Testing Conditional Independence via Rosenblatt Transforms[1]

Kyungchul Song[2]

Department of Economics, University of Pennsylvania

September 5, 2007

Abstract

This paper investigates the problem of testing conditional independence of Y and Z given $\lambda_\theta(X)$ for some unknown $\theta \in \Theta \subset \mathbf{R}^d$, for a parametric function $\lambda_\theta(\cdot)$. For instance, such a problem is relevant in recent literatures of heterogeneous treatment effects and contract theory. First, this paper finds that using Rosenblatt transforms in a certain way, we can construct a class of tests that are asymptotically pivotal and asymptotically unbiased against $\sqrt{n}$-converging Pitman local alternatives. The asymptotic pivotalness is convenient especially because the asymptotic critical values remain invariant over different estimators of the unknown parameter θ. Even when tests are asymptotically pivotal, however, it is often the case that simulation methods to obtain asymptotic critical values are yet unavailable or complicated, and hence this paper suggests a simple wild bootstrap procedure. A special case of the proposed testing framework is to test the presence of quantile treatment effects in a program evaluation data set. Using the JTPA training data set, we investigate the validity of nonexperimental procedures for inferences about quantile treatment effects of the job training program.

Key words and Phrases: Conditional independence, asymptotic pivotal tests, Rosenblatt transforms, wild bootstrap.

JEL Classifications: C12, C14, C52.

[1] Part of this paper's results were circulated in its previous version: "Testing Conditional Independence using Conditional Martingale Transforms."

[2] I am grateful to Yoon-Jae Whang and Guido Kuersteiner who gave me valuable advice at the initial stage of this research. I also thank Robert de Jong, Qi Li, Hyungsik Roger Moon, Joon Park, Bernard Salanié, Norm Swanson and seminar participants at various workshops and conferences for their valuable comments. My special thanks go to Petra Todd who kindly gave me the data set of JTPA. All errors are mine. Address correspondence to: Kyungchul Song, Department of Economics, University of Pennsylvania, 528 McNeil Building, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297.

1 Introduction

Suppose that Y and Z are random variables, and let $\lambda_\theta(X)$ be a real valued function of a random vector X indexed by a parameter $\theta \in \Theta$. For example, we may consider $\lambda_\theta(X) = h(X'\theta)$ for some known function h. This paper investigates the problem of testing conditional independence of Y and Z given $\lambda_\theta(X)$ for some $\theta \in \Theta$:

$$Y \perp Z|\lambda_\theta(X) \text{ for some } \theta \in \Theta. \tag{1}$$

The function $\lambda_\theta(\cdot)$ is known up to a finite dimensional parameter $\theta \in \Theta$.

The conditional independence restriction is often used as part of the identifying restriction of an econometric model. In the literature of program evaluations, testing conditional independence of the observed outcome and the treatment decision given observable covariates can serve as testing lack of average or quantile treatment effects under the assumption of unconfoundedness (e.g. Rosenbaum and Rubin (1983) and Firpo (2006)), or serve as a means to evaluate the assumptions of nonexperimental procedures when a good quality experimental data set is available (Heckman, Ichimura, and Todd (1997)). A conditional independence restriction is, sometimes, a direct implication of an economic theory. For example, in the literature of insurance, the presence of positive conditional dependence between coverage and risk is known to be a direct consequence of adverse selection under information asymmetry. (e.g. Chiappori and Salanié (2000)).

The literature of testing conditional independence for continuous variables appears rather recent and includes relatively few researches as compared to that of other nonparametric or semiparametric tests. Linton and Gozalo (1999) and Delgado and González Manteiga (2001) proposed a bootstrap-based test of conditional independence. Su and White (2003a, 2003b, 2003c) studied several methods of testing conditional independence based on comparing conditional densities, characteristic functions, and empirical likelihoods. Angrist and Kuersteiner (2004) suggested asymptotically pivotal tests of conditional independence when Z is binary.

This paper's framework of hypothesis testing is based on an unconditional-moment formulation of the null hypothesis using test functions that run through an index space. The tests in our framework have the usual local power properties that are shared by other empirical-process based approaches such as Linton and Gozalo (1999), Delgado and González Manteiga (2001), and Angrist and Kuersteiner (2004). In contrast with Linton and Gozalo (1999) and Delgado and González Manteiga (2001), our tests are asymptotically pivotal (or asymptotically distribution free).[3]

[3] Although we can write the conditional independence restrictions as semiparametric conditional moment

Asymptotic pivotal tests have asymptotic critical values that do not change as we move from one data generating process to another within the null hypothesis. In general, the limiting distribution of tests based on an unconditional-moment formulation of the null hypothesis changes as we move from one specification of $\lambda_\theta(X)$ to another, or from one estimator of θ to another. The main contribution of this paper is to propose a class of tests that are asymptotically pivotal for a wide class of different specifications of $\lambda_\theta(X)$ and for different estimators of θ even with different rates of convergence.

There are two important merits of using asymptotic pivotal tests. First, when the tests are asymptotically pivotal and asymptotic critical values can be simulated, one can compute the asymptotic critical values once and for all, and use them for a variety of different specifications of $\lambda_\theta(X)$ or even different tests with the same limiting distribution without having to resort to the resampling method. For example, we can use the existing tables of asymptotic critical values in a special case when Z is binary (see Section 6). The asymptotic pivotal tests are convenient in particular, when the data are large and testing procedure is complicated so that the resampling procedure is cumbersome.

Second, even when the resampling procedure is a viable option, it is still better to use asymptotic pivotal tests in many regular cases. In some cases, the asymptotic critical values are hard to simulate even when the test is asymptotically pivotal. Furthermore, when one is interested in performing multiple tests that have different limiting distributions, simulations should be performed again every time one changes the test statistic, say, from a two-sided test to a one-sided one. Hence a bootstrap procedure may still be an attractive option in this situation. Then, it is well-known that in many regular cases, the bootstrap method applied to asymptotic pivotal tests shows asymptotic refinement property, and hence in this case one does better by using an asymptotic pivotal test in place of asymptotically nonpivotal one. This paper suggests a simple wild bootstrap procedure based on the proposal of Delgado and González Manteiga (2001). The proposed bootstrap procedure is simple and easy to implement as it does not require the computation of nonparametric estimators for each bootstrap sample.

The method we employ to obtain asymptotic pivotal tests is to apply a quantile transform $U = F_\theta(\lambda_\theta(X))$ of $\lambda_\theta(X)$ and employ Rosenblatt transforms that are applied to (Y, U) and

restrictions, the procedure of Song (2007) to obtain asymptotic pivotal tests does not apply to these restrictions.

The property of asymptotic pivotalness is shared, in particular, by Angrist and Kuersteiner (2004) who employed a martingale transform of Khmaladze (1993). However, they focus only on the case of Z being equal to binary and $P\{Z = 1|X\}$ is parametrically specified. Our framework does not require this restriction.

Several tests suggested by Su and White are also asymptotically pivotal, and consistent, but have different asymptotic power properties. Their tests are asymptotically unbiased against Pitman local alternatives that converge to the null at a rate slower than $\sqrt{n}$.

(Z, U) when both Y and Z are continuous, and applied to Y when Z is binary. Since the quantile transform and the conditional distribution functions in the Rosenblatt transform are unknown, we replace them by nonparametric estimators. The use of the quantile transform renders the test invariant to the estimation error of $\hat{\theta}$ in large samples, and the use of the Rosenblatt transform makes the test asymptotically pivotal. It is worth emphasizing that the Rosenblatt transform causes almost no additional computational cost as compared to other existing testing procedures of conditional independence. Most conditional independence tests involve one or two incidences of nonparametric estimation, and so does the method of Rosenblatt transform that we propose here.

As emphasized before, one of the most distinctive aspects of this paper's proposal is that the asymptotic pivotalness of the test is maintained even when we allow the conditioning variable $\lambda_\theta(X)$ to depend on the unknown parameter θ. This is particularly important when the dimension of the random vector X is large relative to the sample size. In such a situation, instead of testing

$$Y \perp Z | X,$$

it is reasonable to impose a single-index restriction upon X and test the hypothesis of the form in (1).[4] One can check the robustness of the result by varying the parametrization used for the single-index $\lambda_\theta(X)$. It is worth noting that the single-index structure in the conditioning variable, though it is set up as a parametric function, has a semiparametric nature in itself. Indeed, for any strictly increasing function $h(\cdot)$, the null hypothesis of $Y \perp Z | \lambda_\theta(X)$ and $Y \perp Z | h(\lambda_\theta(X))$ are equivalent.

The finite sample performance of the tests proposed in the paper are compared via a Monte Carlo simulation study. The tests we focus on are of two types: one using indicator functions in the construction of tests as in Stute (1997) and the other using exponential functions as in Bierens (1990). We investigate their performances under various data generating processes.

As an illustration, this paper applies the proposed method of testing to the problem of evaluating nonexperimental procedures. Since the seminal paper by LaLonde (1986), one of the main questions addressed in the literatures of program evaluation has been whether econometric methods applied to nonexperimental data set are reliable (see e.g. Heckman, Ichimura, and Todd (1997, 1998), Dehejia and Wahba (1999), and Smith and Todd (2001, 2005) and references therein). To address this question in the context of LaLonde (1986), Smith and Todd (2001, 2005), in particular, estimated treatment effects from the combined

[4] In general, $Y \perp Z | X$ does not imply $Y \perp Z | \lambda_\theta(X)$. (See Phillips (1987) or Dawid (1979) for example.) Hence the single-index restriction is an additional element that belongs to a model specification prior to the conditional independence restriction.

data set of the control group (from the randomized-out data) and the comparison group (from the nonexperimental data) using various econometric methods, and checked if the estimates are close to zero as they would if the nonexperimental procedures were valid.

In a similar spirit, we employ a data set from the Job Training Partnership Act (JTPA) training program used by Heckman, Ichimura, Smith and Todd (1998) (from here on, HIST) and evaluate the econometric methods of quantile treatment effects, yet from the perspective of hypothesis testing rather than that of estimation. This approach using the hypothesis testing framework seems reasonable as compared to the use of estimators when the results depend heavily on the specifics of the estimators used. Typically the questions addressed by hypothesis testing are simpler than estimators, and often requires less assumptions accordingly.

We first test whether the outcome variable of an untreated state is conditionally independent from the receipt of treatment given the propensity score. The test fails to reject the null of conditional independence. Second, we consider weaker implications needed for identification of quantile treatment effects. We also fail to reject the implications using the JTPA data. These results appear to be consistent with one part of the mixed results from HIST where conditional mean independence tests are performed with adjustment for estimation of the propensity score. With adjustment for estimation of the propensity score, their test fails to reject the null of conditional mean independence, while without adjustment, their test firmly rejects the null. However, they minimize the result with adjustment for estimation because for lower percentile values of propensity scores, the bias estimates are significantly different from zero. Our findings suggest that empirical evidence against the unconfoundedness assumption or its implications still appears weak in the JTPA data.

This paper is organized as follows. In the next section, we introduce the testing framework formally and discuss examples. In Section 3, we provide the first result that gives an asymptotic representation of the empirical process constituting the test statistic. Section 4 deals with the case where Z is binary. In Section 5, a bootstrap procedure is suggested and asymptotically justified. In Section 6, we propose a variant of the Rosenblatt-transform approach that utilizes the martingale transform of Khmaladze (1993). In Section 7, we present and discuss the results from the Monte Carlo simulation study. Section 8 is devoted to an empirical application that investigate the validity of nonexperimental procedures using the JTPA training data. In Section 9, we conclude. The mathematical proofs are relegated to the appendix.

2 Testing Conditional Independence

2.1 The Null Hypothesis and the Rosenblatt transform

Suppose that we are given a random vector (Y, Z, X) distributed by P and an unknown real valued function $\lambda_\theta(\cdot)$ on $\mathbf{R}^{d_X}$, $\theta \in \Theta \subset \mathbf{R}^d$. We are interested in testing the null hypothesis in (1). For example, we may consider $\lambda_\theta(X) = h(X'\theta)$ for a known function h. Hence a special case is testing conditional independence of Y and Z given $X'\theta$ for some $\theta \in \Theta$. We assume the following for (Y, Z).

Assumption 1C : The random vector (Y, Z) is absolutely continuous with respect to the Lebesgue measure on $\mathbf{R}^2$ with a continuous joint density function.

When either Y or Z is binary, the development of this paper's thesis becomes simpler, as we will see in a later section. The null hypothesis in (1) requires to check the conditional independence of Y and Z given $\lambda_\theta(X)$ for all $\theta \in \Theta$ until we find one θ that satisfies the conditional independence restriction. The condition (ii)(a) in the following simplifies this problem by making it suffice to focus on a specific θ_0 in Θ.

Assumption 2C : (i) The parameter space Θ is compact in $\mathbf{R}^{d_\Theta}$.
(ii) There exists a parameter $\theta_0 \in \Theta$ and $B(\theta_0, \delta) \triangleq \{\theta \in \Theta : ||\theta - \theta_0|| < \delta\}$, $\delta > 0$, such that
(a) $Y \perp Z | \lambda_{\theta_0}(X)$, whenever the null hypothesis of (1) holds,
(b) $\lambda_\theta(\cdot)$, $\theta \in B(\theta_0, \delta)$, is uniformly bounded and for any $\theta_1, \theta_2 \in B(\theta_0, \delta)$,

$$|\lambda_{\theta_1}(x) - \lambda_{\theta_2}(x)| \leq C||\theta_1 - \theta_2||, \text{ for some } C > 0,$$

(c) for each $\theta \in B(\theta_0, \delta)$, $\lambda_\theta(X)$ is absolutely continuous with respect to the Lebesgue measure and has a uniformly bounded density, and
(d) for each $\theta \in B(\theta_0, \delta)$, there exists a σ-finite measure w_θ that defines the conditional density function $f_\theta(y, z, x|\bar{\lambda}_1, \bar{\lambda})$ of (Y, Z, X) given $(\lambda_\theta(X), \lambda_0(X)) = (\bar{\lambda}_1, \bar{\lambda})$ where $f_\theta(y, z, x|\bar{\lambda}_1, \bar{\lambda})$ is twice continuously differentiable in $\bar{\lambda}_1$ in the support of $\lambda_\theta(X)$, with derivatives $f_\theta^{(j)}$, $j = 1, 2$, satisfying $\sup_{\theta \in B(\theta_0, \delta)} \int (f_\theta^{(j)})^{2+\delta'}(y, z, x|\bar{\lambda}_1, \bar{\lambda}) dw_\theta(y, z, x) < C$ for some $\delta', C > 0$.

For example, the parameter θ_0 in the above assumption can be defined as a unique solution to the following problem

$$\theta_0 = \arg\min_{\theta \in \Theta} \sup_{(y,z,\bar{\lambda}) \in \mathbf{R}^3} \left| \mathbf{E}\left[\gamma_z(Z)(\gamma_y(Y) - \mathbf{E}\left[\gamma_y(Y)|\lambda_\theta(X)\right])|\lambda_\theta(X) = \bar{\lambda}\right] \right|, \tag{2}$$

if the solution exists, where $\gamma_y(\cdot)$ and $\gamma_z(\cdot)$ are appropriate functions indexed by $y \in \mathbf{R}$ and $z \in \mathbf{R}$. For functions $\gamma_y(\cdot)$ and $\gamma_z(\cdot)$, one may take indicator functions $\gamma_z(Z) = 1\{Z \leq z\}$ and $\gamma_y(Y) = 1\{Y \leq y\}$. The uniform boundedness condition in (ii)(b) for Λ is innocuous, because by choosing a strictly increasing function Φ on $[0,1]$, we can redefine $\lambda'_\theta = \Phi \circ \lambda_\theta$. The Lipschitz continuity in (ii)(b) can be made to hold by choosing this Φ appropriately. When $\lambda_\theta(X) = X'\theta$ and Φ is chosen to be the standard normal distribution function, the condition (ii)(b) is satisfied. The absolute continuity condition in (c) is satisfied in particular when $\lambda_{\theta_0}(X) = h(X'\theta_0)$ with a continuous function h and there exists a continuous random variable X_j whose coefficient θ_{j0} is not zero. The condition (d) does not require the random vector X to be continuous. For example, the condition allows the σ-finite measure w_θ to have a marginal for X as a counting measure. Stute and Zhu (2005) use an analogous, weaker condition.[5]

Briefly we write $\lambda_0(X) = \lambda_{\theta_0}(X)$ and let F_0 be the distribution function of $\lambda_0(X)$. Also let $U \triangleq F_0(\lambda_0(X))$. Then under Assumptions 1 and 2, we can write the null hypothesis as (e.g. Theorem 9.2.1 in Chung (2001), p.322)

$$H_0 : P\left\{\mathbf{E}(\gamma_y(Y)|U,Z) = \mathbf{E}(\gamma_y(Y)|U), \text{ for all } y \in [0,1]\right\} = 1,$$

for an appropriate class of functions $\{\gamma_y(\cdot) : y \in [0,1]\}$. Let us define

$$\tilde{Z} \triangleq F_{Z|U}(Z|U) \text{ and } \tilde{Y} \triangleq F_{Y|U}(Y|U),$$

where $F_{Z|U}(\cdot|U)$ and $F_{Y|U}(\cdot|U)$ are conditional distribution functions of Y and Z given U respectively. Then $(\tilde{Z}, U)$ is distributed as the joint distribution of two independent uniform $[0,1]$ random variables, and so is $(\tilde{Y}, U)$. The transform of (Z, U) into $(\tilde{Z}, U)$ is called *the Rosenblatt transform*, due to Rosenblatt (1952). Under Assumption 2, we can write the null hypothesis equivalently as :

$$H_0 : P\left\{\mathbf{E}(\gamma_y(\tilde{Y})|U,\tilde{Z}) = \mathbf{E}(\gamma_y(\tilde{Y})), \text{ for all } y \in [0,1]\right\} = 1. \quad (3)$$

The alternative hypothesis is given by the negation of the null:

$$H_1 : P\left\{\mathbf{E}(\gamma_y(\tilde{Y})|U,\tilde{Z}) = \mathbf{E}(\gamma_y(\tilde{Y})), \text{ for all } y \in [0,1]\right\} < 1.$$

The next subsection provides examples that demonstrate the relevancy of this testing problem in empirical researches.

[5] See Condition A(i) of Theorem 2.1 there.

2.2 Examples

2.2.1 Heterogeneous Treatment Effects under Unconfoundedness Condition

In the literature of program evaluations, it is often of interest to see whether participating in a certain program or treatment has an effect on an individual's outcome. (See Rosenbaum and Rubin (1983), HIST, Hirano, Imbens, and Ridder (2003) and references therein.) Let Y_1 represent an outcome when the individual is treated and Y_0 an outcome when not treated, and Z denote a binary variable representing participation in the treatment. Let X be a vector of covariates such that conditional independence

$$(Y_0, Y_1) \perp Z \mid X \tag{4}$$

holds. Then it is well-known that (Rosenbaum and Rubin (1983)) when $P(X) \triangleq \mathbf{P}(Z = 1|X) \in (0,1)$, $(Y_0, Y_1) \perp Z \mid P(X)$. Consider testing the following restriction:

$$Y \perp Z \mid P(X), \tag{5}$$

where $Y = Y_0(1-Z) + Y_1 Z$ represents an observed outcome. Under this restriction, for any measurable function u,

$$\begin{aligned} & \mathbf{E}[u(Y_0) - u(Y_1)|P(X)] \\ = \; & \mathbf{E}[u(Y)|P(X), Z = 0] - \mathbf{E}[u(Y)|P(X), Z = 1] \\ = \; & \mathbf{E}[u(Y)|P(X)] - \mathbf{E}[u(Y)|P(X)] = 0. \end{aligned}$$

Hence this implies that $\mathbf{E}[u(Y_0) - u(Y_1)|P(X)] = 0$. When the conditional independence restriction in (5) holds, there is no evidence not only of the (local) average treatment effect, but also of the (conditional) quantile treatment effects. Therefore, we may be interested in testing $Y \perp Z \mid P(X;\theta)$ where $P(X;\theta)$ is a parametrization of $P(X)$.

In many cases, the condition in (4) is more restrictive than is needed for identifying treatment effects parameters at hand. For example, to identify the mean effect of treatment on the treated or its average version, it often suffices to assume that

$$Y_0 \perp Z \mid P(X)$$

or that Y_0 is conditionally mean independent of Z given $P(X)$.(Heckman, Ichimura, and Todd (1997)). We can test the above null hypothesis when experimental data are available.

2.2.2 Testing Identifying Assumptions for Quantile Treatment Effects

As mentioned before, when experimental data are available, we can test a variety of identifying assumptions that are used to estimate a variety of quantities representing treatment effects. Among the examples are quantile treatment effects that we focus on here (e.g. Abadie, Angrist, and Imbens (2002) and Firpo (2007)). Let $q_{0\tau}$ and $q_{1\tau}$ be the τ-th quantiles of Y_0 and Y_1. Under the assumption that

$$1\{Y_1 \leq q_{1\tau}\} \perp Z \mid X \text{ and } 1\{Y_0 \leq q_{0\tau}\} \perp Z \mid X, \tag{6}$$

the quantile treatment effect $\Delta_\tau = q_{1\tau} - q_{0\tau}$ is identified as $q_{1\tau}$ and $q_{0\tau}$ are identified from (Firpo (2007))

$$\mathbf{E}\left[\frac{Z}{p(X)} 1\{Y \leq q_{1\tau}\}\right] = \tau \text{ and } \mathbf{E}\left[\frac{1-Z}{1-p(X)} 1\{Y \leq q_{0\tau}\}\right] = \tau.$$

Let F_1 and F_0 be distribution functions of Y_1 and Y_0. By a theorem of Rosenbaum and Rubin (1983) and using a quantile transform Y_1 and Y_0 into $F_1(Y_1)$ and $F_0(Y_0)$, the restrictions in (6) imply[6]

$$1\{F_1(Y_1) \leq \tau\} \perp Z \mid P(X) \text{ and } 1\{F_0(Y_0) \leq \tau\} \perp Z \mid P(X),$$

if $P(X) \in (0,1)$. Using experimental data, we can test the conditional independence restrictions in the above.

