, (2003), "Mean-squared-error Calculations for Average Treatment Effects," unpublished manuscript, Department of Economics, UC Berkeley.

Lechner, M, (2002), "Some Practical Issues in the Evaluation of Heterogenous Labour Market Programmes by Matching Methods," *Journal of the Royal Statistical Society*, Series A, 165: 659–82.

Lee, M.-J., (2005), *Micro-Econometrics for Policy, Program, and Treatment Effects* Oxford University Press, Oxford.

Doksum, K., (1974), *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisco.

Neumeyer, N., and H. Dette, (2003), "Nonparametric Comparison of Regression Curves: An Empirical Process Approach," *The Annals of Statistics*, Vol 31, 880-920.

Pinkse, J., and P. Robinson, (1995), "Pooling Nonparametric Estimates of Regression Functions with a Similar Shape," in *Statistical Methods of Econometrics and Quantitative Economics: A Volume in Honour of C.R. Rao*, G.S. Maddala, P.C.B. Phillips and T.N. Srinivisan, eds., 172-197.

Rosenbaum, P., (1997), "The role of a second control group in an observational study", *Statistical Science*, (with discussion), Vol 2., No. 3, 292–316.

Rosenbaum, P., (2001), *Observational Studies*, second edition, Springer Verlag, New York.

Rosenbaum, P., and D. Rubin, (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70: 41-55.

Rubin, D. (1974), "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies," *Journal of Educational Psychology*, 66: 688-701.

Wooldridge, J., (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.

Table 1: Summary Statistics Experimental GAIN Data

	Los Angeles (LA) $N_T = 2995$, $N_C = 1400$		Riverside (RI) $N_T = 4405$, $N_C = 1040$		Alameda (AL) $N_T = 597$, $N_C = 601$		San Diego (SD) $N_T = 6978$, $N_C = 1154$	
	mean	(s.d.)	mean	(s.d.)	mean	(s.d.)	mean	(s.d.)
female	0.94	(0.24)	0.88	(0.33)	0.95	(0.22)	0.84	(0.37)
age	38.52	(8.43)	33.64	(8.20)	34.72	(8.62)	33.80	(8.59)
age-squared/100	15.55	(6.83)	11.99	(5.96)	12.79	(6.41)	12.16	(6.24)
hispanic	0.32	(0.47)	0.27	(0.45)	0.08	(0.26)	0.25	(0.44)
black	0.45	(0.50)	0.16	(0.36)	0.70	(0.46)	0.23	(0.42)
hsdiploma	0.35	(0.48)	0.52	(0.50)	0.59	(0.49)	0.57	(0.50)
1 child	0.33	(0.47)	0.39	(0.49)	0.42	(0.49)	0.43	(0.50)
children under 6	0.10	(0.30)	0.16	(0.37)	0.31	(0.46)	0.13	(0.34)
earnings q-1	0.22	(0.87)	0.45	(1.41)	0.21	(0.85)	0.59	(1.48)
earnings q-2	0.22	(0.88)	0.57	(1.55)	0.21	(0.87)	0.71	(1.68)
earnings q-3	0.23	(0.86)	0.60	(1.60)	0.20	(0.87)	0.76	(1.77)
earnings q-4	0.22	(0.87)	0.61	(1.60)	0.26	(1.02)	0.81	(1.88)
earnings q-5	0.20	(0.88)	0.67	(1.70)	0.25	(1.11)	0.83	(1.92)
earnings q-6	0.19	(0.81)	0.70	(1.76)	0.23	(0.89)	0.84	(1.90)
earnings q-7	0.19	(0.81)	0.71	(1.79)	0.26	(1.05)	0.84	(1.95)
earnings q-8	0.18	(0.80)	0.73	(1.84)	0.22	(1.01)	0.83	(1.96)
earnings q-9	0.18	(0.80)	0.72	(1.83)	0.23	(1.00)	0.83	(1.99)
earnings q-10	0.17	(0.74)	0.73	(1.82)	0.24	(1.09)	0.84	(2.01)
pos earn q-1	0.88	(0.33)	0.78	(0.41)	0.86	(0.34)	0.73	(0.44)
pos earn q-2	0.88	(0.33)	0.76	(0.42)	0.86	(0.34)	0.72	(0.45)
pos earn q-3	0.87	(0.33)	0.76	(0.43)	0.86	(0.34)	0.71	(0.45)
pos earn q-4	0.87	(0.33)	0.75	(0.43)	0.86	(0.34)	0.71	(0.45)
pos earn q-5	0.88	(0.32)	0.74	(0.44)	0.86	(0.35)	0.71	(0.46)
pos earn q-6	0.89	(0.31)	0.74	(0.44)	0.86	(0.35)	0.70	(0.46)
pos earn q-7	0.88	(0.33)	0.74	(0.44)	0.87	(0.34)	0.71	(0.45)
pos earn q-8	0.89	(0.32)	0.73	(0.44)	0.87	(0.33)	0.72	(0.45)
pos earn q-9	0.89	(0.31)	0.74	(0.44)	0.87	(0.33)	0.73	(0.45)
pos earn q-10	0.89	(0.31)	0.74	(0.44	0.87	(0.34	0.73	(0.44
Earnings Year 1	0.36	(1.02)	0.59	(1.24)	0.36	(1.04)	0.64	(1.33)

Table 2: Tests for Zero and Constant Average Treatment Effects

	Zero Cond. Ave TE			Constant Cond. Ave. TE			Zero Ave. TE		
County	chi-sq	(dof)	normal	chi-sq	(dof)	normal	chi-sq	(dof)	normal
LA	34.58	(29)	0.73	34.56	(28)	0.88	0.37	(1)	-0.61
RI	248.09	(29)	28.77	171.22	(28)	19.14	72.46	(1)	8.51
AL	46.68	(29)	2.32	46.52	(28)	2.48	0.04	(1)	0.21
SD	97.51	(29)	9.00	88.14	(28)	8.04	3.64	(1)	1.91