The test can be used to test the identifying restrictions of *conditional* quantile treatment effects. Let $q_{0\tau}(X)$ and $q_{1\tau}(X)$ be defined from the relation

$$\tau = P\{Y_0 \leq q_{0\tau}(X)|P(X)\} \text{ and } \tau = P\{Y_1 \leq q_{1\tau}(X)|P(X)\}.$$

The conditional quantile treatment effect can be defined as $\Delta_\tau(X) = q_{0\tau}(X) - q_{1\tau}(X)$. Let $F_1(Y_1|P(X))$ and $F_0(Y_0|P(X))$ be the conditional quantile transforms of Y_1 and Y_0, i.e., where $F_1(\cdot|P(X))$ and $F_0(\cdot|P(X))$ are conditional distribution functions of Y_1 given $P(X)$ and Y_0 given $P(X)$. Then the identifying restrictions of conditional quantile treatment effects are given by

$$1\{F_1(Y_1|P(X)) \leq \tau\} \perp Z \mid X \text{ and } 1\{F_0(Y_0|P(X)) \leq \tau\} \perp Z \mid X. \tag{7}$$

[6] For any random vector Y, a binary variable Z and random vector X such that $P(X) = P\{Z = 1|X\} \in (0,1)$, it can be shown that $Y \perp Z \mid X$ always implies $Y \perp Z|P(X)$ following steps in the proofs of Theorems 2 and 3 in Rosenbaum and Rubin (1983).

Under these restrictions, we can identify $q_{1\tau}(X)$ and $q_{0\tau}(X)$ from

$$\mathbf{E}\left[\frac{Z}{P(X)}1\{Y \leq q_{1\tau}(X)\}|X\right] = \tau \text{ and } \mathbf{E}\left[\frac{1-Z}{1-P(X)}1\{Y \leq q_{0\tau}(X)\}|X\right] = \tau.$$

The conditional independence restrictions in (7) imply the restrictions

$$1\{F_1(Y_1|P(X)) \leq \tau\} \perp Z \mid P(X) \text{ and } 1\{F_0(Y_0|P(X)) \leq \tau\} \perp Z \mid P(X). \tag{8}$$

It is interesting to observe that the nonparametric components $F_0(Y_0|P(X))$ and $F_1(Y_1|P(X))$ are identified as part of Rosenblatt transforms of $(Y_0, P(X))$ and $(Y_1, P(X))$. As we demonstrate later, testing the restrictions in (8) is a special case of the testing framework that this paper proposes.

2.2.3 Testing Identifying Assumptions for IV Quantile Treatment Effects

Conditional independence assumption is also used in the IV quantile treatment effects model proposed by Chernozhukov and Hansen (2007). Suppose the conditional latent outcome Y_d given $X = x$ is determined through

$$Y_d(x) = q(d, x, U_d),$$

where U_d follows uniform over $[0, 1]$. One of the main conditions used by Chernozhukov and Hansen (2005) is that for some random vector Z, U_d and Z are independent given $X = x$. By Lemma 4.1 of Dawid (1979), this yields a testable implication that $Y_d(X)$ and Z are independent given $X = x$. This implication is directly testable when an experimental data set is available. Using the combined data set of randomized-out participants who applied for the program and who did not apply for the program, we can test

$$Y_0(X) \perp Z \text{ given } X.$$

When X is a vector of large dimension relative to the size of the data set, we can employ a single index restriction and test the null hypothesis of the form in (1).

2.2.4 Contract Theory

In the literature of contract theory, there has been an interest in testing the relevance of asymmetric information in data. Under asymmetric information, it is known that the risk is positively related to the coverage of the contract conditional on all publicly observed variables. This fact motivated several researchers to investigate the presence of such a relationship in various data sets.(See Cawley and Phillipson (1999), Chiappori and Salanié (2000), Chiappori, Jullien, Salanié, and Salanié (2002) and references therein.)

In this situation, the null hypothesis is conditional independence of the risk and the coverage, but the alternative hypothesis is restricted to positive dependence of the risk (or probability of claims in the insurance data) and the coverage of the insurance contract conditional on all observable variables. This paper's framework can be used to deal with this case of one-sided testing.

3 Asymptotic Representation of a Semiparametric Empirical Process

3.1 Asymptotic Representation

The null hypothesis of (3) is a conditional moment restriction. In this paper, we take γ_z and γ_y to be indicator functions:

$$\gamma_z(Z) = 1\{Z \leq z\} \text{ and } \gamma_y(Y) = 1\{Y \leq y\}$$

and write the null hypothesis as

$$H_0 : \mathbf{E}[\beta_u(U_i)\{\gamma_z(\tilde{Z}_i) - z\}\{\gamma_y(\tilde{Y}_i) - y\}] = 0, \ \forall (y, z, u) \in [0, 1]^3 \tag{9}$$

for an appropriate class of functions $\beta_u(\cdot)$. The function β_u can be any function that makes the null hypotheses in (1) and (9) equivalent under Assumptions 1 and 2. For example, one can take $\beta_u(U) = 1\{U \leq u\}$ as in Stute (1997) and Andrews (1997), or $\beta_u(U) = \exp(Uu)$ as in Bierens (1990) and Bierens and Ploberger (1997). For a general discussion of the class of functions for consistent tests, see Stinchcombe and White (1998). We introduce the following assumption for $\beta_u(\cdot)$.

Assumption 3 : The function $\beta_u(\cdot), \ u \in [0, 1]$, is uniformly bounded in $[0, 1]$ and for each $u \in [0, 1]$, $\beta_u(\cdot)$ is of bounded variation.

Assumption 3 is satisfied by most functions used in the literature. In particular, the indicator functions in Stute (1997) and the exponential functions in Bierens (1990) satisfy the condition.

Corresponding to the equivalent formulation of the null hypothesis, a test statistic can be constructed from the process:

$$\nu_n(r) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \beta_u(U_i)(\gamma_z(\tilde{Z}_i) - z)(\gamma_y(\tilde{Y}_i) - y), \; r = (y, z, u) \in [0, 1]^3.$$

Since we do not know $F_0(\cdot)$, $F_{Z|U}(\cdot)$ and $F_{Y|U}(\cdot)$, we need to replace them by the estimators. Let $\hat{\theta}$ be a consistent estimator of θ_0. As for the data and the estimator $\hat{\theta}$, we assume the following.

Assumption 4 : (i) $(Y_i, Z_i, X_i)_{i=1}^n$ is a random sample from P.
(ii) There exists an estimator $\hat{\theta}$ of θ_0 in Assumption 2 such that $||\hat{\theta} - \theta_0|| = o_P(n^{-1/4})$ both under the null hypothesis and under the alternatives.

The estimator $\hat{\theta}$ in (ii) can be obtained from a sample version of the problem in (2). It is worth noting that we do not require the moment conditions for (Y, Z, X). This is mainly due to the fact that the functions indexing the process $\nu_n(r)$ are uniformly bounded. Consistency of $\hat{\theta}$ can be obtained using the approach of Chen, Linton, and van Keilegom (2003). In most cases, it can be shown that $\hat{\theta}$ is $\sqrt{n}$- or $n^{1/3}$-consistent, satisfying the condition (ii).

As for the estimators of $F_0(\cdot)$, $F_{Z|U}(\cdot)$ and $F_{Y|U}(\cdot)$, we define

$$\hat{U}_i \triangleq F_{n,\hat{\theta},i}(\lambda_{\hat{\theta}}(X_i)), \; \hat{Z}_i = \hat{F}_{Z|U,i}(Z_i|\hat{U}_i) \text{ and } \hat{Y}_i = \hat{F}_{Y|U,i}(Y_i|\hat{U}_i)$$

where $F_{n,\hat{\theta},i}(\cdot)$ denotes the empirical distribution function of $\{\lambda_{\hat{\theta}}(X_i)\}_{i=1}^n$ with the omission of the i-th data point $\lambda_{\hat{\theta}}(X_i)$, i.e., $F_{n,\hat{\theta},i}(\lambda_{\hat{\theta}}(X_i)) = \frac{1}{n-1} \sum_{j=1,j\neq i}^{n} 1\{\lambda_{\hat{\theta}}(X_j) \leq \lambda_{\hat{\theta}}(X_i)\}$, and $\hat{F}_{Y|U,i}(y|u)$ and $\hat{F}_{Z|U,i}(z|u)$ are kernel estimators of $F_{Y|U}(y|u)$ and $F_{Z|U}(z|u)$. More specifically,

$$\hat{F}_{Y|U,i}(y|u) = \frac{\sum_{j=1,j\neq i}^{n} 1\{Y_j \leq y\} K_h(\hat{U}_j - u)}{\sum_{j=1,j\neq i}^{n} K_h(\hat{U}_j - u)} \tag{10}$$

where $K_h(x) = K(x/h)/h$, $K(\cdot)$ is a kernel function and h is the bandwidth parameter.[7] We similarly define $\hat{F}_{Z|U,i}(z|u)$. As for the kernel and the bandwidth parameter, we assume the following.

Assumption 5 : (i) K is a symmetric kernel with compact support, twice continuously differentiable with $\int K = 1$, and is nonincreasing on the positive real numbers. (ii) for an arbitrarily small $\varepsilon > 0$, $n^{-1/2+\varepsilon}h^{-2} + n^{1/2+\varepsilon}h^4 \to 0$.

[7] Although we may use the true density function of the uniform random variate U in the denominator of the estimator (e.g. Stute and Zhu (2005), finite sample performances seem better when we use the density estimators.

Assumption 5(i) is also used by Stute and Zhu (2005). The bandwidth condition in (ii) requires bandwidths larger than those prescribed by Stute and Zhu (2005) who requires $n^{-1/2}h^{-1} + n^{1/2}h^2 \to 0$.[8] The main reason for this larger bandwidth is due to our method of deriving asymptotics via linearization of the indicator functions γ_y and γ_z. The feasible version of $\nu_n(r)$ is given by

$$\hat{\nu}_n(r) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \beta_u(\hat{U}_i) \left\{ \gamma_z(\hat{Z}_i) - z \right\} \left\{ \gamma_y(\hat{Y}_i) - y \right\}.$$

Let $l_\infty([0,1]^3)$ denote the space of functions on $[0,1]^3$ that are uniformly bounded, and we endow $l_\infty([0,1]^3)$ with a uniform norm $||\cdot||_\infty$ defined by $||f||_\infty = \sup_{u \in [0,1]^3} |f(u)|$. The notation $\rightsquigarrow$ denotes weak-convergence in $l_\infty([0,1]^3)$ in the sense of Hoffman-Jorgensen (e.g. van der Vaart and Wellner (1996)).

Theorem 1 : *Suppose that* Assumptions 1C–2C, 3-5 *hold. Then the following holds.*

(i) *Both under H_0 and under local alternatives P_n such that* $\sup_{(z,y) \in [0,1]^2} |\tilde{F}_{n,Z|Y}(z|y) - z| = O(n^{-1/2+\varepsilon_1})$ *as $n \to \infty$ for some $\varepsilon_1 \in (0, \varepsilon/2)$, where ε is the constant in* Assumption 5(ii) *and $\tilde{F}_{n,Z|Y}(z|y)$ is the conditional distribution function of $\tilde{Z}_i$ given $\tilde{Y}_i = y$ under P_n,*

$$\sup_{r \in [0,1]^3} |\hat{\nu}_n(r) - \nu_n(r)| = o_P(1). \tag{11}$$

(ii)(a) *Under the null hypothesis, $\hat{\nu}_n \rightsquigarrow \nu$ in $l_\infty([0,1]^3)$, where ν is a Gaussian process whose covariance kernel is given by*

$$c(r_1; r_2) = \left\{ \int \beta_{u_1}(u)\beta_{u_2}(u)du \right\} (z_1 \wedge z_2 - z_1 z_2)(y_1 \wedge y_2 - y_1 y_2).$$

(b) *Under the alternatives, $n^{-1/2}\hat{\nu}_n(r) \rightsquigarrow \mathbf{E}\left[\beta_u(U)(\gamma_z(\tilde{Z}) - z)(\gamma_y(\tilde{Y}) - y)\right]$ in $l_\infty([0,1]^3)$.*

The results of Theorem 1 lead to the asymptotic properties of tests based on the process $\hat{\nu}_n(r)$. The asymptotic representation in (11) shows an interesting fact that the process $\hat{\nu}_n(r)$ is asymptotically equivalent to its infeasible counterpart $\nu_n(r)$. Recall that such a phenomenon is an exception rather than a norm in nonparametric tests. The local alternatives in (i) include Pitman local alternatives that converge to the null hypothesis at the rate of $\sqrt{n}$. It is worth noting that the estimation error in $\lambda_{\hat{\theta}}$ does not play a role in determining the limit behavior of the process $\hat{\nu}_n(r)$ even when the rate of convergence of $\hat{\theta}$ is slower

[8] In Stute and Zhu (2005), this is the condition corresponding to the case of X supported in a compact set. (See Remark 2.5 there.)

than $n^{-1/2}$. In other words, the results of Theorem 1 do not change when we replace $\lambda_{\hat{\theta}}$ by λ_0. This phenomenon is discovered and analyzed by Stute and Zhu (2005) in the context of testing single-index restrictions. See also Escanciano and Song (2007) for similar results under series estimation. This convenient phenomenon is due to the use of the empirical quantile transform of $\{\lambda_{\hat{\theta}}(X_i)\}_{i=1}^n$.

Based on the result in Theorem 1, we can construct a test statistic

$$T_n = \Gamma\hat{\nu}_n \tag{12}$$

by taking an appropriate continuous functional Γ. For example, in the case of two sided tests, we may take

$$\Gamma_{KS}\hat{\nu}_n = \sup_{r\in[0,1]^3} |\hat{\nu}_n(r)| \text{ or } \Gamma_{CM}\hat{\nu}_n = \left(\int_{[0,1]^3} \hat{\nu}_n(r)^2 dr\right)^{1/2}. \tag{13}$$

The first example is of Kolmogorov-Smirnov type and the second one is of Cramér-von Mises type. In the case of one-sided tests in which we test the null of conditional independence against conditional positive dependence, we may take

$$\Gamma_{KS}^{+}\hat{\nu}_n = \sup_{r\in[0,1]^3} \hat{\nu}_n(r) \text{ or } \Gamma_{CM}^{+}\hat{\nu}_n = \left(\int_{[0,1]^3} \max\{\hat{\nu}_n(r), 0\}^2 dr\right)^{1/2}.$$

The asymptotic properties of the tests based on $\Gamma\hat{\nu}_n$ follow from Theorem 1. Indeed, under the null hypothesis,

$$T_n = \Gamma\hat{\nu}_n \to_d \Gamma\nu. \tag{14}$$

The test is asymptotically pivotal, but in most cases other than the case of ν being a standard Brownian sheet, it is not known how to simulate the Gaussian process ν. In a later section, we suggest a bootstrap method.

4 The Case with Binary Z_i

In some applications, either Y or Z is binary. We consider the case where Z is a binary variable taking values in $\{0, 1\}$, and the variables Y and $\lambda_0(X)$ are continuous. In this case, the null hypothesis becomes (after the Rosenblatt transform of Y)

$$\mathbf{E}[\beta_u(U)\{Z - p(U)\}\{\gamma_y(\tilde{Y}) - y\}] = 0 \text{ for all } (y, u) \in [0, 1]^2 \tag{15}$$

by the binary nature of Z. Here $p(U) = P\{Z = 1|U\}$. We assume the following in place of Assumptions 1C-2C.

Assumption 1B : The random variable Y is absolutely continuous with respect to the Lebesgue measure on $\mathbf{R}$ with a continuous joint density function.

Assumption 2B : (i) The parameter space Θ is compact in $\mathbf{R}^{d_\Theta}$.
(ii) There exists a parameter $\theta_0 \in \Theta$ and $B(\theta_0, \delta) \triangleq \{\theta \in \Theta : ||\theta - \theta_0|| < \delta\}$, $\delta > 0$, such that conditions of Assumption 2C(a)-(c) hold and (d) for each $\theta \in B(\theta_0, \delta)$, there exists a σ-finite measure w_θ that defines the conditional density function $f_\theta(y, x|\bar{\lambda}_1, \bar{\lambda})$ of (Y, X) given $(\lambda_\theta(X), \lambda_0(X)) = (\bar{\lambda}_1, \bar{\lambda})$ where $f_\theta(y, x|\bar{\lambda}_1, \bar{\lambda})$ is twice continuously differentiable in $\bar{\lambda}_1$ in the support of $\lambda_\theta(X)$, with the derivatives $f_\theta^{(j)}$, $j = 1, 2$, satisfying

$$\sup_{\theta \in B(\theta_0, \delta)} \int (f_\theta^{(j)})^{2+\delta'}(y, x|\bar{\lambda}_1, \bar{\lambda}) dw_\theta(y, x) < C$$

for some $\delta', C > 0$.

Let $\hat{p}(U)$ be a nonparametric estimator of $p(U)$. For example, we may define

$$\hat{p}(u) = \frac{\sum_{i=1}^n Z_i K_h(\hat{U}_i - u)}{\sum_{i=1}^n K_h(\hat{U}_i - u)}. \tag{16}$$

We can alternatively use series estimation to obtain $\hat{p}(u)$. For the function $p(u)$ and its estimator $\hat{p}(u)$, we assume the following.

Assumption 6 : (i) For some $\varepsilon > 0$, $p(u) \in (\varepsilon, 1 - \varepsilon)$ for all $u \in [0, 1]$.
(ii) As for the estimator $\hat{p}(U)$, $\sup_{u \in [0,1]} |\hat{p}(u) - p(u)| = o_P(1)$.

The conditions for the kernel and the bandwidths in (16) are subsumed in the condition (ii) above. Lower level conditions can be found in the appendix (see Lemma B3 there.) We suggest two step cross-validations for these bandwidths in a section devoted to simulation studies.

We consider the following process

$$\bar{\nu}_n(u, y) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\beta_u(\hat{U}_i)\{Z_i - \hat{p}(\hat{U}_i)\}\{\gamma_y(\hat{Y}_i) - y\}}{\sqrt{\hat{p}(\hat{U}_i) - \hat{p}(\hat{U}_i)^2}}, \tag{17}$$

where $\hat{Y}_i$ and $\hat{U}_i$ are as defined before, and establish the asymptotic representation for this process.

Theorem 2 : *Suppose that* Assumptions 1B-2B, 3-6 *hold. Then the following holds.*

(i) *Both under* H_0 *and local alternatives* P_n *such that* $\sup_{y\in[0,1]}|P_n\{Z_i=1|\tilde{Y}_i=y\}-P_n\{Z_i=1\}|=O(n^{-1/2+\varepsilon_1})$ *for some* $\varepsilon_1\in(0,\varepsilon/2)$, *where* ε *is the constant in* Assumption 5(ii),

$$\sup_{(u,y)\in[0,1]^2}\left|\bar{\nu}_n(u,y)-\frac{1}{\sqrt{n}}\sum_{i=1}^n\frac{\beta_u(U_i)\{Z_i-p(U_i)\}}{\sqrt{p(U_i)-p(U_i)^2}}\left\{\gamma_y(\tilde{Y}_i)-y\right\}\right|=o_P(1).$$

(ii)(a) *Under* H_0, $\bar{\nu}_n\rightsquigarrow\bar{\nu}$ *in* $l_\infty([0,1]^2)$, *where* $\bar{\nu}$ *is a Gaussian process on* $[0,1]^2$ *whose covariance kernel is given by*

$$c(u_1,y_1;u_2,y_2)\triangleq\left\{\int\beta_{u_1}(u)\beta_{u_2}(u)du\right\}\{y_1\wedge y_2-y_1y_2\}.$$

(b) *Under the alternatives,*

$$n^{-1/2}\bar{\nu}_n(u,y)\rightsquigarrow\mathbf{E}\left[\frac{\beta_u(U)\{Z-p(U)\}\{\gamma_y(\tilde{Y})-y\}}{\sqrt{p(U)-p(U)^2}}\right]\ in\ l_\infty([0,1]^2).$$

We can construct test statistics using appropriate functional $\Gamma:\bar{T}_n=\Gamma\bar{\nu}_n$. For example,

$$\Gamma_{KS}\bar{\nu}_n=\sup_{(u,y)\in[0,1]^2}|\bar{\nu}_n(u,y)|\text{ or }\Gamma_{CM}\bar{\nu}_n=\left(\int_{[0,1]^2}\bar{\nu}_n(u,y)^2d(u,y)\right)^{1/2}.$$

When the Gaussian process $\bar{\nu}$ can be simulated, the asymptotic critical values can be read from the distribution of $\Gamma\bar{\nu}$. However, as emphasized before, often this is not the case in general, in particular when β_u is taken to be other than indicator functions.

5 Bootstrap Tests

The wild bootstrap method has been known to perform well in nonparametric and semiparametric tests. (Härdle and Mammen (1993), Stute, González and Presedo (1998), Whang (2000), and Delgado and González Manteiga (2001).) It is also known that under certain settings, the wild bootstrap method shows higher order improvement when the test statistic is asymptotically pivotal (Liu (1988) and Mammen (1993)).

Let $(\{\omega_{i,b}\}_{i=1}^n)_{b=1}^B$ be an i.i.d. sequence of random variables that are bounded, independent of $\{Y_i,Z_i,X_i\}$ and satisfy that $\mathbf{E}(\omega_{i,b})=0$ and $\mathbf{E}(\omega_{i,b}^2)=1$. For example, one can take $\omega_{i,b}$ to be a random variable with two-point distribution assigning masses $(\sqrt{5}+1)/(2\sqrt{5})$ and $(\sqrt{5}-1)/(2\sqrt{5})$ to the points $-(\sqrt{5}-1)/2$ and $(\sqrt{5}+1)/2$. Then, we consider the following bootstrap empirical process:

$$\nu_{n,b}^{*}(r)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_{i,b}\beta_{u}(\hat{U}_{i})\left\{\gamma_{z}(\hat{Z}_{i})-z\right\}\left\{\gamma_{y}(\hat{Y}_{i})-y\right\},\ b=1,\cdots,B.$$

The bootstrap empirical process $\nu_{n,b}^{*}(r)$ is similar to those proposed by Delgado and González Manteiga (2001). Given a functional Γ, we take bootstrap test statistics $T_{n,b}^{*}=\Gamma\nu_{n,b}^{*}$, $b=1,\cdots,B$. The critical values for the tests based on $T_{n,b}^{*}$ are approximated by $c_{\alpha,n,B}=\inf\{t: B^{-1}\Sigma_{b=1}^{B}1\{T_{n,b}^{*}>t\}\geq 1-\alpha\}$. Hence an α-level bootstrap test is obtained by $1\{T_{n}>c_{\alpha,n,B}\}$ where T_{n} is as defined in (12).

When Z_{i} is binary, we use the following bootstrap empirical process:

$$\bar{\nu}_{n,b}^{*}(r)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_{i,b}\beta_{u}(\hat{U}_{i})\frac{Z_{i}-\hat{p}(\hat{U}_{i})}{\sqrt{\hat{p}(\hat{U}_{i})-\hat{p}(\hat{U}_{i})^{2}}}\left\{\gamma_{y}(\hat{Y}_{i})-y\right\},\ b=1,\cdots,B,$$

and construct the bootstrap test statistic $\bar{T}_{n,b}^{*}=\Gamma\bar{\nu}_{n,b}^{*}$, $b=1,\cdots,B$, using an appropriate functional Γ. An approximate critical value can be obtained by $c_{\alpha,n,B}=\inf\{t: B^{-1}\Sigma_{b=1}^{B}1\{\bar{T}_{n,b}^{*}>t\}\geq 1-\alpha\}$, yielding an α-level bootstrap test: $\bar{\varphi}\triangleq 1\{\bar{T}_{n}>c_{\alpha,n,B}\}$, where $\bar{T}_{n}$ is as defined in Section 4. Let $F_{T_{n}^{*}}^{*}$ denote the conditional distribution of the bootstrap test statistic T_{n}^{*} given the sample $\{S_{i}\}_{i=1}^{n}$ and d be a distance metrizing weak convergence on the real line.

Theorem 3: (i) Suppose that the conditions of Theorem 1 hold under H_{0}. Then under H_{0},

$$d\left(F_{T_{n}^{*}}^{*},F_{\Gamma\nu}\right)\to 0\text{ in }P.$$

(ii) Suppose that the conditions of Theorem 2 hold under H_{0}. Then under H_{0},

$$d\left(F_{\bar{T}_{n}^{*}}^{*},F_{\Gamma\bar{\nu}}\right)\to 0\text{ in }P.$$

The wild bootstrap procedure is easy to implement. In particular, it does not require nonparametric estimation for each bootstrap sample.

6 The Method of Martingale Transforms

In this subsection, we develop a variant of the Rosenblatt-transformed test by applying the method of martingale transforms. The resulting tests are still asymptotically pivotal with different limiting distributions. More importantly, the martingale transform involved is extremely simple, adding no additional computational cost to the Rosenblatt-transform based test proposed earlier. The asymptotic power properties of the martingale-transform tests are different from the Rosenblatt-transformed test, and hence can be considered as an alternative. Since the development of the tests along this line can be proceeded in the similar manner as before, we summarize the results here.

First, suppose that Y and Z are continuous. Following Khmaladze (1993), we define $\mathcal{K}$ as follows:[9]

$$\gamma_y^{\mathcal{K}}(\tilde{Y}) \triangleq \gamma_y(\tilde{Y}) + \log(1 - \tilde{Y} \wedge y) \text{ and } \gamma_z^{\mathcal{K}}(\tilde{Z}) \triangleq \gamma_z(\tilde{Z}) + \log(1 - \tilde{Z} \wedge z). \tag{18}$$

Such a transform $\mathcal{K}$ is called a *martingale transform* as the empirical process indexed by the transformed functions weakly converges to a martingale in the univariate case. Using the martingale transform, we can reformulate the null hypothesis as follows:

$$H_0 : \mathbf{E}[\beta_u(U_i)\gamma_z^{\mathcal{K}}(\tilde{Z}_i)\gamma_y^{\mathcal{K}}(\tilde{Y}_i)] = 0, \ \forall (y, z, u) \in [0,1]^3.$$

The test statistic is based on the empirical version of the null hypothesis:

$$\hat{\nu}_n^{MT}(r) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \beta_u(\hat{U}_i)\gamma_z^{\mathcal{K}}(\hat{Z}_i)\gamma_y^{\mathcal{K}}(\hat{Y}_i),$$

where $\hat{Z}_i$ and $\hat{Y}_i$ are estimated Rosenblatt-transforms as before. Observe that the computation of the empirical process $\hat{\nu}_n^{MT}(r)$ is no more complicated than the process $\hat{\nu}_n(r)$. The asymptotic theory along the line of Theorem 1 tells us that[10]

$$\hat{\nu}_n^{MT} \rightsquigarrow \nu^{MT} \text{ in } l_\infty([0,1] \times [0,1)^2)$$

where ν^{MT} is a Gaussian process with covariance kernel $C(r_1; r_2) = \left\{\int \beta_{u_1}(u)\beta_{u_2}(u)du\right\}(z_1 \wedge z_1)(y_1 \wedge y_1)$. Observe that when $\beta_u(U) = 1\{U \leq u\}$, the Gaussian process ν^{MT} is a standard Brownian sheet. The test statistic can be constructed as $T_n^{MT} = \Gamma\hat{\nu}_n^{MT}$ using a known functional Γ. The bootstrap procedure can be followed as analogously before by considering

[9] The martingale transform after the Rosenblatt-transform is a special case of a conditional martingale transform proposed by Song (2007). Indeed, the martingale transform satisfies the properties of conditional isometry and conditional orthogonality where the conditioning variable is U_i here.

[10] The domain of weak convergence is not $[0,1]^3$ but $[0,1] \times [0,1)^2$. This is because when $z = y = 1$, the martingale transform is not defined. The proof of this weak convergence is available upon request.

the following bootstrap empirical process:

$$\nu_{n,b}^{*MT}(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_{i,b} \beta_u(\hat{U}_i) \gamma_z^{\mathcal{K}}(\hat{Z}_i) \gamma_y^{\mathcal{K}}(\hat{Y}_i), \ b = 1, \cdots, B,$$

and taking the empirical critical values of the bootstrap test statistics $T_{n,b}^{MT*} = \Gamma \nu_{n,b}^{*MT}$, $b = 1, \cdots, B$.

When Z is binary and Y is continuous, we may consider the following equivalent formulation of the null hypothesis:

$$H_0 : \mathbf{E}[\beta_u(U_i)(Z - p(U))\gamma_y^{\mathcal{K}}(\tilde{Y}_i)] = 0, \ \forall (y, z, u) \in [0, 1]^3$$

and the following martingale transformed process

$$\bar{\nu}_n^{MT}(u, y) \triangleq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\beta_u(\hat{U}_i) \left\{ Z_i - \hat{p}(\hat{U}_i) \right\} \gamma_y^{\mathcal{K}}(\hat{Y}_i)}{\sqrt{\hat{p}(\hat{U}_i) - \hat{p}(\hat{U}_i)^2}}$$

using the estimator $\hat{p}(\hat{U}_i)$ as before. Following the steps of the proof of Theorem 2, we can show that

$$\bar{\nu}_n^{MT} \rightsquigarrow \bar{\nu}^{MT} \text{ in } l_\infty([0, 1] \times [0, 1))$$

where $\bar{\nu}^{MT}$ is a Gaussian process with covariance kernel $C(r_1; r_2) = \left\{ \int \beta_{u_1}(u) \beta_{u_2}(u) du \right\} (y_1 \wedge y_1)$. It is worth noting that when $\beta_u(U) = 1\{U \leq u\}$, the Gaussian process ν^{MT} is a two-parameter standard Brownian sheet. The test statistic can be constructed as $\bar{T}_n^{MT} = \Gamma \bar{\nu}_n^{MT}$ using a known functional Γ. In the case of Kolmogorov-Smirnov functional

$$\Gamma \nu = \sup_{(u,y) \in [0,1] \times [0,1)} \nu(u, y)$$

and $\beta_u(U) = 1\{U \leq u\}$, the asymptotic critical values are available from the simulations of Dr. Ray Brownrigg and can be found at his website: http://www.mcs.vuw.ac.nz/~ray/Brownian:

significance level	0.5	0.25	0.20	0.10	0.05	0.025	0.01
critical values	1.46	1.81	1.91	2.21	2.46	2.70	3.03

However, as mentioned before, the asymptotic critical values are not available in general for other choices of β_u. Alternatively, one can use a bootstrap test that is obtained by using

$$\bar{\nu}_{n,b}^{*MT}(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \omega_{i,b} \frac{\beta_u(\hat{U}_i) \left\{ Z_i - \hat{p}(\hat{U}_i) \right\} \gamma_y^{\mathcal{K}}(\hat{Y}_i)}{\sqrt{\hat{p}(\hat{U}_i) - \hat{p}(\hat{U}_i)^2}}, \ b = 1, \cdots, B,$$

and taking the empirical critical values of the bootstrap test statistics $\bar{T}_{n,b}^{MT*} = \Gamma \ \bar{\nu}_{n,b}^{*MT}$, $b = 1, \cdots, B$.

7 Simulation Studies

In this section, we present and discuss the results from simulation studies. We consider testing conditional independence of Y_i and Z_i given X_i, where each variable is set to be real-valued. The variable X_i is drawn from the uniform distribution on $[-1, 1]$ and the errors η_i and ε_i are independently drawn from $N(0, 1)$. The variable Z_i is binary taking zero or one by the following rule:

$$Z_i = 1\{X_i + \eta_i > 0\}$$

and the variable Y_i is determined by

$$\begin{aligned} \text{DGP 1}: &\quad Y_i = 0.3X_i + \kappa Z_i + \varepsilon_i, \text{ or} \\ \text{DGP 2}: &\quad Y_i = \sin(0.6X_i + \kappa Z_i) \times (\varepsilon_i - 0.2). \end{aligned}$$

The DGP 1 represents a linear specification with additive errors and DGP 2 represents a nonlinear specification with multiplicative errors. Note that $\kappa = 0$ corresponds to the null hypothesis of conditional independence between Y and Z given X.

Under these specifications, we consider the following two kinds of null hypotheses:

$$\begin{aligned} &\text{Null Hypothesis 1:} \qquad Y \perp Z \mid X \text{ and} \\ &\text{Null Hypothesis 2:} \qquad 1\{\tilde{Y} \leq \tau\} \perp Z \mid X, \end{aligned}$$

where $\tilde{Y}$ denotes the Rosenblatt transform of Y with respect to U, the quantile transform of X. As we shall see in the next section, Null Hypothesis 2 can be viewed as an identifying assumption for quantile treatment effects in program evaluations.

We consider two cases of $\beta_u(U) = \exp(Uu)$ and $\beta_u(U) = 1\{U \leq u\}$. Nonparametric estimations are done using kernel estimation using the kernel $K(u) = (15/16)(1-u^2)^2 1\{|u| \leq 1\}$. The number of Monte Carlo iterations and the number of bootstrap Monte Carlo iterations are set to be 1000. The sample sizes are 100, 200, and 500. As for taking the

Kolmogorov-Smirnov functional, we use the product of two sets of ten equally spaced grid points, $\{0, 0.1, 0.2, \cdots, 0.9\}^2$. The nominal size is set to be 0.05.

In choosing bandwidths, we perform cross-validations in two steps. We consider the following cross-validation criterion for the bandwidth of $\hat{p}$:

$$CV_Z(h) = \sum_{i=1}^{n} \{Z_i - \hat{p}_{h,i}(\hat{U}_i)\}^2$$

where $\hat{p}_{h,i}(u)$ is a leave-one-out kernel estimator of $p(u)$ with bandwidth h. We choose h^* such that minimizes $CV_Z(h)$ and define $\hat{p}_i(\hat{U}_i) = \hat{p}_{h^*,i}(\hat{U}_i)$. As for the bandwidth for $\hat{Y}_i$, we consider the following two cross-validation criteria:

$$\begin{aligned} \text{(CV-KS) } CV_Y^{KS}(h) &= \sup_{y \in \mathbf{R}} \left| \sum_{i=1}^{n} \frac{\{Z_i - \hat{p}_i(\hat{U}_i)\}^2 \{1\{Y_i \leq y\} - \hat{F}_{Y|U,i,h}(y|\hat{U}_i)\}^2}{\hat{p}_i(\hat{U}_i) - \hat{p}_i^2(\hat{U}_i)} \right| \text{ and} \\ \text{(CV-CM) } CV_Y^{CM}(h) &= \int \left| \sum_{i=1}^{n} \frac{\{Z_i - \hat{p}_i(\hat{U}_i)\}^2 \{1\{Y_i \leq y\} - \hat{F}_{Y|U,i,h}(y|\hat{U}_i)\}^2}{\hat{p}_i(\hat{U}_i) - \hat{p}_i^2(\hat{U}_i)} \right|^2 dy, \end{aligned}$$

where $\hat{F}_{Y|U,i,h}(y|\hat{U}_i)$ is the estimator defined in (10) at the bandwidth h. Note that the weighting reflects our normalization of the process to obtain asymptotically pivotal tests. Then we choose h^* that minimizes $CV_Y^{KS}(h)$ or $CV_Y^{CM}(h)$. As in Stute and Zhu (2005), we choose $h_1 = h^* \times n^{1/5 - 2/9}$ as the bandwidth for $\hat{F}_{Y|U,i,h}(y|\hat{U}_i)$ to fulfill the undersmoothing. As the wild bootstrap method does not require nonparametric estimation for each set of bootstrap samples, it suffices that we perform the cross-validations once and for all for the original data set.

[INCLUDE TABLES 1-4 HERE]

Tables 1 and 2 contain the empirical size and power of the tests using $\beta_u(U) = \exp(Uu)$ and $\beta_u(U) = 1\{U \leq u\}$ under DGP 1 and DGP 2. Both tests show good size and power properties. The empirical powers increase conspicuously as the sample size increases. Tables 3 and 4 present the results from Null Hypothesis 2 under DGP 1 and DGP 2. Comparing the results from CV-KS and those from CV-CM, it appears that both methods of cross-validation performs well in terms of size and power. Overall, the proposed methods of test and cross-validation appears to work well.

8 An Empirical Application : Evaluating the Evaluator using the JTPA Training Data

In the literature of program evaluations, the availability of experimental data has proved to be crucial in providing answers to questions that are often impossible to address using only nonexperimental data. Since the seminal paper by LaLonde (1986), such experimental data have been often used to evaluate the existing econometric methods that are applied to nonexperimental data. See e.g. Dehejia and Wahba (1999) and Smith and Todd (2002, 2005). In particular, Heckman, Ichimura, and Todd (1997) and HIST performed an extensive analysis of econometric models using the JTPA training data set.[11] In the empirical study we use the same data set and propensity scores from the JTPA training program as used by HIST and test certain implications from the identifying assumptions for nonexperimental procedures.

In this section, we assume the set-up of Section 2.2.1. The first test we focus on is the following null hypothesis:

$$H_0^{CI}\colon\ Y_0 \perp Z \mid P(X).$$

This is one implication from the unconfoundedness condition in (4). In a similar vein, HIST performed testing conditional mean independence of Y_0 and Z given $P(X)$.

The second test we focus on is an identifying assumption for quantile treatment effects:

$$H_0^{QNT}\colon\ 1\left\{Y_0 \leq q_{0\tau}(U)\right\} \perp Z\ |U,$$

where $U = F_P(X)$ the quantile transform of $P(X)$, with $F_P(x) = P\{P(X) \leq P(x)\}$.

Let $\hat{P}(X)$ be an estimator of $P(X)$ after a certain parametric specification of $P(X)$. This estimator $\hat{P}(X)$ corresponds to $\lambda_{\hat{\theta}}(X)$ in our notations in the previous sections. We take the empirical quantile transform of $\hat{P}(X_i)$ to obtain $\hat{U}_i = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} 1\{\hat{P}(X_j) \leq \hat{P}(X_i)\}$. Using this $\hat{U}_i$, we obtain the estimator $\tilde{Y}_i$ of the Rosenblatt-transformed outcomes $\tilde{Y}_i$ and the conditional distribution function $\hat{p}(\hat{U}_i)$ of $p(U_i) = P\{Z = 1|U_i\}$. We take the test statistic to be $\bar{T}_n = \Gamma\bar{\nu}_n$ where $\bar{\nu}_n$ is the process defined in (17) and Γ is the Kolmogorov-Smirnov functional.

We use the same data set used for Tables VIA and VIB in HIST. In particular, we employ the same estimates for the propensity score as obtained by HIST. Unlike their conditional mean independence tests, the tests constructed along the formulation of unconditional moment restrictions are not in general asymptotically invariant to the estimation error from the propensity scores. So the adjustment for estimation error of propensity score in this

[11] See Orr, et. al. (1995) for the description of the data.

case is not merely a matter of second order correction as is the case in HIST, but a matter of first order asymptotics. However, the tests proposed by this paper are designed to be asymptotically pivotal regardless of the estimation error in the propensity score estimation as long as the estimation error is stochastically bounded. Hence the testing procedures in this paper are asymptotically valid regardless of the estimation of the propensity score.

[INCLUDE TABLES 5-8 HERE]

The results from testing Null Hypothesis 1 are given in Tables 5-6 which contain p-values from both the tests using $\beta_u(U) = \exp(Uu)$ and $\beta_u(U) = 1\{U \leq u\}$. In both the cases of testing for each quarter or testing jointly, the tests do not reject the null hypothesis of conditional independence restriction at the significance level 5% for most quantiles and quarters. Tables 7-8 contain results from testing Null Hypothesis 2.[12] Across different quantiles τ, the test does not reject the null of conditional independence either, except for a few quarter-quantile pairs at quantiles around 0.6-0.7 in the case of $\beta_u(U) = 1\{U \leq u\}$.

Our findings appear to contradict one conclusion from HIST that the underlying assumptions of matching estimator, that is the assumption of unconfoundedness, do not seem to be supported by the JTPA data. HIST obtained mixed results from testing conditional mean independence depending on the use of adjustment for estimation error in the propensity score, and looked for evidence alternatively from pointwise bias estimates. They concluded that the JTPA data did not support the assumption of conditional mean independence because the bias estimates at various values for the propensity score were found to be significantly different from zero at low values of the propensity score.

To look closely into this apparent conflict, note that the bias considered by HIST is defined as

$$B(p) = \mathbf{E}[Y_0|P(X) = p, Z = 1] - \mathbf{E}[Y_0|P(X) = p, Z = 0].$$

Under H_0^{CI}, $B(p) = 0$ for all values of p. But HIST find that for low values of p, estimates of $\hat{B}(p)$ are significantly different from zero, and conclude that the data do not support H_0^{CI}. We can translate this conflict into that of testing conditional independence. The argument of HIST is analogous to testing the null hypothesis

$$H_{0,p}^{Point} : Y_0 \perp Z \; |P(X) = p$$

individually for each p and reject the stronger hypothesis, H_0^{CI}, based on the rejection of $H_{0,p}^{Point}$ at certain values of p. However, in general, when we reject the null hypothesis pointwise, this does not support the rejection of the joint hypothesis by itself. In order to reject a

[12] Here we report only the results with CV-KS. We have obtained similar results using CV-CM.

stronger null hypothesis, we need stronger evidence.[13] In our context, since H_0^{CI} is stronger than $H_{0,p}^{Point}$, in order to reject H_0^{CI}, we need evidence that is stronger than is just enough to reject $H_{0,p}^{Point}$. Our findings suggest that such strong enough evidence against H_0^{CI} is not yet discovered.[14]

9 Conclusion

A class of asymptotically pivotal tests of conditional independence have been proposed and investigated. This paper demonstrate that one can obtain asymptotically pivotal tests by using the Rosenblatt-transform in an appropriate manner. As it is often difficult to compute asymptotic critical values even when the tests are asymptotically pivotal, we propose bootstrap tests. The finite sample performances of the tests are affirmed by Monte Carlo simulation studies. In an empirical application, using the JTPA training data set and the propensity score specifications of HIST we test identifying assumptions for quantile treatment effects. From our testing results, we find no evidence against such identifying assumptions.

10 Appendix: Mathematical Proofs

10.1 Main Results

In this section, we present the proofs of the main results. Throughout the proofs, the notation C denotes a positive absolute constant that may assume different values in different contexts. For a class $\mathcal{F}$ of measurable functions, $N(\varepsilon, \mathcal{F}, L_r(Q))$ and $N_{[]}(\varepsilon, \mathcal{F}, L_r(Q))$ denote the covering and bracketing numbers of $\mathcal{F}$ with respect to the $L_r(Q)$-norm. (See van der Vaart and Wellner (1996) for their definitions.) Similarly we define $N(\varepsilon, \mathcal{F}, ||\cdot||_\infty)$ and $N_{[]}(\varepsilon, \mathcal{F}, ||\cdot||_\infty)$ to be the covering and bracketing numbers with respect to $||\cdot||_\infty$. We present two preliminary lemmas which are useful for many purposes.

Lemma A1 : *Let Λ_n be a sequence of classes of measurable functions such that for each $\lambda \in \Lambda_n$, $\lambda(X)$ is absolutely continuous with respect to the Lebesgue measure and has a uniformly bounded density. Let $\mathcal{T}$ be a class of uniformly bounded functions of bounded variation. Then for the class $\mathcal{G}_n = \{\tau \circ \lambda : (\tau, \lambda) \in \mathcal{T} \times \Lambda_n\}$ of measurable functions, it is satisfied that for any $r \geq 1$, and any probability measure Q,*

$$\log N_{[]}(C_2\varepsilon, \mathcal{G}_n, L_r(Q)) \leq \log N(\varepsilon^r, \Lambda_n, ||\cdot||_\infty) + C_1/\varepsilon,$$

where C_1 and C_2 are positive constants depending only on r.

[13]This ssue is closely related to the classical issue of multiple tests. In the context of multiple hypothesis testing, the p-values from the individual tests should be properly adjusted. See Lehman and Romano (2005) and references therein.

[14]In terms of bias estimates, to compare the bias estimate and the result of conditional mean independence, one must view $B(p)$ as a bias *function* of p and consider a confidence *band* for its nonparametric estimator uniformly over p, instead of pointwise confidence intervals.

Proof of Lemma A1 : Let F_λ be the distribution function of $\lambda(X)$. Then, observe that for any $\lambda_1, \lambda_2 \in \Lambda$,

$$\begin{aligned} \sup_x |F_{\lambda_1}(\lambda_1(x)) - F_{\lambda_2}(\lambda_2(x))| &\leq \sup_x |F_{\lambda_1}(\lambda_1(x) + \Delta_\infty) - F_{\lambda_1}(\lambda_1(x) - \Delta_\infty)| \\ &\leq C\Delta_\infty, \text{ where } \Delta_\infty = ||\lambda_1 - \lambda_2||_\infty \end{aligned}$$

because the density of $\lambda(X)$ is uniformly bounded. From now on, we identify $\lambda(x)$ with its quantile transform $F_\lambda(\lambda(x))$ without loss of generality so that $\lambda(X)$ is uniformly distributed on $[0,1]$. Since a function of bounded variation can be written as the difference of two monotone functions, we lose no generality by assuming that each $\tau \in \mathcal{T}$ is decreasing. Hence by the result of Birman and Solomjak (1967) (see also van der Vaart (1996) for a more general result), for any $r \geq 1$

$$\log N_{[]}(\varepsilon, \mathcal{T}, L_r(Q)) \leq \frac{C_2}{\varepsilon}, \tag{19}$$

for any probability measure Q and for a constant $C_2 > 0$. Here the constant C_2 does not depend on Q.

We choose $\{\lambda_1, \cdots, \lambda_{N_1}\}$ such that for any $\lambda \in \Lambda_n$, there exists $j \in \{1, \cdots, N_1\}$ with $\|\lambda_j - \lambda\|_\infty < \varepsilon^r/2$. Take a positive integer $M_\varepsilon \leq 2/\varepsilon^r + 1$ and choose a set $\{c_1, \cdots, c_{M_\varepsilon}\}$ such that $c_1 = 0$ and

$$c_{m+1} = c_m + \varepsilon^r/2, \; m = 1, \cdots, M_\varepsilon - 1$$

so that $0 = c_1 \leq c_2 \leq \cdots \leq c_{M_\varepsilon} \leq 1$. Define $\tilde{\lambda}_j(x)$ as follows.

$$\tilde{\lambda}_j(x) = c_m \text{ when } \lambda_j(x) \in [c_m, c_{m+1}), \text{ for some } m \in \{1, 2, \cdots, M_\varepsilon - 1\}.$$

For each $j_1 \in \{1, \cdots, N_1\}$, let Q_{j_1} be the distribution of $\tilde{\lambda}_{j_1}(X)$ under Q. Then we choose $\{(\tau_k, \Delta_k)\}_{k=1}^{N_2(j_1)}$ such that for any $\tau \in \mathcal{T}$, there exists $(\tau_{j_2}, \Delta_{j_2})$, $j_2 \in \{1, \cdots, N_2(j_1)\}$ such that $|\tau(\bar{\lambda}) - \tau_{j_2}(\bar{\lambda})| \leq \Delta_{j_2}(\bar{\lambda})$ with $\int \Delta_{j_2}(\bar{\lambda})^r Q_{j_1}(d\bar{\lambda}) < \varepsilon^r$. Now, take any $g \in \mathcal{T}$ such that $g \triangleq \tau \circ \lambda$ and choose λ_{j_1} such that $\|\lambda_{j_1} - \lambda\|_\infty < \varepsilon^r/2$ and, $(\tau_{j_2}, \Delta_{j_2})$ such that $|\tau(\bar{\lambda}) - \tau_{j_2}(\bar{\lambda})| \leq \Delta_{j_2}(\bar{\lambda})$ with $\int \Delta_{j_2}(\bar{\lambda})^r Q_{j_1}(d\bar{\lambda}) < \varepsilon^r$. We fix these j_1 and j_2. We extend the definition of Δ_{j_2} with its domain now equal to $\mathbf{R}$ by setting $\Delta_{j_2}(\bar{\lambda}) = 0$ for all $\bar{\lambda} \in \mathbf{R} \backslash [0,1]$.

Consider

$$\begin{aligned} |g(x) - (\tau_{j_2} \circ \tilde{\lambda}_{j_1})(x)| &\leq |(\tau \circ \lambda)(x) - (\tau \circ \tilde{\lambda}_{j_1})(x)| + |(\tau \circ \tilde{\lambda}_{j_1})(x) - (\tau_{j_2} \circ \tilde{\lambda}_{j_1})(x)| \\ &\leq |(\tau \circ \lambda)(x) - (\tau \circ \tilde{\lambda}_{j_1})(x)| + (\Delta_{j_2} \circ \tilde{\lambda}_{j_1})(x). \end{aligned} \tag{20}$$

The range of $\tilde{\lambda}_{j_1}$ is finite and $||\lambda - \tilde{\lambda}_{j_1}||_\infty \leq ||\lambda - \lambda_{j_1}||_\infty + ||\lambda_{j_1} - \tilde{\lambda}_{j_1}||_\infty \leq \varepsilon^r/2 + \varepsilon^r/2 = \varepsilon^r$. Hence $|(\tau \circ \lambda)(x) - (\tau \circ \tilde{\lambda}_{j_1})(x)|$ is bounded by $\tau(\tilde{\lambda}_{j_1}(x) - \varepsilon^r) - \tau(\tilde{\lambda}_{j_1}(x) + \varepsilon^r)$, or by

$$\tau_{j_2}(\tilde{\lambda}_{j_1}(x) - \varepsilon^r) - \tau_{j_2}(\tilde{\lambda}_{j_1}(x) + \varepsilon^r) + \Delta_{j_2}(\tilde{\lambda}_{j_1}(x) - \varepsilon^r) + \Delta_{j_2}(\tilde{\lambda}_{j_1}(x) + \varepsilon^r).$$

We analyze the difference $\tau_{j_2}(\tilde{\lambda}_{j_1}(x) - \varepsilon^r) - \tau_{j_2}(\tilde{\lambda}_{j_1}(x) + \varepsilon^r)$ which is written as $A_1(x) + A_2(x) + A_3(x) + A_4(x)$ where

$$\begin{aligned} A_1(x) &= \tau_{j_2}(\tilde{\lambda}_{j_1}(x) - \varepsilon^r) - \tau_{j_2}(\tilde{\lambda}_{j_1}(x) - \varepsilon^r/2), \\ A_2(x) &= \tau_{j_2}(\tilde{\lambda}_{j_1}(x) - \varepsilon^r/2) - \tau_{j_2}(\tilde{\lambda}_{j_1}(x)) \\ A_3(x) &= \tau_{j_2}(\tilde{\lambda}_{j_1}(x)) - \tau_{j_2}(\tilde{\lambda}_{j_1}(x) + \varepsilon^r/2) \text{ and} \\ A_4(x) &= \tau_{j_2}(\tilde{\lambda}_{j_1}(x) + \varepsilon^r/2) - \tau_{j_2}(\tilde{\lambda}_{j_1}(x) + \varepsilon^r). \end{aligned}$$

We write $A_1(x)$ as

$$\sum_{m=1}^{M_\varepsilon - 1} \left\{\tau_m^U(j_2) - \tau_m^L(j_2)\right\} \times 1\left\{c_m \le \lambda_{j_2}(x) < c_{m+1}\right\},$$

where $\tau_m^U(j_2) = \tau_{j_2}(c_m - \varepsilon^r)$ and $\tau_m^L(j_2) = \tau_{j_2}(c_m - \varepsilon^r/2)$. Since τ_{j_2} is decreasing, we have $\tau_m^L(j_2) \le \tau_m^U(j_2)$, and since $c_{m+1} = c_m + \varepsilon^r/2$, we have

$$\tau_{m+1}^U(j_2) = \tau_{j_2}(c_{m+1} - \varepsilon^r) = \tau_{j_2}(c_m - \varepsilon^r/2) = \tau_m^L(j_2),\ m = 1, \cdots, M_\varepsilon - 1.$$

Hence, we conclude

$$\tau_{M_\varepsilon - 1}^L(j_2) \le \cdots \le \tau_{m+1}^U(j_2) = \tau_m^L(j_2) \le \tau_m^U(j_2) = \tau_{m-1}^L(j_2) \le \cdots \le \tau_1^U(j_2).$$

Suppose that $\tau_1^U(j_2) = \tau_{M_\varepsilon-1}^L(j_2)$. Then $A_1(x) = 0$ and hence the $L_r(Q)$-norm of A_1 is trivially zero. Suppose that $\tau_1^U(j_2) > \tau_{M_\varepsilon-1}^L(j_2)$. Note that since τ is uniformly bounded, we have $\tau_1^U(j_2) - \tau_{M_\varepsilon-1}^L(j_2) < C < \infty$ for an absolute constant $C > 0$. Define

$$\tilde{\Delta}_{j_1,j_2}(x) \triangleq \sum_{m=1}^{M_\varepsilon - 1} \frac{\tau_m^L(j_2) - \tau_m^U(j_2)}{\tau_1^L(j_2) - \tau_{M_\varepsilon-1}^U(j_2)} \times 1\left\{c_m \le \lambda_{j_2}(x) < c_{m+1}\right\}.$$

Let $p_m(j_2) = P\left\{c_m \le \lambda_{j_2}(x) < c_{m+1}\right\}$. Since $\tilde{\Delta}_{j_1,j_2}(x) \le 1$, we have $\tilde{\Delta}_{j_1,j_2}^r(x) \le \tilde{\Delta}_{j_1,j_2}(x)$ so that

$$\mathbf{E}\tilde{\Delta}_{j_1,j_2}^r(X) \le \sum_{m=1}^{M_\varepsilon - 1} \frac{\tau_m^L(j_2) - \tau_m^U(j_2)}{\tau_1^L(j_2) - \tau_{M_\varepsilon-1}^U(j_2)} \times p_m(j_2) = \varepsilon^r/2,$$

because $p_m(j_2) = \varepsilon^r/2$ for $m \in \{1, \cdots, M_\varepsilon - 1\}$. Thus we conclude that the $L_r(Q)$-norm of A_1 is bounded by $C\varepsilon$. We can deal with the functions $A_j(x)$, $j = 2, 3, 4$, precisely in the same manner.

From (20), we can bound $|g(x) - (\tau_{j_2} \circ \tilde{\lambda}_{j_1})(x)|$ by

$$\begin{aligned} & A_1(x) + A_2(x) + A_3(x) + A_4(x) + (\Delta_{j_2} \circ \tilde{\lambda}_{j_1})(x) + \Delta_{j_2}(\tilde{\lambda}_{j_1}(x) - \varepsilon^r) + \Delta_{j_2}(\tilde{\lambda}_{j_1}(x) + \varepsilon^r) \qquad (21) \\ = \ & \Delta_{j_1,j_2}^*(x), \text{ say.} \end{aligned}$$

Now, let us compute $[\mathbf{E}\{\Delta_{j_1,j_2}^*(X)\}^r]^{1/r}$. The $L_r(Q)$-norm of the first four functions is bounded by $C\varepsilon$ as we proved before. By the choice of Δ_{j_2}, $\mathbf{E}[\Delta_{j_2}^r(\tilde{\lambda}_{j_1}(X))] = \int \Delta_j(\bar{\lambda})^r Q_{j_2}(d\bar{\lambda}) < \varepsilon^r$. Let us turn to the last two terms in (21). Note that

$$\begin{aligned} \mathbf{E}\left[\Delta_{j_2}^r(\tilde{\lambda}_{j_1}(X) - \varepsilon^r)\right] &= \sum_{m=1}^{M_\varepsilon - 1} \Delta_{j_2}^r(c_m - \varepsilon^r) p_m(j_2) = \sum_{m=1}^{M_\varepsilon - 1} \Delta_{j_2}^r(c_m - \varepsilon^r)\varepsilon^r/2 \\ &= \sum_{m=1}^{M_\varepsilon - 3} \Delta_{j_2}^r(c_m)\varepsilon^r/2 \le \mathbf{E}[\Delta_{j_2}^r(\tilde{\lambda}_{j_1}(X))] < \varepsilon^r. \end{aligned}$$

The third equality is due to our setting $\Delta_{j_2}(c) = 0$ for $c \in \mathbf{R}\backslash[0,1]$. Similarly, $\mathbf{E}[\Delta_{j_2}^r(\tilde{\lambda}_{j_1}(X) - \varepsilon^r)] < \varepsilon^r$. Combining these results, $\mathbf{E}[\{\Delta_{j_1,j_2}^*(X)\}^r] \le C_1^r \varepsilon^r$, for some constant $C_1 > 0$, yielding the result that

$$\log N_{[]}(C_1\varepsilon, \mathcal{G}_n, L_r(Q)) \le \log N(\varepsilon^r, \Lambda_n, ||\cdot||_\infty) + C_2/\varepsilon.$$

■

Let $F_{n,\lambda}(\bar{\lambda}) = \frac{1}{n}\sum_{i=1}^{n} 1\left\{\lambda(X_i) \leq \bar{\lambda}\right\},\ \lambda \in \Lambda_n$. Then Lemma A1 immediately yields a bracketing entropy bound for a class of functions in which $\beta_u\left(F_{n,\lambda}(\lambda(\cdot))\right)$ realizes. In the case of an indicator function $\beta_u(\bar{u}) = 1\{u \leq \bar{u}\}$, we cannot apply the framework of Chen, Linton and van Keilegom (2003) because the nonparametric function $F_{n,\lambda}(\bar{\lambda})$ is a step function and hence L_p uniform continuity with respect to the sup norm fails.

Corollary A1 : *Let* Λ_n *be a class of functions as in* Lemma A1 *and let* $\{\beta_u(\cdot)\}_{u\in[0,1]}$ *be a class of functions satisfying* Assumption 3*. Then, there exists a class of functions* $\mathcal{G}_n$ *such that*

$$P\left\{\beta_u\left(F_{n,\lambda}(\lambda(\cdot))\right) \in \mathcal{G}_n,\ \text{for all } (\lambda, u) \in \Lambda_n \times [0,1]\right\} = 1$$

and for any $r \geq 1$*, and any probability measure* Q*,* $\log N_{[]}(C_2\varepsilon, \mathcal{G}_n, L_r(Q)) \leq \log N(\varepsilon^r, \Lambda_n, ||\cdot||_\infty) + C_1/\varepsilon$*, where* C_1 *and* C_2 *are positive constants depending only on* r*.*

The following lemma is useful to establish the bracketing entropy bound of a function space in which conditional distribution function estimators realize in. See the proof of Lemma A3. Note that the framework of Andrews (1994) does not directly apply when the realized functions are discontinuous. The lemma can be used conveniently in particular when the conditional distribution function estimators contain unknown finite dimensional or infinite dimensional parameters in the conditioning variable.

Lemma A2 : *We introduce three classes of functions. First, let* $\mathcal{F}_n$ *be a sequence of classes of uniformly bounded functions* $\phi(\cdot,\cdot) : \mathbf{R} \times \mathcal{S} \to [0,1]$*, such that for each* $v \in \mathcal{S}$*,* $\phi(\cdot, v)$ *is monotone, and for each* $\varepsilon > 0$*,*

$$\sup_{(y,v)\in\mathbf{R}\times\mathcal{S}}\sup_{\eta\in[0,\varepsilon]} |\phi(y, v+\eta) - \phi(y, v-\eta)| < M_n\varepsilon \tag{22}$$

for some sequence $M_n > 1$*, where* $\mathcal{S}$ *is a totally bounded subset of the real line. Second, let* $\mathcal{G}$ *be a class of measurable functions* $G : \mathbf{R}^{d_X} \to \mathcal{S}$*. Lastly let* $\mathcal{J}_n^{\mathcal{G}} = \{\phi(\cdot, G(\cdot)) : (\phi, G) \in \mathcal{F}_n \times \mathcal{G}\}$*.*

Then for any probability measure P *and for any* $p > 1$*,*

$$\log N_{[]}(\varepsilon, \mathcal{J}_n^{\mathcal{G}}, L_p(P)) \leq CM_n^{2/(p-1)}/\varepsilon^{p/(p-1)} - C\log(\varepsilon) + \log N_{[]}(C\varepsilon/M_n, \mathcal{G}, L_p(P))$$

for some $C > 0$*.*

Proof of Lemma A2 : Fix $\varepsilon > 0$ and let $N_\varepsilon(v) = N(\varepsilon, \mathcal{F}_n(v), L_p(Q_v))$, where $\mathcal{F}_n(v) = \{\phi(\cdot, v) : \phi \in \mathcal{F}_n\}$ and Q_v denotes the conditional measure of Y given $V = v$. Take a partition $\mathcal{S} = \cup_{k=1}^{J_\varepsilon} B(b_k)$ where $B(b_k)$ is a set contained in an ε-interval centered at b_k and $J_\varepsilon \leq C\varepsilon^{-d_X}$. For each b_k, take $\{(f_{k,j}, \Delta_{k,j})\}_{j=1}^{N_\varepsilon(b_k)}$ such that for any $f \in \mathcal{F}_n(b_k)$, there exists $j \in \{1, \cdots, N_\varepsilon(b_k)\}$ such that $|f(y) - f_{k,j}(y)| \leq \Delta_{k,j}(y)$ and $\int |\Delta_{k,j}(y)|^p P(dy) \leq \varepsilon^p$. Given $\phi \in \mathcal{F}_n$, we let $\tilde{f}_j(y,v) = \sum_{k=1}^{J_\varepsilon - 1} f_{k,j}(y, b_k) 1\{v \in B_\varepsilon(b_k)\}$ where $f_{k,j}(y, b_k)$ is such that $|\phi(y, b_k) - f_{k,j}(y)| \leq \Delta_{k,j}(y)$ and $\int |\Delta_{k,j}(y)|^p P(dy) \leq \varepsilon^p$. By the result of Birman and Solomjak

(1967), the smallest number of such j's are bounded by $\exp(C/\varepsilon)/\varepsilon$. Then

$$\begin{aligned} \left|\phi(y,v)-\tilde{f}_j(y,v)\right| &= \left|\sum_{k=1}^{J_\varepsilon-1}\{\phi(y,v)-f_{k,j}(y,b_k)\}1\{v\in B_\varepsilon(b_k)\}\right| \\ &\leq \left|\sum_{k=1}^{J_\varepsilon-1}\{\phi(y,v)-\phi(y,b_k)\}1\{v\in B_\varepsilon(b_k)\}\right| \\ &\quad+\left|\sum_{k=1}^{J_\varepsilon-1}\{\phi(y,b_k)-f_{k,j}(y,b_k)\}1\{v\in B_\varepsilon(b_k)\}\right|. \end{aligned} \tag{23}$$

The last term is bounded by $\sum_{k=1}^{J_\varepsilon-1}\Delta_{k,j}(y,b_k)1\{v\in B_\varepsilon(b_k)\}$. Since

$$\{\phi(y,v)-\phi(y,b_k)\}1\{v\in B_\varepsilon(b_k)\}\leq CM_n\varepsilon 1\{v\in B_\varepsilon(b_k)\}$$

we can bound the second to the last term in (23) by $CM_n\varepsilon$. Hence, $\left|\phi(y,v)-\tilde{f}_j(y,v)\right|\leq\bar{\Delta}_j(y)$ where

$$\bar{\Delta}_j(y)=\sum_{k=1}^{J_\varepsilon-1}\Delta_{k,j}(y,b_k)1\{v\in B_\varepsilon(b_k)\}+CM_n\varepsilon.$$

Take $q=(1-1/p)^{-1}$. By Hölder inequality,

$$\left(\sum_{k=1}^{J_\varepsilon-1}\Delta_{k,j}(y)1\{v\in B_\varepsilon(b_k)\}\right)^p\leq\left\{\sum_{k=1}^{J_\varepsilon-1}\Delta_{k,j}^p(y,b_k)\right\}\left\{\sum_{k=1}^{J_\varepsilon-1}1\{v\in B_\varepsilon(b_k)\}\right\}^{p/q}=\sum_{k=1}^{J_\varepsilon-1}\Delta_{k,j}^p(y,b_k).$$

Now, note that $\int\bar{\Delta}_j^p(y)P(dy,dv)$ is bounded by

$$C\sum_{k=1}^{J_\varepsilon-1}\int\Delta_{k,j}^p(y,b_k))P(dy)+CM_n^p\varepsilon^p\leq CJ_\varepsilon\varepsilon^p+CM_n^p\varepsilon^p\leq C\varepsilon^{p-1}+CM_n^p\varepsilon^p\leq CM_n^p\varepsilon^{p-1},$$

yielding the inequality: $\log N_{[]}(CM_n\varepsilon^{(p-1)/p},\mathcal{F}_n,L_p(P))\leq C/\varepsilon-C\log(\varepsilon)$.

Now, take $(G_k,\Delta_k)_{k=1}^{N_1}$ such that for any $G\in\mathcal{G}$, there exists (G_j,Δ_j) such that $|G-G_j|\leq\Delta_j$ and $\mathbf{E}_P\Delta_j^p(X)<\varepsilon^p$. For each $j\in\{1,\cdot\cdot\cdot,N_1\}$, let Q_j be the distribution of $(Y,G_j(X))$ under P. Take $(\phi_k,\tilde{\Delta}_k)_{k=1}^{N_2(j)}$ such that for any $\phi\in\mathcal{J}_n$, there exists $(\phi_k,\tilde{\Delta}_k)$ such that $|\phi(y,v)-\phi_k(y,v)|<\tilde{\Delta}_k(y,v)$ and $\int\tilde{\Delta}_k^p(y,v)Q_j(dy,dv)<\varepsilon^p$. Now,

$$\begin{aligned} |\phi(y,G(x))-\phi_k(y,G_j(x))| &\leq |\phi(y,G(x))-\phi(y,G_j(x))|+|\phi(y,G_j(x))-\phi_k(y,G_j(x))| \\ &\leq M_n\Delta_j(x)+\tilde{\Delta}_k(y,G_j(x))=\bar{\Delta}_{j,k}(y,x),\text{ say.} \end{aligned}$$

Since $\mathbf{E}\bar{\Delta}_{j,k}^p(Y,X)\leq C(M_n+C)^p\varepsilon^p\leq CM_n^p\varepsilon^p$, we conclude that

$$\begin{aligned} \log N_{[]}\left(CM_n\varepsilon,\mathcal{J}_n^{\mathcal{G}},L_p(P)\right) &\leq C\log N_{[]}(\varepsilon,\mathcal{F}_n,L_p(P))+C\log N_{[]}(\varepsilon,\mathcal{G},L_p(P)) \\ &\leq C(M_n/\varepsilon)^{p/(p-1)}+C\log(M_n/\varepsilon)+C\log N_{[]}(\varepsilon,\mathcal{G},L_p(P)). \end{aligned}$$

By redefining $\varepsilon'=CM_n\varepsilon$, we obtain the wanted result. ■

Proof of Theorem 1 : Let $\mathcal{G}_n$ be a class of functions such that

$$P\left\{\beta_u(F_{n,\theta,i}(\lambda_\theta(\cdot))) \in \mathcal{G}_n \text{ for all } (u,\theta) \in [0,1] \times B(\theta_0,\delta)\right\} = 1$$

and satisfies that $\log N_{[]}(\varepsilon, \mathcal{G}_n, ||\cdot||_r) < C/\varepsilon$ and let $\Lambda_n \triangleq \{\lambda_\theta : \theta \in B(\theta_0, Cn^{-1/4})\}$. The existence of such $\mathcal{G}_n$ is guaranteed by Corollary A1 above, combined with the fact that $\log N(\varepsilon, \Lambda_n, ||\cdot||_r) < C \log \varepsilon$. This inequality follows from the assumption that λ_θ is Lipschitz continuous in θ and Θ is compact in $\mathbf{R}^{d_\Theta}$ (c.f. Assumption 2(i), (ii)(b)). Since $\frac{1}{nh}\sum_{j=1,j\neq i}^n K_h(\hat{U}_j - \hat{U}_i) > \eta$, for a small $\eta \in (0,1)$, with probability approaching one (Lemma B3), we confine our attention to such an event.

(i) Let ε be the constant in Assumption 5(ii) and ε_1 the constant in the condition for local alternatives. Choose $\delta_n = n^{-\varepsilon/2}$. Then, observe that $n^{\varepsilon_1}\delta_n \to 0$, $(n^{-1/2+\varepsilon/2}h^{-1} + n^{\varepsilon/2}h^2)\delta_n^{-1} \to 0$ and $\sqrt{n}(n^{-1/2}h^{-1} + h^2)^2\delta_n^{-1} = \left(n^{-1/2+\varepsilon/2}h^{-2} + n^{1/2+\varepsilon/2}h^4\right) \to 0$. Let $\gamma_{z,\delta_n}(v) = \Phi((z-v)/\delta_n)$ and $\gamma'_{z,\delta_n}(v) = -\phi((z-v)/\delta_n)/\delta_n$ where Φ and ϕ are a standard normal distribution function and its density function. Likewise we define $\gamma_{y,\delta_n}(\cdot)$ and $\gamma'_{y,\delta_n}(\cdot)$. The $o_P(1)$ terms in the following are uniform over $(r,\beta) \in [0,1]^3 \times \mathcal{G}_n$. Let $\gamma^\perp_{z,\delta_n}(\tilde{Z}_i) = \gamma_{z,\delta_n}(\tilde{Z}_i) - \mu_{\delta_n}(z)$ and $\gamma^\perp_{y,\delta_n}(\tilde{Y}_i) = \gamma_{y,\delta_n}(\tilde{Y}_i) - \mu_{\delta_n}(y)$, where $\mu_{\delta_n}(z) = \mathbf{E}[\gamma_{z,\delta_n}(\tilde{Z}_i)]$ and $\mu_{\delta_n}(y) = \mathbf{E}[\gamma_{y,\delta_n}(\tilde{Y}_i)]$. Similarly define $\gamma^\perp_{z,\delta_n}(\hat{Z}_i) = \gamma_{z,\delta_n}(\hat{Z}_i) - \mu_{\delta_n}(z)$ and $\gamma^\perp_{y,\delta_n}(\hat{Y}_i) = \gamma_{y,\delta_n}(\hat{Y}_i) - \mu_{\delta_n}(y)$.

We write $\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\gamma_z^\perp(\hat{Z}_i)\gamma_y^\perp(\hat{Y}_i)$ as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\left\{\gamma_z^\perp(\hat{Z}_i)\gamma_y^\perp(\hat{Y}_i) - \gamma^\perp_{z,\delta_n}(\hat{Z}_i)\gamma^\perp_{y,\delta_n}(\hat{Y}_i)\right\} + \frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\gamma^\perp_{z,\delta_n}(\hat{Z}_i)\gamma^\perp_{y,\delta_n}(\hat{Y}_i). \tag{24}$$

By Lemma A3 below, the first sum is asymptotically equivalent to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{\beta_u(U_i)\Delta_{z,y,n}(\tilde{Z}_i,\tilde{Y}_i) - \mathbf{E}\left[\beta_u(U_i)\Delta_{z,y,n}(\tilde{Z}_i,\tilde{Y}_i)\right]\right\} + \sqrt{n}\mathbf{E}\left[\beta_u(U_i)\Delta_{z,y,n}(\tilde{Z}_i,\tilde{Y}_i)\right]$$

where $\Delta_{z,y,n}(\tilde{Z}_i,\tilde{Y}_i) = \gamma_z^\perp(\tilde{Z}_i)\gamma_y^\perp(\tilde{Y}_i) - \gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i)$. We can easily show that the first sum is $o_P(1)$ by using the usual stochastic equicontinuity arguments. The last term is 0 under the null hypothesis and $o(1)$ under the local alternatives. Therefore, it suffices to establish the asymptotic representation in the theorem for the last sum in (24). The asymptotic representation follows immediately once we show the following three claims.

Claim 1 : $\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\{\gamma_{y,\delta_n}(\hat{Y}_i) - \gamma_{y,\delta_n}(\tilde{Y}_i)\}\mu_{\delta_n}(z) = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta_u(U_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i)\mu_{\delta_n}(z) + o_P(1)$.

Claim 2 :

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\gamma^\perp_{z,\delta_n}(\hat{Z}_i)\{\gamma_{y,\delta_n}(\hat{Y}_i) - \gamma_{y,\delta_n}(\tilde{Y}_i)\} &= o_P(1), \text{ and} \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\{\gamma_{y,\delta_n}(\hat{Y}_i) - \gamma_{y,\delta_n}(\tilde{Y}_i)\} &= o_P(1),\end{aligned}$$

Claim 3 : $\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta_{u,\delta_n}(U_{n,\theta,i})\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \beta_u(U_i)\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i) + o_P(1)$.

Similarly as in the proof of Claim 1, we can show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\{\gamma_{z,\delta_n}(\hat{Z}_i)-\gamma_{z,\delta_n}(\tilde{Z}_i)\}\mu_{\delta_n}(y)=-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_u(U_i)\gamma_{z,\delta_n}^{\perp}(\tilde{Z}_i)\mu_{\delta_n}(y)+o_P(1). \tag{25}$$

Let us see how the three claims lead to the wanted result of the theorem. First, write

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\gamma_{z,\delta_n}^{\perp}(\hat{Z}_i)\gamma_{y,\delta_n}^{\perp}(\hat{Y}_i) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\gamma_{z,\delta_n}^{\perp}(\hat{Z}_i)\left\{\gamma_{y,\delta_n}(\hat{Y}_i)-\gamma_{y,\delta_n}(\tilde{Y}_i)\right\}\\ +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\left\{\gamma_{z,\delta_n}(\hat{Z}_i)-\gamma_{y,\delta_n}(\tilde{Z}_i)\right\}\gamma_{y,\delta_n}^{\perp}(\tilde{Y}_i) &+ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\gamma_{z,\delta_n}^{\perp}(\tilde{Z}_i)\gamma_{y,\delta_n}^{\perp}(\tilde{Y}_i).\end{aligned}$$

By applying Claim 2 and (25), we reduce the above sums to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\gamma_{z,\delta_n}^{\perp}(\tilde{Z}_i)\gamma_{y,\delta_n}^{\perp}(\tilde{Y}_i)+o_P(1).$$

The proof is complete by Claim 3 and the fact that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_u(U_i)\gamma_{z,\delta_n}^{\perp}(\tilde{Z}_i)\gamma_{y,\delta_n}^{\perp}(\tilde{Y}_i)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_u(U_i)\gamma_{z}^{\perp}(\tilde{Z}_i)\gamma_{y}^{\perp}(\tilde{Y}_i)+o_P(1),$$

as can be shown using arguments similar to the proof of Lemma A3 below.

Proof of Claim 1: Let γ''_{z,δ_n} denote the second order derivative of γ_{z,δ_n}. Take M to be a positive integer greater than $1/(2\varepsilon)-1$ where ε is the constant in Assumption 5(ii). By Taylor expansion,

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\mu_{\delta_n}(z)\{\gamma_{y,\delta_n}(\hat{Y}_i)-\gamma_{y,\delta_n}(\tilde{Y}_i)\} \qquad (26)\\ =\;&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\mu_{\delta_n}(z)\gamma'_{y,\delta_n}(\tilde{Y}_i)\left\{\hat{Y}_i-\tilde{Y}_i\right\}+\sum_{m=2}^{M-1}\frac{1}{m!\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\mu_{\delta_n}(z)\gamma_{y,\delta_n}^{(m)}(\tilde{Y}_i)\left\{\hat{Y}_i-\tilde{Y}_i\right\}^m\\ &+\frac{1}{M!}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\mu_{\delta_n}(z)\gamma_{y,\delta_n}^{(M)}(\tilde{Y}_i^*)\left\{\hat{Y}_i-\tilde{Y}_i\right\}^M\end{aligned}$$

where $\tilde{Y}_i^*$ lies on the line segment between $\hat{Y}_i$ and $\tilde{Y}_i$. Note that $\max_{1\le i\le n}\{\hat{Z}_i-\tilde{Z}_i\}^m=O_P(n^{-m/2}h^{-m}+h^{2m})$ by Lemma B3 below. We analyze the second sum. Let $\mathcal{F}_{Y,n}$ be the class of functions that contain $\hat{F}_{Y|U}(\cdot|F_{n,\hat{\theta},i}(\cdot))$ with probability one and have uniform distance from $F_{Y|U}(\cdot|F_0(\cdot))$ bounded by $Cn^{-1/4}$. By Lemmas A1 and A2 above, we can take such classes to satisfy

$$\log N_{[]}(\varepsilon,\mathcal{F}_{Y,n},L_p(P))\le C(h^{-2}\varepsilon^{-1})^{p/(p-1)}.$$

As for the second sum in (26), we consider the following process

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\mu_{\delta_n}(z)\gamma_{y,\delta_n}^{(m)}(\tilde{Y}_i)\left\{F_Y(Y_i,X_i)-\tilde{Y}_i\right\}^{m-1}\left\{\hat{Y}_i-\tilde{Y}_i\right\}$$

indexed by $(\beta, F_Y) \in \tilde{\mathcal{G}}_n \times \mathcal{F}_{Y,n}$ and $(y,z,u) \in [0,1]^3$. By Lemma B1 below, the above sum is equal to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{E}\left[\beta(X_i)\mu_{\delta_n}(z)\gamma_{y,\delta_n}^{(m)}(\tilde{Y}_i)\left\{F_Y(Y_i,X_i)-\tilde{Y}_i\right\}^{m-1}\left\{\gamma_y(\tilde{Y}_i)-\tilde{Y}_i\right\}|U_i\right]_{y=\tilde{Y}_i}+o_P(1).$$

Using Fubini Theorem, we can check that the leading sum above is a mean zero process. Note that $\sup_{F_Y\in\mathcal{F}_{Y,n}}\sup_{(y,x)}|F_Y(y,x)-F_{Y|U}(y|F_0(x))| = O_P(n^{-1/4})$ by the construction of $\mathcal{F}_{Y,n}$. Hence the second sum in (26) is $o_P(1)$. The third sum in (26) is readily shown to be $O_P(\delta_n^{-(M+1)}n^{1/2}\{n^{-1/2}h^{-1}+h^2\}^M) = o_P(1)$ by the choice of δ_n and M.

Now, we turn to the first sum in (26). By Lemma B1, the sum is asymptotically equivalent to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_u(U_i)\mathbf{E}\left[\mu_{\delta_n}(z)\gamma'_{y,\delta_n}(\tilde{Y}_i)\left\{\gamma_y(\tilde{Y}_i)-\tilde{Y}_i\right\}|U_i\right]_{z=\tilde{Z}_i}.$$

The proof is complete by noting that

$$\mathbf{E}\left[\mu_{\delta_n}(z)\gamma'_{y,\delta_n}(\tilde{Y}_i)\left\{1\{y\le\tilde{Y}_i\}-\tilde{Y}_i\right\}|U_i\right]_{y=\tilde{Y}_i} = \mu_{\delta_n}(z)\int\gamma'_{y,\delta_n}(\bar{y})\left\{1\{\tilde{Y}_i\le\bar{y}\}-\bar{y}\right\}d\bar{y},$$

where by integration by parts, the last term is equal to $-\mu_{\delta_n}(z)\left\{\gamma_{y,\delta_n}(\tilde{Y}_i)-\mu_{\delta_n}(y)\right\}$. The equality uses the fact that $\tilde{Y}_i$ is independent of U_i.

Proof of Claim 2 : We deal with only the first statement. Note that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\gamma_{z,\delta_n}(\hat{Z}_i)\left\{\gamma_{y,\delta_n}(\hat{Y}_i)-\gamma_{y,\delta_n}(\tilde{Y}_i)\right\} \quad (27)\\
=\;&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\gamma_{z,\delta_n}(\tilde{Z}_i)\left\{\gamma_{y,\delta_n}(\hat{Y}_i)-\gamma_{y,\delta_n}(\tilde{Y}_i)\right\}\\
&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\left\{\gamma_{z,\delta_n}(\hat{Z}_i)-\gamma_{z,\delta_n}(\tilde{Z}_i)\right\}\left\{\gamma_{y,\delta_n}(\hat{Y}_i)-\gamma_{y,\delta_n}(\tilde{Y}_i)\right\}.
\end{aligned}$$

The first sum is equal to (by expanding the terms and following the same steps in the proof of Claim 1)

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)\gamma_{z,\delta_n}(\tilde{Z}_i)\gamma'_{y,\delta_n}(\tilde{Y}_i)\left\{\hat{Y}_i-\tilde{Y}_i\right\}+o_P(1).$$

By Lemma B1 and the stochastic equicontinuity arguments as in the proof of Claim 1, the sum above is asymptotically equivalent to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{E}\left[c_n(S_i;r)\left\{1\{y'\le Y_i\}-\tilde{Y}_i\right\}|U_i\right]_{y'=Y_i}, \quad (28)$$

where $c_n(S_i;r) = \beta_u(U_i)\gamma_{z,\delta_n}(\tilde{Z}_i)\gamma'_{y,\delta_n}(\tilde{Y}_i)$. Now, write $\mathbf{E}\left[c_n(S_i;r)\left\{1\{y' \leq Y_i\} - \tilde{Y}_i\right\}|U_i\right]_{y'=Y_i}$ as

$$\begin{aligned}
&\beta_u(U_i)\mu_{\delta_n}(z)\mathbf{E}\left[\gamma'_{y,\delta_n}(\tilde{Y}_i)\left\{1\{y' \leq \tilde{Y}_i\} - \tilde{Y}_i\right\}|U_i\right]_{y'=\tilde{Y}_i} \\
&+\beta_u(U_i)\mathbf{E}\left[\{\mathbf{E}\left[\gamma_{z,\delta_n}(\tilde{Z}_i)|U_i,\tilde{Y}_i\right] - \mu_{\delta_n}(z)\}\gamma'_{y,\delta_n}(\tilde{Y}_i)\left\{1\{y' \leq \tilde{Y}_i\} - \tilde{Y}_i\right\}|U_i\right]_{y'=\tilde{Y}_i}.
\end{aligned} \tag{29}$$

Because $\tilde{Y}_i$ is independent of U_i, the expectation in the first term is written as

$$\int_{\tilde{Y}_i}^1 \gamma'_{y,\delta_n}(u)du - \int_0^1 \gamma'_{y,\delta_n}(u)udu = -\{\gamma_{y,\delta_n}(\tilde{Y}_i) - \mu_{\delta_n}(y)\}.$$

We turn to the last term in (29). Let $\mu_{\delta_n}(U_i,z|y) = \mathbf{E}\left[\gamma_{z,\delta_n}(\tilde{Z}_i)|U_i,\tilde{Y}_i = y\right]$ and write

$$\begin{aligned}
&\mathbf{E}\left[\{\mu_{\delta_n}(U_i,z|y) - \mu_{\delta_n}(z)\}\gamma'_{y,\delta_n}(\tilde{Y}_i)\left\{1\{y' \leq \tilde{Y}_i\} - \tilde{Y}_i\right\}|U_i\right]_{y'=\tilde{Y}_i} \\
=& \int_0^1 \int_{-y/\delta_n}^{y/\delta_n} \{\mu_{\delta_n}(U_i,z|y) - \mu_{\delta_n}(z)\}\phi(\bar{y})\left\{1\{\tilde{Y}_i \leq y + \delta_n\bar{y}\} - (y+\delta_n\bar{y})\right\} f(\bar{z}, y+\delta_n\bar{y}|U_i)d\bar{z}d\bar{y} \\
=& \int_0^1 \int_{-\infty}^{\infty} \{\mu_{\delta_n}(U_i,z|y) - \mu_{\delta_n}(z)\}\left\{1\{\tilde{Y}_i \leq y + \delta_n\bar{y}\} - (y+\delta_n\bar{y})\right\} f(\bar{z}, y|U_i)d\bar{z}d\bar{y} + O_P(n^{-1/2+\varepsilon_1}\delta_n),
\end{aligned}$$

where $f(\bar{z},\bar{y}|U_i)$ denotes the conditional density of $(\tilde{Z}_i,\tilde{Y}_i)$ given U_i. The last equality can be seen to follow easily by the continuity of the joint distribution of $\tilde{Y}_i$ and $\tilde{Z}_i$ (see Assumption 1). The term $O_P(n^{-1/2+\varepsilon_1}\delta_n)$ is due to the chosen local alternatives. Note that $f(\bar{z},y|U_i) = f(\bar{z}|y,U_i)$ so that the above is equal to

$$\begin{aligned}
&\int_{-\infty}^{\infty} \mathbf{E}\left[\{\mu_{\delta_n}(U_i,z|y) - \mu_{\delta_n}(z)\}\left\{1\{y' \leq \tilde{Y}_i + \delta_n\bar{y}\} - (\tilde{Y}_i + \delta_n\bar{y})\right\}|U_i, \tilde{Y}_i = y\right]_{y'=\tilde{Y}_i} d\bar{y} \\
=& \int_{-\infty}^{\infty} \left\{1\{\tilde{Y}_i \leq y + \delta_n\bar{y}\} - (y+\delta_n\bar{y})\right\}\{\mu_{\delta_n}(U_i,z|y) - \mu_{\delta_n}(z)\}d\bar{y}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbf{E}\left[\{\mu_{\delta_n}(U_i,z|y) - \mu_{\delta_n}(z)\}\gamma'_{y,\delta_n}(\tilde{Y}_i)\left\{1\{y' \leq \tilde{Y}_i\} - \tilde{Y}_i\right\}|U_i\right]_{y'=\tilde{Y}_i} \\
=& \int_{-\infty}^{\infty} \frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{1\{\tilde{Y}_i \leq y + \delta_n\bar{y}\} - (y+\delta_n\bar{y})\right\}\{\mu_{\delta_n}(U_i,z|y) - \mu_{\delta_n}(z)\}d\bar{y} + O_P(n^{\varepsilon_1}\delta_n).
\end{aligned}$$

Note that leading sum is mean zero, and under the null and local alternatives it is $o_P(1)$. Therefore, the first sum in (27) becomes $-\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta_u(U_i)\mu_{\delta_n}(z)\left\{\gamma_{y,\delta_n}(\tilde{Y}_i) - \mu_{\delta_n}(y)\right\} + o_P(1)$.

Let us turn to the second term in (27). Expanding the term similarly as before, we observe that the term is asymptotically equivalent to

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\gamma'_{z,\delta_n}(\tilde{Z}_i)\gamma'_{y,\delta_n}(\tilde{Y}_i)\left\{\hat{Z}_i - \tilde{Z}_i\right\}\left\{\hat{Y}_i - \tilde{Y}_i\right\}1_{i,\delta_n}(r).$$

By Lemma B3 below and Assumption 5(ii), we find that $\sup_{1\leq i\leq n}|\hat{Z}_i - \tilde{Z}_i| = o_P(n^{-1/4})$ and $\sup_{1\leq i\leq n}|\hat{Y}_i - \tilde{Y}_i| = o_P(n^{-1/4})$. Hence the above sum is equal to $o_P(1)$.

Proof of Claim 3 : Let $\tilde{\mathcal{G}}_n$ be the class $\mathcal{G}_n$ in Corollary A1 with $\beta_{u_2}(u_1) = u_1$ with Λ_n as defined at the beginning of the proof of this theorem. Let $\mathcal{G}'_n = \{G \in \tilde{\mathcal{G}}_n : ||G - F_0 \circ \lambda_0||_\infty \leq C\varepsilon_n\}$ where ε_n is a decreasing sequence $\varepsilon_n = o(n^{-1/4})$. Without loss of generality, we assume that $\beta_u(\cdot)$ is increasing. Define $d_{u,G}(x) = \beta_u(G(X_i)) - \beta_u(G_0(X_i))$ where $G(\cdot) \in \mathcal{G}'_n$. We consider the following:

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{i=1}^n d_{u,G}(X_i)\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i) \qquad (30) \\ = \; & \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ d_{u,G}(X_i)\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i) - \mathbf{E}[d_{u,G}(X_i)\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i)] \right\} \\ & + \sqrt{n}\mathbf{E}[d_{u,G}(X_i)\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i)]. \end{aligned}$$

Note that $\mathbf{E}[d^2_{u,G}(X_i)] \leq \mathbf{E}[\{\beta_u(U_i + C\varepsilon_n) - \beta_u(U_i - C\varepsilon_n)\}^2] = o(1)$. By Lemma A1 and checking the bracketing entropy for the class

$$\{d_{u,G}(\cdot)\gamma_{z,\delta_n}(\cdot)\{\gamma_{y,\delta_n}(\cdot) - \mu_{\delta_n}(y)\} : (G, r) \in \mathcal{G}'_n \times [0,1]^3\},$$

we can show that the first process is stochastically equicontinuous in $(G, r) \in \mathcal{G}'_n \times [0,1]^3$ and hence it is $o_P(1)$. The last term in (30) is equal to

$$\sqrt{n}\mathbf{E}[\{\beta_u(U_{n,\theta,i}) - \beta_u(U_i)\}\mathbf{E}[\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i)|U_{\theta,i}, U_i]|(Y_k, X_k, Z_k)^n_{k=1,k\neq i} = (y_k, x_k, z_k)^n_{k=1,k\neq i}],$$

where $(y_k, x_k, z_k)^n_{k=1,k\neq i}$ constitute the function G. Hence the above is bounded by

$$\sqrt{n}\mathbf{E}\left[\{\beta_u(U_i + C\varepsilon_n) - \beta_u(U_i - C\varepsilon_n)\}|\mathbf{E}[\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i)|U_{\theta,i}, U_i]|\right]. \qquad (31)$$

Under the null hypothesis or the local alternatives in the theorem,

$$\begin{aligned} & \left|\mathbf{E}[\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i)|U_{\theta,i}, U_i]\right| \\ = \; & \left|\mathbf{E}[\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i)|U_{\theta,i}, U_i] - \mathbf{E}[\gamma^\perp_{z,\delta_n}(\tilde{Z}_i)\gamma^\perp_{y,\delta_n}(\tilde{Y}_i)|U_i]\right| + o(n^{-1/4}). \end{aligned}$$

By applying Assumption 2 (ii)(d) to Lemma A2(ii) of Song (2006), we deduce that the leading term is $o(n^{-1/4})$. Hence the last term in (30) is $o(1)$.

(ii) Since the products of the indicator functions constitute P-Donsker classes, the weak convergence result immediately follows. The local shift result follows from the uniform law of large numbers. ■

Lemma A3 : $\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\left\{\gamma^\perp_z(\hat{Z}_i)\gamma^\perp_y(\hat{Y}_i) - \gamma^\perp_{z,\delta_n}(\hat{Z}_i)\gamma^\perp_{y,\delta_n}(\hat{Y}_i)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \beta_u(U_i)\Delta_{z,y,n}(\tilde{Z}_i, \tilde{Y}_i) + o_P(1)$ uniformly over $(\beta, r) \in \mathcal{G}_n \times (0,1]^2 \times [0,1]$.

Proof of Lemma A3 : Let $\eta_n = n^{-1/2}h^{-1} + h^2$. For some sufficiently large $M > 0$, let $1_n = 1\{\max_{1\leq i\leq n}|\hat{Y}_i - \tilde{Y}_i| < M\eta_n, \max_{1\leq i\leq n}|\hat{Z}_i - \tilde{Z}_i| < M\eta_n\}$. Then, since by Lemma B3 $\hat{F}_{Y|U}(\cdot|\cdot)$ converges uniformly to $F_{Y|U}(\cdot|\cdot)$ at the rate of $O_P(\eta_n)$, we obtain $P\{1_n = 1\} < \varepsilon$ by choosing arbitrarily large M. Hence it suffices to consider only the event $1_n = 1$ which we assume throughout the proof. First we show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\left\{\gamma^\perp_z(\hat{Z}_i)\gamma^\perp_y(\hat{Y}_i) - \gamma^\perp_{z,\delta_n}(\hat{Z}_i)\gamma^\perp_{y,\delta_n}(\hat{Y}_i)\right\} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\Delta_{z,y,n}(\tilde{Z}_i, \tilde{Y}_i) + o_P(1), \qquad (32)$$

uniformly over $(\beta, r) \in \mathcal{G}_n \times [0,1]^3$. We write

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \gamma_z^{\perp}(\hat{Z}_i)\gamma_y^{\perp}(\hat{Y}_i) - \gamma_{z,\delta_n}^{\perp}(\hat{Z}_i)\gamma_{y,\delta_n}^{\perp}(\hat{Y}_i) - \Delta_{z,y,n}(\tilde{Z}_i, \tilde{Y}_i) \right\} \\
= \; & \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \gamma_y(\hat{Y}_i) - \gamma_y(\tilde{Y}_i) \right\} \gamma_z^{\perp}(\hat{Z}_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \gamma_{y,\delta_n}(\hat{Y}_i) - \gamma_{y,\delta_n}(\tilde{Y}_i) \right\} \gamma_{z,\delta_n}^{\perp}(\hat{Z}_i) \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \gamma_z(\hat{Z}_i) - \gamma_z(\tilde{Z}_i) \right\} \gamma_y^{\perp}(\tilde{Y}_i) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \gamma_{z,\delta_n}(\hat{Z}_i) - \gamma_{z,\delta_n}(\tilde{Z}_i) \right\} \gamma_{y,\delta_n}^{\perp}(\tilde{Y}_i).
\end{aligned}
$$

We focus only on the first difference of two sums which we write

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \eta_{y,\delta_n}(\hat{Y}_i) - \eta_{y,\delta_n}(\tilde{Y}_i) \right\} \gamma_z^{\perp}(\hat{Z}_i) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \gamma_{y,\delta_n}(\hat{Y}_i) - \gamma_{y,\delta_n}(\tilde{Y}_i) \right\} \eta_{z,\delta_n}(\hat{Z}_i), \qquad (33)
$$

where $\eta_{y,\delta_n}(\cdot) = \gamma_y(\cdot) - \gamma_{y,\delta_n}(\cdot)$. Write $\frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \eta_{y,\delta_n}(\hat{Y}_i) - \eta_{y,\delta_n}(\tilde{Y}_i) \right\} \gamma_z^{\perp}(\hat{Z}_i)$ as

$$
\begin{aligned}
& \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \eta_{y,\delta_n}(\hat{Y}_i) - \eta_{y,\delta_n}(\tilde{Y}_i) \right\} \gamma_z^{\perp}(\hat{Z}_i) 1 \left\{ |\tilde{Y}_i - y| \leq \eta_n \right\} \\
& + \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \eta_{y,\delta_n}(\hat{Y}_i) - \eta_{y,\delta_n}(\tilde{Y}_i) \right\} \gamma_z^{\perp}(\hat{Z}_i) 1 \left\{ |\tilde{Y}_i - y| > \eta_n \right\}.
\end{aligned}
$$

Observe that for any $u, y \in [0,1]$, we can write $\left| \eta_{y,\delta_n}(u) - \eta_{y,\delta_n}(u') \right| 1 \{ |u - y| \leq \eta_n \}$ as

$$
\begin{aligned}
& \left| 1 - \Phi\left(\frac{y-u}{\delta_n} \right) - \gamma_y(u') + \Phi\left(\frac{y-u'}{\delta_n} \right) \right| 1 \{ y < u \leq y + \eta_n \} \\
& + \left| -\Phi\left(\frac{y-u}{\delta_n} \right) - \gamma_y(u') + \Phi\left(\frac{y-u'}{\delta_n} \right) \right| 1 \{ y - \eta_n \leq u \leq y \} \\
\leq \; & \left\{ \Phi\left(\frac{y-u+\eta_n}{\delta_n} \right) - \Phi\left(\frac{y-u}{\delta_n} \right) \right\} 1 \{ y < u \leq y + \eta_n \} \\
& + \left\{ \Phi\left(\frac{y-u-\eta_n}{\delta_n} \right) - \Phi\left(\frac{y-u}{\delta_n} \right) \right\} 1 \{ y - \eta_n \leq u \leq y \}.
\end{aligned}
$$

The inequality can be checked by drawing the graph of $\eta_{y,\delta_n}(u)$. Taking integral and applying change of variables, the first term becomes

$$
\delta_n \int_{-\eta_n/\delta_n}^{0} \left\{ \Phi\left(u + \frac{\eta_n}{\delta_n} \right) - \Phi(u) \right\} du = O\left(\delta_n \{ \exp(0) - \exp(-(\eta_n/\delta_n)^2) \} \right) \leq C\eta_n^3/\delta_n^2.
$$

We deal with the second term similarly to deduce that these last two terms are bounded by $C\frac{\eta_n}{\delta_n} 1 \{ |u-y| \leq \eta_n \}$. Therefore,

$$
\begin{aligned}
& \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \beta(X_i) \left\{ \eta_{y,\delta_n}(\hat{Y}_i) - \eta_{y,\delta_n}(\tilde{Y}_i) \right\} \gamma_z^{\perp}(\hat{Z}_i) 1 \left\{ |\tilde{Y}_i - y| \leq \eta_n \right\} \right| \\
\leq \; & \frac{C\eta_n^2}{\delta_n^2} \left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(1 \left\{ \left| \tilde{Y}_i - y \right| \leq \eta_n \right\} - \mathbf{E} 1 \left\{ \left| \tilde{Y}_i - y \right| \leq \eta_n \right\} \right) \right\} + \frac{C\eta_n^3 \sqrt{n}}{\delta_n^2} = o_P(1).
\end{aligned}
$$

Now note that $1\{u > y + \eta_n\} \left|\eta_{y,\delta_n}(u) - \eta_{y,\delta_n}(u')\right|$ is bounded by

$$1\{u > y + \eta_n\} \left|\eta_{y,\delta_n}(u) - \eta_{y,\delta_n}(u - \eta_n)\right| \leq 1\{u > y + \eta_n\} \left\{\Phi\left(\frac{y-u}{\delta_n} + \frac{\eta_n}{\delta_n}\right) - \Phi\left(\frac{y-u}{\delta_n}\right)\right\}.$$

Reasoning similarly with $1\{u < y - \eta_n\} \left|\eta_{y,\delta_n}(u) - \eta_{y,\delta_n}(u')\right|$, we deduce that

$$\begin{aligned} & \left|\frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\left\{\eta_{y,\delta_n}(\hat{Y}_i) - \eta_{y,\delta_n}(\tilde{Y}_i)\right\}\gamma_z^\perp(\hat{Z}_i)1\left\{|\tilde{Y}_i - y| > \eta_n\right\}\right| \qquad (34) \\ \leq\ & \frac{1}{\sqrt{n}}\sum_{i=1}^n 1\left\{\tilde{Y}_i > y + \eta_n\right\}\left|\eta_{y,\delta_n}(\tilde{Y}_i - \eta_n) - \eta_{y,\delta_n}(\tilde{Y}_i)\right| + \frac{1}{\sqrt{n}}\sum_{i=1}^n 1\left\{\tilde{Y}_i < y - \eta_n\right\}\left|\eta_{y,\delta_n}(\tilde{Y}_i) - \eta_{y,\delta_n}(\tilde{Y}_i + \eta_n)\right|. \end{aligned}$$

Write the first sum as

$$\begin{aligned} & \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(1\left\{\tilde{Y}_i > y + \eta_n\right\}\left|\eta_{y,\delta_n}(\tilde{Y}_i - \eta_n) - \eta_{y,\delta_n}(\tilde{Y}_i)\right| - \mathbf{E}\left[1\left\{\tilde{Y}_i > y + \eta_n\right\}\left|\eta_{y,\delta_n}(\tilde{Y}_i - \eta_n) - \eta_{y,\delta_n}(\tilde{Y}_i)\right|\right]\right) \\ & + \sqrt{n}\mathbf{E}\left[1\left\{\tilde{Y}_i > y + \eta_n\right\}\left|\eta_{y,\delta_n}(\tilde{Y}_i - \eta_n) - \eta_{y,\delta_n}(\tilde{Y}_i)\right|\right] \end{aligned}$$

Since $\eta_n/\delta_n \to 0$, we can show that the first term is $o_P(1)$ using the usual arguments of stochastic equicontinuity. The second sum is equal to

$$\begin{aligned} & \sqrt{n}\mathbf{E}\left[1\left\{\tilde{Y}_i > y + \eta_n\right\}\left\{\Phi\left(\frac{y-\tilde{Y}_i}{\delta_n} + \frac{\eta_n}{\delta_n}\right) - \Phi\left(\frac{y-\tilde{Y}_i}{\delta_n}\right)\right\}\right] \\ =\ & \sqrt{n}\int_{y+\eta_n}^1 \left\{\Phi\left(\frac{y-u}{\delta_n} + \frac{\eta_n}{\delta_n}\right) - \Phi\left(\frac{y-u}{\delta_n}\right)\right\}du = \delta_n\sqrt{n}\int_{\eta_n/\delta_n}^{(1-y)/\delta_n}\left\{\Phi\left(-u + \frac{\eta_n}{\delta_n}\right) - \Phi(-u)\right\}du \\ =\ & O\left(\sqrt{n}\delta_n\left\{\exp(0) - \exp(-\eta_n^2/\delta_n^2)\right\}\right) = O\left(\sqrt{n}\left(\frac{\eta_n}{\delta_n}\right)^2\delta_n\right) = O\left(\frac{\sqrt{n}\eta_n^2}{\delta_n}\right) = o(1). \end{aligned}$$

The last equality is due to the fact that $\sqrt{n}(n^{-1/2}h^{-1} + h^2)^2\delta_n^{-1} \to 0$. Similarly, we can show that the second term in (34) is $o_P(1)$. Hence we conclude that the first sum in (33) is $o_P(1)$. We turn to the second sum there. First, write the sum as

$$\begin{aligned} & \frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\left\{\gamma_{y,\delta_n}(\hat{Y}_i) - \gamma_{y,\delta_n}(\tilde{Y}_i)\right\}\left\{\eta_{z,\delta_n}(\hat{Z}_i) - \eta_{z,\delta_n}(\tilde{Z}_i)\right\} \qquad (35) \\ & + \frac{1}{\sqrt{n}}\sum_{i=1}^n \beta(X_i)\left\{\gamma_{y,\delta_n}(\hat{Y}_i) - \gamma_{y,\delta_n}(\tilde{Y}_i)\right\}\eta_{z,\delta_n}(\tilde{Z}_i). \end{aligned}$$

Similarly as before, we can show that the first sum is $o_P(1)$. Take a sequence $\tilde{\eta}_n$ such that $\tilde{\eta}_n/\delta_n \to \infty$ and $\tilde{\eta}_n \to 0$. Then write

$$\begin{aligned} \eta_{y,\delta_n}(u) &= 1\{u \leq y\}\left\{1 - \Phi\left(\frac{y-u}{\delta_n}\right)\right\} - 1\{u > y\}\Phi\left(\frac{y-u}{\delta_n}\right) \\ &= 1\{y - \tilde{\eta}_n < u \leq y\}\left\{1 - \Phi\left(\frac{y-u}{\delta_n}\right)\right\} - 1\{u > y + \tilde{\eta}_n\}\Phi\left(\frac{y-u}{\delta_n}\right) \\ &\quad + 1\{u \leq y - \tilde{\eta}_n\}\left\{1 - \Phi\left(\frac{y-u}{\delta_n}\right)\right\} - 1\{y + \tilde{\eta}_n \geq u > y\}\Phi\left(\frac{y-u}{\delta_n}\right). \end{aligned}$$

From this we deduce that

$$|\eta_{y,\delta_n}(u)| \leq 1\{y-\tilde{\eta}_n < u \leq y+\tilde{\eta}_n\}/2 + 1\{u > y+\tilde{\eta}_n\}\Phi\left(-\frac{\tilde{\eta}_n}{\delta_n}\right) + 1\{u \leq y-\tilde{\eta}_n\}\left\{1-\Phi\left(\frac{\tilde{\eta}_n}{\delta_n}\right)\right\}.$$

Hence the L_2-norm of $\eta_{y,\delta_n}(u)$ with respect the uniform distribution $[0,1]$ vanishes as n goes to infinity. Using this fact and following the steps of the proof of Claim 2, we can show that the second sum in (35) is $o_P(1)$.

Now, for the completion of the proof, it suffices to show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{\beta(X_i)-\beta_u(U_i)\}\Delta_{z,y,n}(\tilde{Z}_i,\tilde{Y}_i) = o_P(1).$$

The proof can be proceeded exactly in the same manner as in the proof of Claim 3. We omit the details. ■

Proof of Theorem 2 : (i) Let $\mathcal{G}_n'$ be as in the proof of Lemma A2 and let $e(p;z)=(z-p)/\sqrt{p-p^2}$. For a fixed sequence $D_{n,i}=\{(z_k,u_k)\}_{k=1,k\neq i}^n$, define

$$p_n(u;D_{n,i}) = \frac{\sum_{j=1,j\neq i}^n z_j K_h(u_j-u)}{\sum_{j=1,j\neq i}^n K_h(u_j-u)}.$$

Let $\mathcal{D}_n$ be the set of sequences $\{(z_k,u_k)\}_{k=1,k\neq i}^n \in \{0,1\}^{n-1}\times[0,1]^{n-1}$ such that

$$\inf_{u\in[0,1]}\frac{1}{nh}\sum_{j=1,j\neq i}^{n}K_h(u_j-u) > \eta$$

for a small $\eta\in(0,1)$ and $||p_n(\cdot;\{(z_k,u_k)\}_{k=1,k\neq i}^n)-p(\cdot)||_\infty \leq \varepsilon_n$ for a sequence ε_n decreasing at a slower rate than the convergence rate of $\hat{p}$. Let $\mathcal{P}_n=\{p_n(\cdot;D_n):D_n\in\mathcal{D}_n\}$ and $\mathcal{P}_n^{\mathcal{G}}=\{p_n(G(\cdot);D_n):(D_n,G)\in\mathcal{D}_n\times\mathcal{G}_n'\}$. Since by Lemma B3, $\hat{p}(\cdot)$ converges uniformly to $p(\cdot)$ at the rate of $o_P(n^{-1/4})$ and $\frac{1}{nh}\sum_{j=1,j\neq i}^n K_h(\hat{U}_j-u)$ is uniformly consistent for $f(u)=1$, the probability $P\left\{\hat{p}\in\mathcal{P}_n^{\mathcal{G}}\right\}$ is bounded by

$$P\left\{\hat{p}\in\mathcal{P}_n^{\mathcal{G}}\right\}+P\left\{\frac{1}{nh}\sum_{j=1,j\neq i}^{n}K_h(\hat{U}_j-u)<\eta\right\}\to 1 \text{ as } n\to\infty.$$

Let us compute the bracketing entropies of $\mathcal{P}_n$ and $\mathcal{P}_n^{\mathcal{G}}$. Since $|p_n(u-\varepsilon)-p(u+\varepsilon)|\leq Ch^{-2}\varepsilon$, Lemma A2 and Lemma A1 above gives

$$\begin{aligned}
\log N_{[]}(\varepsilon,\mathcal{P}_n,L_p(P)) &\leq -C\log(h)-C\log(\varepsilon).\\
\log N_{[]}(\varepsilon,\mathcal{P}_n^{\mathcal{G}},L_p(P)) &\leq \log N_{[]}((h^{-2}\varepsilon)^p,\Lambda_n,||\cdot||_\infty)+C_1h^{-2}/\varepsilon.
\end{aligned}$$

Now, consider for $(\beta,\tilde{p})\in\mathcal{G}_n\times\mathcal{P}_n^{\mathcal{G}}$,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)e(\tilde{p}(X_i);Z_i)\left\{\gamma_{y,\delta_n}(\hat{Y}_i)-\mu_{\delta_n}(y)\right\} \qquad (36)\\
=\ &\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)e(\tilde{p}(X_i);Z_i)\left\{\gamma_{y,\delta_n}(\hat{Y}_i)-\gamma_{y,\delta_n}(\tilde{Y}_i)\right\}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta(X_i)e(\tilde{p}(X_i);Z_i)\left\{\gamma_{y,\delta_n}(\tilde{Y}_i)-\mu_{\delta_n}(y)\right\}.
\end{aligned}$$

Similarly as in the proof of Claim 2, the leading sum in the second line is equal to

$$-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_u(U_i)\mathbf{E}\left[e(p(U_i);Z_i)|U_i\right]\left\{\gamma_{y,\delta_n}(\tilde{Y}_i)-\mu_{\delta_n}(y)\right\}+o_P(1)=o_P(1)$$

leaving us to deal with the last sum in (36). Using similar arguments in the proof of Claim 3 along with the fact that $e(p;z)$ is infinitely differentiable in p on $(\varepsilon,1-\varepsilon)$ with uniformly bounded derivatives, we can show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_u(G(X))\left\{e(\bar{p}(G(X));Z_i)-e(p(U_i);Z_i)\right\}\left\{\gamma_{y,\delta_n}(\tilde{Y}_i)-y\right\}=o_P(1),$$

where $\bar{p}\in\mathcal{P}_n$. The proof is complete.

(ii) The proof is similar to that of Theorem 1(ii) and is omitted. ■

Proof of Theorem 3 : We consider only the case of

$$\bar{\nu}^*_{n,b}(r)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_{i,b}\beta_u(\hat{U}_i)\frac{Z_i-\hat{p}(U_i)}{\sqrt{\hat{p}(U_i)-\hat{p}(U_i)^2}}\left\{\gamma_y(\hat{Y}_i)-y\right\}.$$

The other case can be dealt with similarly given the result of Theorem 1. Following the proof of Theorem 2, we can show that

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_{i,b}\beta_u(\hat{U}_i)\frac{Z_i-\hat{p}(U_i)}{\sqrt{\hat{p}(U_i)-\hat{p}(U_i)^2}}\left\{\gamma_y(\hat{Y}_i)-y\right\}\\ =\;&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\omega_{i,b}\beta_u(U_i)\frac{Z_i-p(U_i)}{\sqrt{p(U_i)-p(U_i)^2}}\left\{\gamma_y(\tilde{Y}_i)-y\right\}+o_P(1)\end{aligned}\tag{37}$$

in probability. The weak convergence of the above process follows by the almost sure conditional multiplier central limit theorem of Ledoux and Talagrand (1988) (e.g. see Theorem 2.9.7 of van der Vaart and Wellner (1996), p.183) and it is easy to see that the covariance function of the above is equal to that of

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_u(U_i)\frac{Z_i-p(U_i)}{\sqrt{p(U_i)-p(U_i)^2}}\left\{\gamma_y(\tilde{Y}_i)-y\right\},$$

under the null hypothesis, because ω_i is centered, i.i.d., bounded, independent of $(X_i,Y_i,Z_i)_{i=1}^n$ and $\mathbf{E}\omega_i^2=1$. ■

10.2 Uniform Asymptotic Representation of a Semiparametric Empirical Process

Let Ψ_n and Φ_n be classes of functions $\psi:\mathbf{R}^{d_S}\to\mathbf{R}$ and $\varphi:\mathbf{R}^{d_W}\times\mathbf{R}^{d_W}\to\mathbf{R}$ that satisfy Assumptions B1 and B2 below. Then we introduce a kernel estimator of $g_\varphi(u,w)=\mathbf{E}\left[\varphi(W,w)|U=u\right]$ as follows:

$$\hat{g}_{\varphi,\theta,i}(u,w)=\frac{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\varphi(W_j,w)K_h\left(U_{n,\theta,j}-u\right)}{\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}K_h\left(U_{n,\theta,j}-u\right)}.\tag{38}$$

Let us define a shrinking neighborhood of θ_0 : $\Theta_n=\{\theta\in\Theta:||\theta-\theta_0||\leq Cn^{-1/4}\}$ for a fixed constant C.

Assumption B1 : The functions $g_{\varphi,\theta}(u,w)$ are twice continuously differentiable in u with the derivatives $g^{(1)}_{\varphi,\theta}(u,w)$ and $g^{(2)}_{\varphi,\theta}(u,w)$ satisfying $\mathbf{E}[\sup_{\varphi\in\Phi,\ \theta\in\Theta}|g^{(1)}_{\varphi,\theta}(U_i,W_i)|^{2+\delta}]$ and $\mathbf{E}[\sup_{\varphi\in\Phi,\ \theta\in\Theta}|g^{(2)}_{\varphi,\theta}(U_i,W_i)|^{2+\delta}] < \infty$ for some $\delta > 0$.

Assumption B2 : For classes Φ_n and Ψ_n, there exist $b_\Phi, b_\Psi \in [0,2)$ and sequences b_n and d_n such that $b_\Psi \vee b_\Phi \in [\frac{p}{p-1+p\delta}, \frac{2}{1+2\delta})$ for some $\delta > 0$, $p > 2$, and

$$\log N_{[]}(\varepsilon, \Phi_n, ||\cdot||_p) < b_n \varepsilon^{-b_\Phi},\ \log N_{[]}(\varepsilon, \Psi_n, ||\cdot||_p) < d_n \varepsilon^{-b_\Psi}$$

and envelopes $\tilde{\varphi}$ and $\tilde{\psi}$ for Φ_n and Ψ_n satisfy that $\mathbf{E}[\sup_{w\in\mathbf{R}^{d_W}}|\tilde{\varphi}(W,w)|^p|X] < \infty$, $\mathbf{E}[\sup_{w\in\mathbf{R}^{d_W}}|\tilde{\varphi}(w,W)|^p|X] < \infty$, and $\mathbf{E}[|\tilde{\psi}(S)|^p|X] < \infty$, a.s., for some $\varepsilon > 0$.

Assumption B3 : (i) As for the kernel K, suppose that Assumption 5(i) holds. (ii) For δ in Assumption B2 and for arbitrarily small $\varepsilon > 0$, $h \to 0$,

$$n^{-\frac{1}{b_\Psi \vee b_\Phi}+\delta+\frac{1}{2}}(b_n \vee d_n)^{\frac{1}{2}} \to 0, \text{and } n^{-1/2+\varepsilon}h^{-1} \to 0.$$

We establish that under these conditions, the processes $\hat{\Delta}_n(\theta,\varphi,\psi)$ and $\Delta_n(\theta,\varphi,\psi)$, $(\theta,\varphi,\psi) \in \Theta_n \times \Phi_n \times \Psi_n$ defined by

$$\hat{\Delta}_n(\theta,\varphi,\psi) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(S_i)\left\{\hat{g}_{\varphi,\theta,i}(U_{n,\theta,i},W_i) - g_\varphi(U_i,W_i)\right\}$$

are asymptotically equivalent uniformly over $(\theta,\varphi,\psi) \in \Theta_n \times \Phi_n \times \Psi_n$. Similar results in the case of series estimation can be found in Lemma 1U of Escanciano and Song (2007) and Lemma 1U of Song (2007). A related, nonuniform result was also obtained by Stute and Zhu (2005) (SZ hereafter).[15]

Lemma B1 : Suppose that Assumption 5 holds for the kernel and the bandwidth and Assumptions B1-B2 hold. Then

$$\sup_{(\theta,\varphi,\psi)\in\Theta_n\times\Phi_n\times\Psi_n}\left|\hat{\Delta}_n(\theta,\varphi,\psi) - \frac{1}{\sqrt{n}}\sum_{i=1}^n b_{\psi,\varphi}(W_i,U_i)\right| = o_P(1),$$

where $b_{\psi,\varphi}(w,u) = \mathbf{E}\left[\psi(S_i)\{\varphi(w,W_i) - g_\varphi(U_i,W_i)\}|U_i = u\right].$

Proof of Lemma B1 : Define

$$\hat{\rho}_{\varphi,\theta,i}(u,w) = \frac{1}{n-1}\sum_{j=1,j\neq i}^n K_h(U_{n,\theta,j}-u)\varphi(W_j,w)$$

and write $\hat{g}_{\varphi,\theta,i}(U_{n,\theta,i},W_i) - g_\varphi(U_i,W_i)$ as

$$\begin{aligned} & \hat{\rho}_{\varphi,\theta,i}(U_{n,\theta,i},W_i)/\hat{f}_{\theta,i}(U_{n,\theta,i}) - \rho_\varphi(U_i,W_i)/f_0(U_i) \\ = & \left[\hat{\rho}_{\varphi,\theta,i}(U_{n,\theta,i},W_i) - g_\varphi(U_i,W_i)\hat{f}_{\theta,i}(U_{n,\theta,i})\right]\left\{1+\tilde{\delta}_{n,\theta,i}\right\}/f_\theta(U_{n,\theta,i}) \\ = & \left[\hat{\rho}_{\varphi,\theta,i}(U_{n,\theta,i},W_i) - g_\varphi(U_i,W_i)\hat{f}_{\theta,i}(U_{n,\theta,i})\right]\left\{1+\tilde{\delta}_{n,\theta,i}\right\}, \end{aligned}$$

[15] Note that we do not need the condition $n^{1/2}h^2 \to \infty$ used by SZ. This condition was used to prove Lemma 4.3 there. The formulation of our lemma is different and does not require a counterpart of Lemma 4.3.

where $\tilde{\delta}_{n,\theta,i} = \delta_{n,\theta,i} + \delta^2_{n,\theta,i}(1-\delta_{n,\theta,i})^{-1}$ and $\delta_{n,\theta,i} = f_0(U_i) - \hat{f}_{\theta,i}(U_{n,\theta,i})$ and f_0 and f_θ are the density of a uniform random variate. Let $\Delta_n = \max_{1\le i\le n}\sup_{\theta\in\Theta_n}|f_0(U_i) - \hat{f}_{\theta,i}(U_{n,\theta,i})|$. Since $f_0(u) = 1$, we have

$$\Delta_n = \max_{1\le i\le n}\sup_{\theta\in\Theta_n}|f_0(U_{n,\theta,i}) - \hat{f}_{\theta,i}(U_{n,\theta,i})| \le \max_{1\le i\le n}\sup_{\theta\in\Theta_n}\sup_{u\in[0,1]}\left|f_0(u) - \hat{f}_{\theta,i}(u)\right| = o_P(1)$$

by Lemma B3 below, so that $\max_{1\le i\le n}\sup_{\theta\in\Theta_n}\left|\tilde{\delta}_{n,\theta,i}\right| \le \Delta_n + \Delta_n^2/(1-\Delta_n) = o_P(1)$. Hence it suffices to show that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(S_i)\left[\hat{\rho}_{\varphi,\theta,i}(U_{n,\theta,i},W_i) - g_\varphi(U_i,W_i)\hat{f}_{\theta,i}(U_{n,\theta,i})\right] \\
=\ & \frac{1}{\sqrt{n}}\sum_{i=1}^n b_{\psi,\varphi}(W_i,U_i) + o_P(1).
\end{aligned}$$

Since $f_0(\cdot) = 1$, we can write $\hat{\rho}_{\varphi,\theta,i}(U_{n,\theta,i},W_i) - g_\varphi(U_i,W_i)\hat{f}_{\theta,i}(U_{n,\theta,i})$ as

$$\hat{\rho}_{\varphi,\theta,i}(U_{n,\theta,i},W_i) - g_\varphi(U_i,W_i) + g_\varphi(U_i,W_i)\left[f_0(U_i) - \hat{f}_{\theta,i}(U_{n,\theta,i})\right].$$

As for the first difference, observe that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(S_i)\left[\hat{\rho}_{\varphi,\theta,i}(U_{n,\theta,i},W_i) - g_\varphi(U_i,W_i)\right] \\
=\ & \frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\ne i}^n \psi(S_i)\left[\varphi(W_j,W_i)K_h(U_{n,\theta,j} - U_{n,\theta,i}) - g_\varphi(U_i,W_i)\right] \\
=\ & \frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\ne i}^n \psi(S_i)\left\{\varphi(W_j,W_i) - g_\varphi(U_i,W_i)\right\}K_h(U_{n,\theta,j} - U_{n,\theta,i}) \\
&-\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(S_i)g_\varphi(U_i,W_i)\left[f_0(U_i) - \hat{f}_{\theta,i}(U_{n,\theta,i})\right].
\end{aligned}$$

Therefore, $\frac{1}{\sqrt{n}}\sum_{i=1}^n \psi(S_i)\left[\hat{\rho}_{\varphi,\theta,i}(U_{n,\theta,i},W_i) - g_\varphi(U_i,W_i)\hat{f}_{\theta,i}(U_{n,\theta,i})\right]$ is equal to

$$\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\ne i}^n \psi(S_i)\Delta_{\varphi,ij}K_h(U_{n,\theta,j} - U_{n,\theta,i}) + o_P(1).$$

where $\Delta_{\varphi,ij} = \varphi(W_j,W_i) - g_\varphi(U_i,W_i)$. Now, write the double sum as

$$\begin{aligned}
&\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\ne i}^n \psi(S_i)\Delta_{\varphi,ij}\left\{K_h(U_{n,\theta,j} - U_{n,\theta,i}) - K_h(U_j - U_i)\right\} \\
&+\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^n\sum_{j=1,j\ne i}^n \psi(S_i)\Delta_{\varphi,ij}K_h(U_j - U_i) = A_{1n} + A_{2n},\ \text{say.}
\end{aligned}$$

Let us consider A_{1n} first. Following the steps in the proof of Lemma 4.7 in SZ, we can write it as

$$\begin{aligned}
& \frac{1}{(n-1)nh^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\psi(S_i)\Delta_{\varphi,ij}K'\left(\frac{U_j-U_i}{h}\right)\sqrt{n}\left\{U_{n,j}-U_j-(U_{n,i}-U_i)\right\} \\
& +\frac{1}{(n-1)\sqrt{n}h^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\psi(S_i)\Delta_{\varphi,ij}K'\left(\frac{U_j-U_i}{h}\right)\left\{U_{\theta,j}-U_j-(U_{\theta,i}-U_i)\right\}+o_P(1) \\
= \;& B_{1n}+B_{2n},\text{ say.}
\end{aligned}$$

Due to Lemma The derivation requires the counterparts of Lemmas 4.1 and 4.2 in SZ which are proved in Lemmas B3 and B4 below. The term B_{2n} corresponds to (4.13) in SZ and can be shown to be $o_P(1)$ in a similar manner as there, leaving us to deal with B_{1n}.

We turn to B_{1n} which we write as

$$\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}q_n(S_j,S_i;\pi)+o_P(1)$$

where $q_n(S_j,S_i;\pi)=\psi(S_i)\Delta_{\varphi,ij}\frac{1}{h^2}K'\left(\frac{U_j-U_i}{h}\right)\{G_1(U_j)-U_j-G_2(U_i)+U_i\}$, with $\pi=(\psi,\varphi,G_1,G_2)\in\Psi_n\times\Phi_n\times\tilde{\mathcal{G}}_n\times\tilde{\mathcal{G}}_n$ and $\tilde{\mathcal{G}}_n$ is the class $\mathcal{G}_n$ in Corollary A1 with $\beta_{u_2}(u_1)=u_1$ such that for all $G\in\tilde{\mathcal{G}}_n$, $\sup_{u\in[0,1]}|G(u)-u|<Cn^{-1/2+\varepsilon}$, and Λ_n being a singleton of $F_0\circ\lambda_0$. Observe that by the uniform central limit theorem, $F_{n,j}(u),F_{n,i}(u)\in\tilde{\mathcal{G}}_n$ with probability approaching one. The leading sum in the preceding display is a U-process with a kernel depending on n. Let

$$\begin{aligned}
\bar{q}_n(S_j,S_i;\pi) &= \mathbf{E}\left[q_n(S_j,S_i;\pi)|S_j\right]+\mathbf{E}\left[q_n(S_j,S_i;\pi)|S_i\right]-2\mathbf{E}\left[q_n(S_j,S_i;\pi)\right]\text{ and} \\
u_n(S_j,S_i;\pi) &= q_n(S_j,S_i;\pi)-\bar{q}_n
\end{aligned}$$

and write the above double sum as

$$\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\bar{q}_n(S_j,S_i;\pi)+\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}u_n(S_j,S_i;\pi). \tag{39}$$

First note that $\mathbf{E}\left[q_n(S_j,S_i;\pi)|S_i\right]=0$ because $\mathbf{E}\left[\Delta_{\varphi,ij}|S_i,U_j\right]=0$. Hence the first double sum above becomes

$$\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n}\mathbf{E}\left[q_n(S_j,S_i;\pi)|S_j\right],$$

which is equal to(after change of variables)

$$\frac{1}{\sqrt{n}}\sum_{j=1,j\neq i}^{n}\int b_{\psi,\varphi}(W_j,U_j+hv)\frac{1}{h}K'(v)\left\{G_1(U_j)-U_j-G_2(U_j+hv)+U_j+hv\right\}dv.$$

The sum above is a mean zero process with its variance bounded by $C(h^{-1}n^{-1/2+\varepsilon})^2\to 0$. The convergence uniform over $(\psi,\varphi,G_1,G_2)\in\Psi_n\times\Phi_n\times\tilde{\mathcal{G}}_n\times\tilde{\mathcal{G}}_n$ can be obtained using the usual stochastic equicontinuity arguments. Later we show that the second sum in (39) vanishes in probability. Thus we conclude that $A_{1n}=o_P(1)$.

Now let us turn to A_{2n} for which we let $q_n^*(S_j,S_i;\pi)=\psi(S_i)\Delta_{\varphi,ij}\frac{1}{h}K\left(\frac{U_j-U_i}{h}\right)$ and consider the following

Hoeffding decomposition:

$$\frac{\sqrt{n}}{n-1}\sum_{j=1,j\neq i}^{n}\mathbf{E}\left[\mathbf{E}\left[q_n^*(S_j,S_i;\pi)|U_i,S_j\right]|S_j\right]=\frac{\sqrt{n}}{n-1}\sum_{j=1,j\neq i}^{n}\int b_{\psi,\varphi}(W_j,u)\frac{1}{h}K\left(\frac{U_j-u}{h}\right)du$$

$$=\frac{\sqrt{n}}{n-1}\sum_{j=1,j\neq i}^{n}\int b_{\psi,\varphi}(W_j,U_j-u'h)K\left(u'\right)du'=\frac{\sqrt{n}}{n-1}\sum_{j=1,j\neq i}^{n}b_{\psi,\varphi}(W_j,U_j).$$

We are left with the last sum in (39) which is a degenerate U-process. Let us define $\mathcal{J}_n=\{q_n(\cdot,\cdot;\pi):\pi\in\Pi_n\}$, $\Pi_n=\Theta_n\times\Phi_n\times[0,1]$. Using the bracketing entropy bound for $\mathcal{J}_n$ in Lemma B2 below, we can apply Theorem 1 (i) in Turki-Moalla (1998), p. 878, to obtain [16]

$$\sup_{\pi\in\Phi_n\times\mathcal{B}\times\Lambda}\left\|\frac{1}{(n-1)\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}u_n(S_j,S_i;\pi))\right\|=o_{a.s}(n^{-\frac{1}{b_\Psi\vee b_\Phi}+\delta+\frac{1}{2}}(b_n\vee d_n)^{\frac{1}{2}})=o_{a.s}(1)$$

by Assumption B2. ■

Lemma B2 : For the class $\mathcal{J}_n$ defined in the proof of Lemma B1, the following holds:

$$\log N_{[]}(\varepsilon,\mathcal{J}_n,||\cdot||_p)\leq C\varepsilon^{-(b_\Psi\vee b_\Phi)}(b_n\vee d_n),$$

where C is a constant.

Proof of Lemma B2 : Define $\kappa_h(u_1,u_2)=K_h\left(u_1-u_2\right)$ and write

$$\mathcal{J}_n=\left\{\psi(\cdot)\kappa_h(\cdot,\cdot)\left\{\varphi(\cdot)-g_\varphi(\cdot)\right\}\left\{G_1(\cdot)-G_2(\cdot)\right\}:(\psi,\varphi,G_1,G_2)\in\Psi_n\times\Phi_n\times\tilde{\mathcal{G}}_n\times\tilde{\mathcal{G}}_n\right\}.$$

We can take its envelope as $\bar{J}_n(S_i,S_j)=2\bar{\psi}(S_i)\kappa_h(U_i,U_j)\bar{\varphi}(\cdot)$. From tedious calculations,

$$\begin{aligned}\log N_{[]}(\varepsilon,\mathcal{J}_n,||\cdot||_p) &\leq \log N_{[]}(C\varepsilon/||\bar{J}_n||_{2p},\Psi_n,||\cdot||_{2p})+\log N_{[]}(C\varepsilon/||\bar{J}_n||_{2p},\Phi_n,||\cdot||_{2p})\\ &\leq C(b_n\vee d_n)\left\{\varepsilon/||\bar{J}_n||_{2p}\right\}^{-(b_\Psi\vee b_\Phi)}.\end{aligned}$$

Since $||\kappa_h||_{2p}$ is a constant uniformly over $h>0$ by assumptions, we obtain the wanted result. ■

Lemma B3 : Let $\mathcal{S}_W$ be the support of W and suppose that $\Phi_n'=\{\varphi(\cdot,w);(\varphi,w)\in\Phi_n\times\mathcal{S}_W\}$ satisfies the same bracketing entropy condition as that for Φ_n in Assumption B2. Then

$$\max_{1\leq i\leq n}\sup_{(\varphi,\theta)\in\Phi_n\times\Theta_n}\sup_{(u,w)\in[0,1]\times\mathbf{R}^{d_W}}|\hat{g}_{\varphi,\theta,i}(u,w)-g_\varphi(u,w)|=O_P(n^{-1/2}h^{-1}b_n^{1/2})+O_P(h^{-2}).$$

Proof of Lemma B3 : Let $\hat{\rho}_{\varphi,\theta,i}(u)$ be as defined in the proof of Lemma B1. Note that it suffices to show that

$$\max_{1\leq i\leq n}\sup_{(\varphi,\theta)\in\Phi_n\times\Theta_n}\sup_{u\in[0,1]}\left|\hat{\rho}_{\varphi,\theta,i}(u,w)-g_\varphi(u,w)\right|=o_P(1)$$

because by putting $\varphi=1$, the above implies the uniform consistency of the density estimator $\hat{f}_{\theta,i}(u)$, so that the above leads to the wanted result of the lemma. (Recall the arguments in the beginning of the proof

[16] In Theorem 1(i) in Turki-Moalla (1998), we take $\lambda=\frac{1}{b_\Phi\vee b_\Psi}-\delta$ for δ in Assumption B2 and apply Lemma B2.

of Lemma B1.) Let $\tilde{\mathcal{G}}_n$ be the class $\mathcal{G}_n$ in Corollary A1 with $\beta_{u_2}(u_1) = u_1$ and $\Lambda_n = \{\lambda_\theta : \theta \in \Theta_n\}$. Let $\varphi_w(\cdot) = \varphi(W_i, w)$. For each $G \in \tilde{\mathcal{G}}_n$, write

$$\begin{aligned}\hat{\rho}_{\varphi,\theta,i}(u) - g_\varphi(u) &= \frac{1}{n}\sum_{i=1}^n \{K_h(G(X_i)-u)\varphi(W_i,w) - \mathbf{E}[K_h(G(X_i)-u)\varphi(W_i,w)]\} \\ &\quad + \frac{1}{n}\sum_{i=1}^n \{\mathbf{E}[K_h(G(X_i)-u)\varphi(W_i,w)] - g_\varphi(u,w)\} \\ &= V_n(G,u) + B_n(G,u) \text{ say.}\end{aligned}$$

We consider V_n first. Note that $K(u)$ is absolutely integrable and has Fourier transform $\tilde{K}(r) = (2\pi)\int e^{iru}K(u)du$ that satisfies $\int |\tilde{K}(r)|dr < \infty$. Let us consider the variance part first. Following Parzen (1962), we can bound $\max_{1\le i\le n}\sup_{(\varphi,\theta)\in\Phi_n\times\Theta_n}\sup_{(u,w)\in[0,1]\times\mathbf{R}^{d_W}} \left|\hat{\rho}_{\varphi,\theta,i}(u) - g_\varphi(u)\right|$ by

$$\begin{aligned}&\sup_{G\in\tilde{\mathcal{G}}_n}\sup_{u\in[0,1]}\int \left|e^{-iru/h}\right| \left|\frac{1}{nh}\sum_{j=1}^n \left\{e^{-irG(X_j)/h}\varphi(W_i,w) - \mathbf{E}[e^{-irG(X_j)/h}\varphi(W_i,w)]\right\}\right| \tilde{K}(r)dr \quad (40) \\ &\le \int \sup_{G\in\tilde{\mathcal{G}}_n}\left|\frac{1}{n}\sum_{j=1}^n \left\{e^{-irG(X_j)}\varphi(W_i,w) - \mathbf{E}[e^{-irG(X_j)}\varphi(W_i,w)]\right\}\right| \tilde{K}(hr)dr.\end{aligned}$$

We define $\Psi_n(r) = \mathbf{E}\left[\sup_{G\in\tilde{\mathcal{G}}_n}\left|\frac{1}{n}\sum_{j=1}^n \left\{e^{-irG(X_j)}\varphi(W_i,w) - \mathbf{E}[e^{-irG(X_j)}\varphi(W_i,w)]\right\}\right|\right]$. Then the L_1 norm of the last term in (40) is bounded by $h^{-1}\int \Psi_n(r)|\tilde{K}(r)|dr$. Later we will establish that

$$\sup_{r\in[-\pi,\pi]} \Psi_n(r) = O(n^{-1/2}b_n^{1/2}), \quad (41)$$

from which we deduce that

$$\max_{1\le i\le n}\sup_{(\varphi,\theta)\in\Phi_n\times\Theta_n}\sup_{u\in[0,1]}\left|\hat{\rho}_{\varphi,\theta,i}(u) - g_\varphi(u)\right| = O_p(n^{-1/2}h^{-1}b_n^{1/2}). \quad (42)$$

For the bias part B_n, we use Corollary A1 and follow the steps in (A.17) and (A.18) in Andrews (1995), p.591, to show that

$$\sup_{G\in\tilde{\mathcal{G}}_n}\sup_{u\in[0,1]}|B_n(G,u)| = O_p(h^{-2}).$$

We turn to the statement in (41). Define $\mathcal{V}_n = \{\exp(-irG(\cdot))\varphi(\cdot) : (r,G,\varphi) \in [0,1]\times\tilde{\mathcal{G}}_n\times\Phi_n'\times\mathcal{S}_W\}$. Note that $|e^{-ir_1G(x)} - e^{-ir_2G(x)}| \le |r_1 - r_2|$ and

$$|\exp[-irG_1(x)] - \exp[-irG_2(x)]| \le |G_1(x) - G_2(x)|.$$

By Theorem 2.7.11. of Van der Vaart and Wellner (1996), p.164, and by applying Corollary A1 above,

$$\begin{aligned}\log N_{[]}(C\varepsilon,\mathcal{V}_n,||\cdot||_p) &\le \log N(\varepsilon,\tilde{\mathcal{G}}_n,||\cdot||_{2p}) + \log N(\varepsilon,\Phi_n,||\cdot||_{2p}) \\ &\le \log N(\varepsilon,\Lambda_n,||\cdot||_{2p}) + \log N(\varepsilon,\Phi_n,||\cdot||_{2p}) + C/\varepsilon \\ &\le -C\log(\varepsilon) + b_n\varepsilon^{-b_\Phi} + C/\varepsilon.\end{aligned}$$

We obtain (41) by using Assumption B2 and the maximal inequality (Pollard (1989)). ■

Lemma B4 : (i) $\max_{1\leq i\leq n}\sup_{\theta:\theta\in\Theta_n}|U_{\theta,i}-U_i|=o_P(n^{-1/4})$.
(ii) $\sup_{\theta:\theta\in\Theta_n}\sup_x|F_{n,\theta,i}(\lambda_\theta(x))-F_\theta(\lambda_\theta(x))-\{F_{n,i}(\lambda_0(x))-F_0(\lambda_0(x))\}|=o_P\left(n^{-3/4}\right)$.

Proof of Lemma B4 : (i) Let $\Delta_n=\sup_{\theta\in\Theta_n}||\lambda_\theta-\lambda_0||_\infty$. Note that

$$\begin{aligned}|F_\theta(\lambda_\theta(X_i))-U_i| &\leq |\mathbf{E}[1\{\lambda_\theta(X_j)\leq\lambda_\theta(X_i)\}-1\{\lambda_0(X_j)\leq\lambda_0(X_i)\}|X_i]|\\ &\leq |\mathbf{E}[1\{\lambda_0(X_i)-2\Delta_n\leq\lambda_0(X_j)\leq\lambda_0(X_i)+2\Delta_n\}|X_i]|\leq C\Delta_n\end{aligned}$$

because the density of $\lambda_0(X_i)$ is uniformly bounded. Since $\Delta_n=o_P(n^{-1/4})$, we obtain the wanted result.
(ii) The LHS term is bounded by

$$\sup_{\theta:\theta\in\Theta_n}\sup_{\bar{\lambda}\in\mathcal{S}}\sup_{v\in[-\delta_n,\delta_n]}\left|\frac{1}{n}\sum_{j=1}^n\left[\Delta_\theta(X_j;\bar{\lambda},v)-\mathbf{E}\Delta_\theta(X_j;\bar{\lambda},v)\right]\right|$$

where $\Delta_\theta(X_j;\bar{\lambda},v)=1\left\{\lambda_\theta(X_j)\leq\bar{\lambda}+v\right\}-1\left\{\lambda_0(X_j)\leq\bar{\lambda}\right\}$. Observe that for each $(\theta_1,\bar{\lambda}_1,v_1)\in\Theta_n\times\mathcal{S}\times[-\delta_n,\delta_n]$,

$$\begin{aligned}&\left\{\mathbf{E}\left[\sup_{\theta_2\in\Theta_n:||\theta_1-\theta_2||<\delta}\sup_{\bar{\lambda}_2\in\mathcal{S}:|\bar{\lambda}_1-\bar{\lambda}_2|<\delta}\sup_{v_2\in[-\delta_n,\delta_n]:|v_1-v_2|<\delta}\left|\Delta_{\theta_1}(X_j;\bar{\lambda}_1,v_1)-\Delta_{\theta_2}(X_j;\bar{\lambda}_2,v_2)\right|^2\right]\right\}^{1/2}\\ \leq\ &\left\{2\mathbf{E}\left[1\left\{\bar{\lambda}_1+v_1-2\delta\leq\lambda_0(X_j)\leq\bar{\lambda}_1+v_1+2\delta\right\}\right]\right\}^{1/2}\leq C\delta^{1/2}.\end{aligned}$$

Therefore, for $\mathcal{H}=\left\{\Delta_\theta(\cdot;\bar{\lambda},v):(\theta,\bar{\lambda},v)\in\Theta_n\times\mathcal{S}\times[-\delta_n,\delta_n]\right\}$, where $\delta_n=o(n^{-1/4})$,

$$\begin{aligned}\log N_{[]}(\varepsilon,\mathcal{H},||\cdot||_2) &\leq \log N_{[]}(C\varepsilon^2,\Theta_n\times\mathcal{S}\times[-\delta_n,\delta_n],||\cdot||_{(1)})\\ &\leq \log N(C\varepsilon^2,\Theta_n,||\cdot||)+\log N(C\varepsilon^2,\mathcal{S},|\cdot|)+\log N(C\varepsilon^2,[-\delta_n,\delta_n],|\cdot|)\leq C\log\varepsilon.\end{aligned}$$

where $||(\theta_1,\bar{\lambda}_1,v_1)-(\theta_2,\bar{\lambda}_2,v_2)||_{(1)}=||\theta_1-\theta_2||+|\bar{\lambda}_1-\bar{\lambda}_2|+|v_1-v_2|$. Hence $\mathcal{H}$ is of polynomial discrimination (Pollard (1984)) and hence P-Donsker. This implies that

$$\begin{aligned}&\sup_{\theta:\theta\in\Theta_n}\sup_{\bar{\lambda}\in\mathcal{S}}\sup_{v\in[-\delta_n,\delta_n]}\left|\frac{1}{n}\sum_{j=1}^n\left[\Delta_\theta(X_j;\bar{\lambda},v)-\mathbf{E}\Delta_\theta(X_j;\bar{\lambda},v)\right]\right|\\ =\ &O_P\left(n^{-1/2}\left\|\left\{\mathbf{E}\left[\sup_{(\theta,\bar{\lambda},v)\in\Theta_n\times\mathcal{S}\times[-\delta_n,\delta_n]}\Delta_{\theta_1}^2(X_j;\bar{\lambda}_1,v_1)\right]\right\}^{1/2}\right\|\right)=o_P\left(n^{-3/4}\right).\end{aligned}$$

■

References

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

[2] Altonji, J. and R. Matzkin (2005), "Cross-section and panel data estimators for non-separable models with endogenous regressors," *Econometrica* 73, 1053-1102.

[3] Andrews, D. W. K (1994), "Empirical process method in econometrics," in *The Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden, Amsterdam: North-Holland.

[4] Andrews, D. W. K (1995), "Nonparametric kernel estimator for semiparametric models," *Econometric Theory* 11, 560-595.

[5] Andrews, D. W. K (1997), "A conditional Kolmogorov test," *Econometrica* 65, 1097-1128.

[6] Angrist, J. D. and G. M. Kuersteiner (2004), "Semiparametric causality tests using the policy propensity score," Working paper.

[7] Bahadur, R. R. (1960), "Stochastic comparison of tests," *Annals of Mathematical Statistics* 31, 276-295.

[8] Bierens, H. J. (1990), "A consistent conditional moment test of functional form," *Econometrica* 58, 1443-1458.

[9] Bierens, H. J. and W. Ploberger (1997), "Asymptotic theory of integrated conditional moment tests," *Econometrica* 65, 1129-1151.

[10] Birman and Solomjak (1997), "Piecewise-polynomial approximation of functions of the classes W_p^α," *Mathematics of the USSR-Sbornik* 2, 295-317.

[11] Cawley, J. and T. Phillipson (1999), "An empirical examination of information barriers to trade insurance," *American Economic Review* 89, 827-846.

[12] Chen, X. (2006), "Large sample sieve estimation of semi-nonparametric models", forthcoming in *Handbook of Econometrics*, Vol. 6, eds J. Heckman and E. Leamer.

[13] Chen, X., O. Linton, and I. van Keilegom (2003), "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica* 71, 1591-1608.

[14] Chernozhukov, V. and C. Hansen (2007), "An IV model of quantile treatment effects," *Econometrica* 73, 245-261.

[15] Chiappori, P-A, and B. Salanié (2000), "Testing for asymmetric information in insurance markets," *Journal of Political Economy* 108, 56-78.

[16] Chiappori, P-A, B. Jullien, B. Salanié, and F. Salanié (2002), "Asymmetric information in insurance: general testable implications," mimeo.

[17] Chung, K. L. (2001), *A Course in Probability Theory*, Academic Press, 3rd Ed.

[18] Dawid, P. (1979), "Conditional independence in statistical theory," *Journal of the Royal Statistical Society.* Series B. 41, 1-31.

[19] Dehejia, R. and S. Wahba (1999), "Causal effects in nonexperimental studies: reevaluating the evaluation of training programs," *Journal of the American Statistical Association* 94, 1053-1062.

[20] Delgado, M. A., W. González Manteiga (2001), "Significance testing in nonparametric regression based on the bootstrap," *Annals of Statistics* 29, 1469-1507.

[21] Escanciano, J. and K. Song (2007), "Asymptotic optimal test of single index restrictions with a focus on average partial effects," Working paper.

[22] Firpo, S. (2006), "Efficient semiparametric estimation of quantile treatment effects," *Econometrica* 75, 259-276.

[23] Hahn, J. (1998), "On the role of the propensity score in efficient semiparametric estimation of average treatment effects," *Econometrica* 66, 315-331.

[24] Härdle, W. and E. Mammen (1993), "Comparing nonparametric versus parametric regression fits," *Annals of Statistics* 21, 1926-1947.

[25] Heckman, J. J., J. Smith, and N. Clements (1997), "Making the most out of program evaluations and social experiments: accounting for heterogeneity in program impacts," *Review of Economic Studies*, 64, 421-471.

[26] Heckman, J. J., H. Ichimura, and P. Todd (1997), "Matching as an econometric evaluation estimator: evidence from evaluating a job training programme," *Review of Economic Studies* 64, 605-654.

[27] Heckman, J. J., H. Ichimura, J. Smith, and P. Todd (1998), "Characterizing selection bias using experimental data," *Econometrica* 66, 1017-1098.

[28] Hirano, K., G. W. Imbens, and G. Ridder (2003), "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71, 1161-1189.

[29] Janssen, A. (2000), "Global power functions of goodness of fit tests," *Annals of Statistics* 28, 239-253.

[30] Khmaladze, E. V. (1993), "Goodness of fit problem and scanning innovation martingales," *Annals of Statistics* 21, 798-829.

[31] LaLonde, R. (1986), "Evaluating the econometric evaluations of training programs with experimental data," *American Economic Review* 76, 604-620.

[32] Lehmann E. L. and J. P. Romano (2005), *Testing Statistical Hypotheses*, 3rd Edition, Springer, New York.

[33] Ledoux, M. and M. Talagrand (1988), "Un critére sur les pertite boules dans le théorème limite central," *Probability Theory and Related Fields* 77, 29-47.

[34] Linton, O. and P. Gozalo (1997), "Conditional independence restrictions: testing and estimation", Cowles Foundation Discussion Paper 1140.

[35] Liu, R. Y. (1988), "Bootstrap procedures under non i.i.d models," *Annals of Statistics* 16, 1696-1708.

[36] Mammen, E. (1993), "Bootstrap and wild bootstrap for high dimensional linear models," *Annals of Statistics* 21, 255-285.

[37] Phillips, P. C. B. (1988), "Conditional and unconditional statistical independence," *Journal of Econometrics*, 38, 341-348.

[38] Pollard, D. (1984) *Convergence of Stochastic Processes*. Springer-Verlag, New York.

[39] Pollard, D. (1989) A maximal inequality for sums of independent processes under a bracketing entropy condition. Unpublished manuscript.

[40] Orr, L, H. Bloom, S. Bell, W. Lin, G. Cave, and F. Doolittle (1995): *The National JTPA Study: Impacts, Benefits and Costs of Title II-A*. Bethesda, MD: Abt Associates.

[41] Rosenbaum, P. and D. Rubin (1983), "The central role of the propensity score in observational studies for causal effects," *Biometrika* 70, 41-55.

[42] Rosenblatt, M. (1952), "Remarks on a multivariate transform," *Annals of Mathematical Statistics* 23, 470-472.

[43] Smith, J. A. and P. E. Todd (2001), "Reconciling conflicting evidence on the performance of propensity-score matching methods," *American Economic Review* 91, 112-118.

[44] Smith, J. A. and P. E. Todd (2005), "Does matching overcome LaLonde's critique of nonexperimental estimators?" *Journal of Econometrics* 125, 305-353.

[45] Song, K. (2006), "Uniform convergence of series estimators over function spaces," Working Paper.

[46] Song, K. (2007), "Testing semiparametric conditional moment restrictions using conditional martingale transforms," Working Paper.

[47] Stinchcombe, M. B. and H. White (1998), "Consistent specification testing with nuisance parameters present only under the alternative," *Econometric Theory* 14, 295-325.

[48] Stute, W. (1997), "Nonparametric model checks for regression," *Annals of Statistics* 25, 613-641.

[49] Stute, W., W. González Manteiga, and M. Presedo Quindimil (1998): "Bootstrap approximations in model checks for regression," *Journal of the American Statistical Association* 93, 141-149.

[50] Stute, W. and L. Zhu (2005), "Nonparametric checks for single-index models," *Annals of Statistics* 33, 1048-1083.

[51] Su, L. and H. White (2003a), "Testing conditional independence via empirical likelihood," Discussion Paper, University of California San Diego.

[52] Su, L. and H. White (2003b), "A nonparametric Hellinger metric test for conditional independence," Discussion Paper, University of California San Diego.

[53] Su, L. and H. White (2003c), "A characteristic-function-based test for conditional independence," Discussion Paper, University of California San Diego.

[54] Turki-Moalla, K. (1998): "Rates of convergence and law of the iterated logarithm for U-processes," *Journal of Theoretical Probability* 11, 869-906.

[55] van der Vaart, A. (1996) New Donsker classes. *Annals of Probability* 24, 2128-2140.

[56] van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag.

[57] Whang, Y. (2000), "Consistent bootstrap tests of parametric regression functions," *Journal of Econometrics* 98, 27-46.

Table 1: Empirical Size and Power of Tests using Simulated Data and CV-KS

		Exp		Ind.	
DGP	κ	$n = 100$	$n = 300$	$n = 100$	$n = 300$
DGP 1	0	0.062	0.049	0.060	0.048
	0.5	0.432	0.885	0.395	0.875
DGP 2	0	0.082	0.056	0.084	0.052
	0.5	0.339	0.874	0.344	0.786

Table 2: Empirical Size and Power of Tests using Simulated Data and CV-CM

		Exp		Ind.	
DGP	κ	$n = 100$	$n = 300$	$n = 100$	$n = 300$
DGP 1	0	0.057	0.057	0.057	0.062
	0.5	0.430	0.881	0.405	0.894
DGP 2	0	0.079	0.047	0.072	0.038
	0.5	0.328	0.862	0.300	0.766

Table 3: Empirical Size and Power of Tests from Simulated Data for Quantile Treatment Effects via CV-KS

			Exp		Ind	
DGP	κ	quantile	$n = 100$	$n = 300$	$n = 100$	$n = 300$
DGP 1	0	0.2	0.080	0.063	0.087	0.051
		0.5	0.093	0.068	0.084	0.063
		0.8	0.083	0.069	0.077	0.071
	0.5	0.2	0.297	0.700	0.366	0.730
		0.5	0.422	0.832	0.431	0.833
		0.8	0.383	0.745	0.353	0.725
DGP 2	0	0.2	0.090	0.059	0.080	0.058
		0.5	0.088	0.057	0.088	0.063
		0.8	0.088	0.069	0.089	0.070
	0.5	0.2	0.509	0.936	0.416	0.872
		0.5	0.234	0.408	0.306	0.572
		0.8	0.187	0.405	0.141	0.374

Table 4: Empirical Size and Power of Tests from Simulated Data for Quantile Treatment Effects via CV-CM

			Exp		Ind	
DGP	κ	quantile	$n = 100$	$n = 300$	$n = 100$	$n = 300$
DGP 1	0	0.2	0.081	0.049	0.080	0.052
		0.5	0.088	0.055	0.910	0.046
		0.8	0.074	0.057	0.086	0.054
	0.5	0.2	0.282	0.698	0.359	0.742
		0.5	0.405	0.816	0.403	0.842
		0.8	0.349	0.746	0.320	0.735
DGP 2	0	0.2	0.078	0.067	0.073	0.068
		0.5	0.079	0.062	0.064	0.064
		0.8	0.067	0.076	0.085	0.055
	0.5	0.2	0.527	0.959	0.388	0.908
		0.5	0.214	0.455	0.294	0.623
		0.8	0.145	0.388	0.109	0.349

Table 5: The p-values from Testing $Y_0 \perp Z \,|p(X)$ using the JTPA data via CV-KS

Quarters	Exp	Ind
1	0.843	0.129
2	0.657	0.402
3	0.480	0.536
4	0.931	0.524
5	0.750	0.128
6	0.712	0.162
Joint	0.903	0.338

Table 6: The p-values from Testing $Y_0 \perp Z \,|p(X)$ using the JTPA data via CV-CM

Quarters	Exp	Ind
1	0.877	0.099
2	0.596	0.262
3	0.567	0.187
4	0.976	0.340
5	0.480	0.230
6	0.723	0.242
Joint	0.778	0.262

Table 7 : The p-values from Testing $1\{Y_0 \leq q_{0\tau}(X)\} \perp Z \mid p(X)$ using the JTPA data with $\beta_u(U) = \exp(Uu)$ and CV-KS

Quarters\Quantiles	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.764	0.634	0.318	0.480	0.779	0.461	0.526	0.922	0.171
2	0.680	0.589	0.890	0.837	0.564	0.118	0.106	0.950	0.962
3	0.715	0.733	0.518	0.858	0.886	0.205	0.152	0.109	0.213
4	0.744	0.726	0.392	0.427	0.782	0.594	0.403	0.670	0.683
5	0.753	0.740	0.708	0.395	0.709	0.521	0.504	0.832	0.663
6	0.685	0.733	0.930	0.312	0.385	0.744	0.531	0.248	0.073

Table 8 : The p-values from Testing $1\{Y_0 \leq q_{0\tau}(X)\} \perp Z \mid p(X)$ using the JTPA data with $\beta_u(U) = 1\{U \leq u\}$ and CV-KS.

Quarters\Quantiles	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
1	0.986	0.345	0.111	0.111	0.099	0.034	0.127	0.687	0.509
2	0.975	0.316	0.675	0.226	0.195	0.015	0.099	0.755	0.926
3	0.991	0.986	0.745	0.338	0.441	0.158	0.109	0.126	0.365
4	0.995	0.981	0.709	0.794	0.412	0.228	0.170	0.590	0.445
5	0.995	0.995	0.644	0.351	0.296	0.027	0.293	0.230	0.562
6	0.969	0.987	0.667	0.533	0.344	0.033	0.022	0.087	0.060